

# A Video Transmission System Based on Human Visual Model for Satellite Channel

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

\* Hughes Network Systems, Inc., 11717 Exploration Lane

Germantown, Maryland 20876, USA, Tel: +1-301-601-6494, Fax: +1-301-428-7177

Email: [yjiang@hns.com](mailto:yjiang@hns.com)

+ Institute for Systems Research, University of Maryland,  
College Park, MD 20742, USA

## Abstract

This paper presents a practical architecture for joint source-channel coding of human visual model based video transmission over satellite channel. Perceptual distortion model 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 proposed. 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.

Designated Symposia: Advanced Signal Processing for Communications

Key Topics: Video Processing in Multimedia Communication

## A Video Transmission System Based on Human Visual Model for Satellite Channel

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

\* Institute for Systems Research, University of Maryland, College Park, MD 20742, USA  
+ Hughes Network Systems Inc., 11717 Exploration Lane, Germantown, MD 20876, USA  
Tel: +1-301-601-6494, Fax: +1-301-428-7177, Email: [yjiang@hns.com](mailto:yjiang@hns.com)

**Abstract:** This paper presents a practical architecture for joint source-channel coding of human visual model based video transmission over satellite channel. Perceptual distortion model 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 proposed. 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.

### I. 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 [5]. Several modern human visual models are developed, such as just-noticeable-distortion (JND) [5][12], visible difference predictor (VDP) [8] and three-component image model [9]. The JND model provides each pixel with 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][3][4] 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][2]. 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. [15] gives a survey on 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 [16] 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 Ramsay 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.

### II THE JND MODEL BASED VIDEO CODEC

Figure 1 and Figure 2 show the JND model based video encoder and decoder respectively. In 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

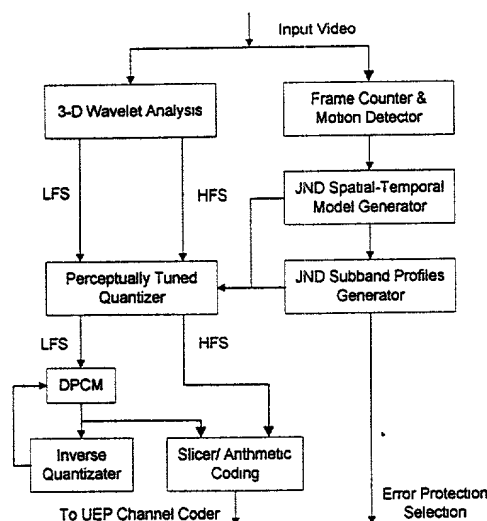


Figure 1 JND Based Video Encoder

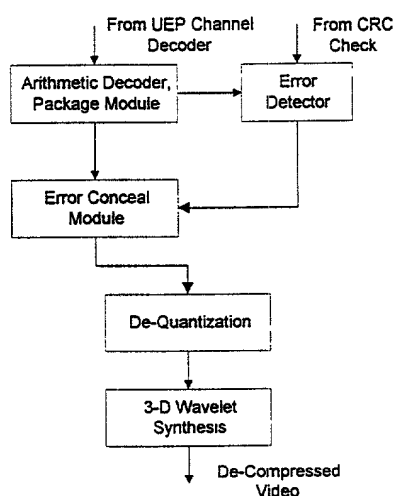


Figure 2 JND Based Video Decoder

spatio-temporal JND profile from analyzing local video signals and the distortion allocation algorithm that determines the JND profile for each subband. The Perceptually 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.

### 1. 3-D Wavelet Analysis

The two-tap Haar wavelet is adopted to proceed temporal analysis, Antonini (7,9) wavelet [10] 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.

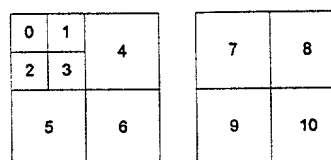


Figure 3 Subbands after 3-D Wavelet Decomposition

### 2. Frame Counter & Motion Detector

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

### 3. 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 [6][7], we use the same syntax, please refer [6][7] for explanation:

$$JND_{s-t}(x, y, n) \equiv f_3(\text{ild}(x, y, n)) \cdot JND_s(x, y, n)$$

$$JND_s(x, y, n) \equiv \max\{f_1(\text{mg}(x, y, n)), f_2(\text{mg}(x, y, n))\} \quad (1)$$

$$0 \leq x < W, 0 \leq y < H$$

### 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 [12]. The perceptual quality of the reconstructed image is accordingly expected to degrade evenly if bit-rate is reduced. MND is expressed as:

$$MND(x, y) \equiv JND(x, y) \cdot d(x, y) \quad (2)$$

where  $0 \leq x < W$ ,  $0 \leq y < H$ ,  $W$  and  $H$  are the width and height of an image respectively,  $d(x, y)$  is the distortion index at point  $(x, y)$ . We define the energy of MND of a small area indexed by  $(i, j)$  as:

$$\sum_{(x,y) \in r_{ij}} MND^2(x, y) \equiv \sum_{(x,y) \in r_{ij}} JND^2(x, y) \cdot \delta(i, j) \quad (3)$$

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:

$$\Delta_G = \frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L \varepsilon(k, l) \quad (4)$$

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 ( $r_y(k, l)$ ), i.e.,  $R_{kl} = \bigcup_{i=1, M, j=1, N} r_y(k, l)$ .

$\varepsilon(k, l)$  is defined as:

$$\varepsilon(k, l) = \text{median}(\delta(i, j) | r_y(k, l) \in R_{kl}, 1 \leq i \leq M, 1 \leq j \leq N) \quad (5)$$

The larger  $\Delta_G$  is, the larger the subjective perceptual distortion is. Compared with PSNR or MSE,  $\Delta_G$  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.

### 5. 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 [7][13].

First, the global object distortion index  $\Delta_G$  is given, which typically ranges from 1.0 to 10.0, where 1.0 stands for just noticeable distortion. Second, each subband is partitioned into non-overlapped 8 by 8 blocks ( $r_y(k, l)$ ). For each block  $r_y(k, l)$ , the step size of the quantizer is maximized under the condition that quantization error energy is less or equal to the MND energy in this block that has the distortion index  $\delta(i, j)$  equal to  $\Delta_G$ , i.e.,

$$\max_{\tau_y(k, l)} \left| \sum_{(x, y) \in r_y(k, l)} MND^2(x, y) \geq \sum_{(x, y) \in r_y(k, l)} w(x, y)^2 + \sum_{(x, y) \in r_y(k, l)} [w(x, y) - \hat{w}(x, y)]^2 \right| \quad (6)$$

where the energy of MND defined as (3),  $w(x, y)$  is wavelet coefficient,  $\hat{w}(x, y)$  is quantized wavelet coefficient.  $\tau_y(k, l)$  is the quantization step size of  $r_y(k, l)$ . A quantization table that leads uniform error energy over all subbands is setup for each subband. It is

packaged in the header of the bit stream of this subband.

### 6. Arithmetic Coding and Slicing Algorithm

Arithmetic coding [14] 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 such small bit stream carry the same amount of "distortion sensitivity". If we want to segment the subband  $S_i$  into  $I$  small bit streams, we can define  $I$  sets  $G_i$  of point  $(x, y)$ , such that for each set  $G_i$  ( $i = 1, \dots, I$ ):  $S_i = \bigcup_j G_j$  and

$$G_i = \left\{ (x, y) \left| \begin{array}{l} \sum_{(x, y) \in G_i} \frac{1}{JND^2(x, y)} = \frac{1}{I} \sum_{(x, y) \in S_i} \frac{1}{JND^2(x, y)} \\ (x, y) \in S_i, (x, y) \notin G_j, j \neq i \end{array} \right. \right\} \quad (7)$$

## III. ERROR CONCEALMENT

### 1. Channel Model

The satellite channel can be well modeled as an additive white Gaussian noise (AWGN) channel; the received base-band time-domain signal is as follows:

$$y(t) = \sum_{n=0}^{N-1} [(a_n g(t - nT) + j a_{qn} g(t - nT - \tau T)) \exp[j(2\pi f t + \theta)]] + n(t) \quad (8)$$

where  $g(t) = g_T(t) \otimes c(t) \otimes f(t)$ ,  $g_T(t)$  is transmitter shaping function,  $c(t)$  is channel response,  $f(t)$  is prefilter,  $n(t)$  is AWGN with two-sided power spectral density  $N_0/2$ ,  $a_n = a_m + j a_{qn}$  is data symbol from complex plane ( $a_n = A(\pm 1 \pm j)$ ) for OQPSK signaling,  $|a_n|^2 = E_s$ .  $\tau$  is delay factor, which is 0.5 for OQPSK. In our system simulation, we use square root raise cosine shaping filter with rolloff factor equal to 0.25.

### 2. RCPC Code, Unique Word and Their Performances

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 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 [16] 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 [16], we order the information bits from 11 subbands according to their source significance information (SSI) as  $S_{10} \dots S_0$ . The ordered information bits are shifted into the shift registers of a  $1/N$ , memory  $M$  convolutional encoder. During the  $n_{10}$  information bits from subband 10 the puncturing matrix  $a(l_{10})$  (we borrow the syntax used in [16]) 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  $n_i$  is an integer of the puncturing period  $P$ .

In order to keep the distortion constant and achieve stable Quality of Service (QoS), the source-coded data is variable bit rate (VBR) information. Some signs should be added in the data stream to inform the receiver the end of each subband. Highly correlated data pattern unique word (UW) is widely used in TDMA systems to identify the start of received bursts. UW could be used here to inform the receiver the beginning of a new subband. The following function is applied for unique word detection:

$$E_{uw} = \sum_{i=1}^n S_i \oplus U_i \quad (9)$$

where  $\oplus$  is inverse modulo 2 summation,  $n$  is the length of a UW,  $S \equiv (S_1, \dots, S_n)$  is received bit after hard-decision,  $U \equiv (U_1, \dots, U_n)$  is previously stored UW data pattern. If  $E_{uw}$  is larger than some threshold  $n - \varepsilon$ , the UW is said to be detected. The overall bit error probability is analyzed as following. The bit error rate upper bound of a rate  $P/(P+1)$  convolutional code is given by [17]:

$$P_b < \frac{1}{P} \sum_{k=d}^{\infty} C_k P_k \quad (10)$$

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

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 [17]:

$$P_b < \frac{1}{P} \left. \frac{dT(D, N)}{dN} \right|_{N=1, D=\sqrt{p(1-p)}} \quad (11)$$

where  $p$  is the transition probability of BSC. Suppose the UW length  $n$  is reasonable long,  $\varepsilon$  is properly selected, and the aperture technique is applied, the UW false detection probability can be ignored. A bit error probability upper bound is:

$$P_b < \sum_{i=0}^{\infty} \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=0}^{\infty} \binom{n}{i} p^i (1-p)^{n-i} \frac{1}{P} \sum_{k=d}^{\infty} C_k P_k \quad (12)$$

The first item above is UW miss detection probability. For OQPSK signal over AWGN channel,  $p$  is lower bounded by  $Q(\sqrt{\frac{E_s}{N_0}})$ . It is easy to achieve this bit error rate with the cutting-edge modem techniques.

If one would like to simplify the system design, constant bit rate (CBR) video stream is also a choice, i.e., each subband has fixed output data length. Only One frame unique word is required to synchronize the whole frame. Figure 4 shows the frame formats for convolutional encoded VBR and CBR video stream.

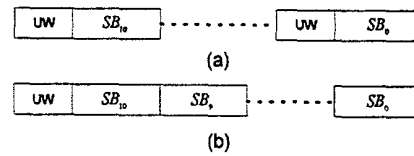


Figure 4 Frame Format (a) VBR (b) CBR

### 3. Rate Allocation Scheme

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  $l$  ( $l=0, \dots, 10$ ) as follows:

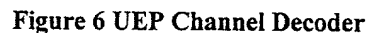
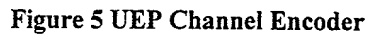
$$D_l = \frac{1}{H_l W_l} \sum_{(x,y) \in S_l} JND(x,y)^2 \quad (13)$$

where  $S_l$  is the set of pixels of subband  $l$ ,  $H_l$  and  $W_l$  are height and weight of it separately.  $D_l$  is an indication of the robustness of  $S_l$  to 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 "Calendar-Train".

$l$	$D_l$	$l$	$D_l$	$l$	$D_l$
0	4.5	4	7.5	8	27.6
1	7.6	5	7.5	9	27.6
2	7.6	6	8.7	10	42.8
3	8.8	7	7.4		

Table 1 Average Distortion  $D_l$  for Each Subband

From simulation we can see that  $D_l$  divides  $S_l$  into four categories,  $\{S_0\}$ ,  $\{S_1, S_2, S_3, S_4, S_5, S_7\}$ ,  $\{S_6, S_9\}$ ,  $\{S_{10}\}$ , which is intuitive for subband coding.



## 5. Error Concealment in Video Decoder

If errors are detected via CRC decoding in the blocks in subband  $S_i$ , these blocks are discarded. If no error is detected in error-corrupted received sequence, the arithmetic decoder can sometimes detect some conflicts during decoding, therefore find bad block and discard it. In worst that case errors are not detected, they will not spread to the whole subband or the whole frame due to slicing.

#### IV. Simulation Result

Figure 8 shows that our distortion measure (4) is better than PSNR in the sense that it reflects the subjective visual quality of image/video better. Figure 8 shows frame 1 in the decoded sequence of “Claire”. The PSNR of (a) and (b) are almost the same, but  $\Delta_G$  indicates that the distortion of (a) is smaller than that of (b) as we can tell from observation (e.g. shoulder, hair and cheek).

6

University of Maryland  
IS-99-030

Because the difference between  $S_{10}$  and  $S_8, S_9$  is relatively small, we finally assign 11 subbands into three error protection groups:  $\{S_0\}$ ,  $\{S_1, S_2, S_3, S_4, S_5, S_7\}$ ,  $\{S_6, S_8, S_9, S_{10}\}$ .

Index	4	5	6	7	8	9
Rate	8/18	8/16	8/14	8/12	8/10	8/9

Table 2 Rate Index of RCPC

Table 2 shows the coding rate index. The sequence of "Calendar-Train" are coded and transmitted over AWGN channel at different SNR. Figure 9 and Figure 10 show the distortion of first 10 decoded frames with different protection schemes at different SNR (In Figure 9 and Figure 10, the legend 3dB (4,7,8) means that the  $E_b/N_0$  is 3dB and we use channel coding rate 8/18 for subband 0, rate 8/12 for subband 1 to 7, and rate 8/10 for subband 8 to 10.)

In Figure 9 the original frames are encoded with the object distortion measure  $\Delta_c=1$ , which means the compression brings just noticeable distortion in the pictures. Even if the channel is ideal, the distortion  $\Delta_c$  goes larger gradually ( $\Delta_c=1.19$  for Frame 0 and 1.77 for Frame 5) in the following frames. The reason is that the JND model is initially got from the first two frames and is not renewed for the subsequent frames, which brings bias in encoding. The larger the frame number is, the more the bias accumulates

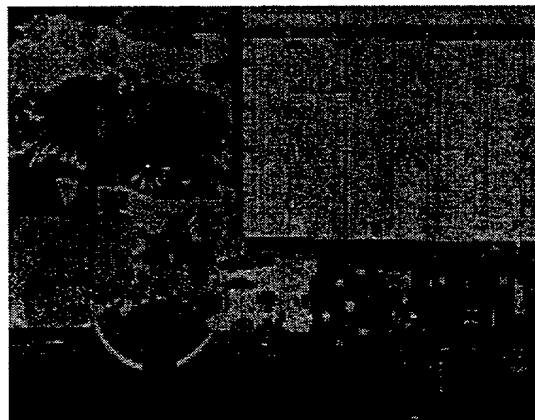
In Figure 10 the original frames are encoded with the object distortion measure  $\Delta_c=5$ , which means the perceptual distortion is 5 times of the just noticeable distortion. Intuitively, the distortion in Figure 7 should be less than that in Figure 8, however it is not the truth. Because the slices generated by source encoder are longer when  $\Delta_c=1$  (finer step size is chosen and more data is transmitted), the probability that a slice is corrupted by errors increases as the number of bits in this slice increases at the same bit error rate (BER). So if a video service requires better quality, the corresponding better channel protection scheme should be chosen.

Figure 11 shows frame no. 3 in recovered sequence of "Calendar-Train". Some areas corrupted by the channel noise can be observed.

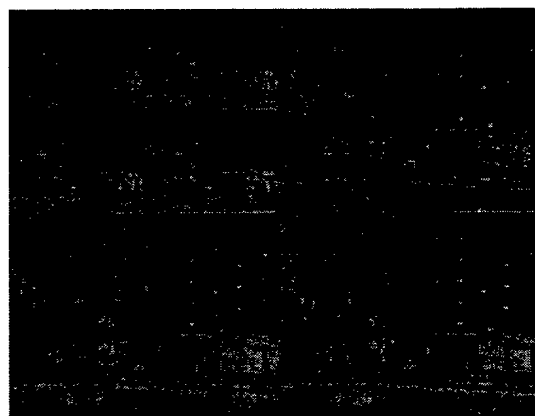
## V. CONCLUSION

We present a satellite video transmission system based on wavelet analysis and human vision model. The joint source channel coding scheme is investigated. We

propose a new performance index based on the JND model. The quantizer and slicer are perceptually optimized. Since we focus on 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.



(a)



(b)

Figure 7 (a) Frame 1 of "Calendar-Train" (b) JND Subband Profile for Subband 0 to 6

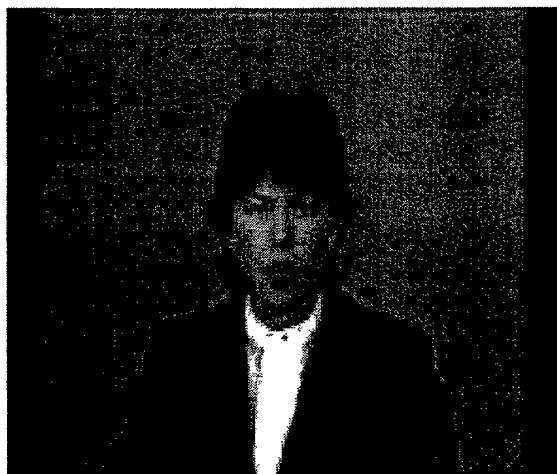


Figure 8-a Decoded Frames of "Claire",  $\Delta_G = 2.38$ ,  
PSNR=30.80dB

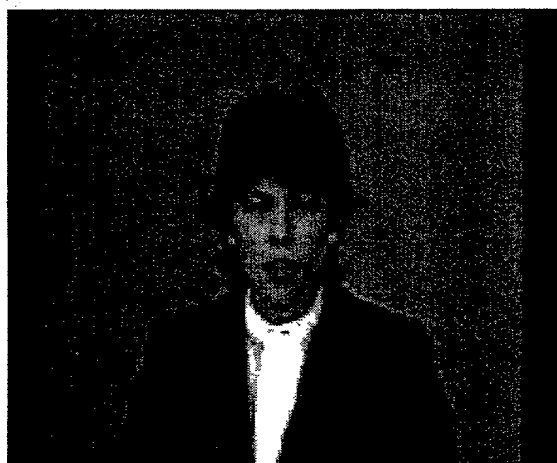


Figure 8-b Decoded Frame of "Claire",  $\Delta_G = 3.07$ ,  
PSNR=30.15dB

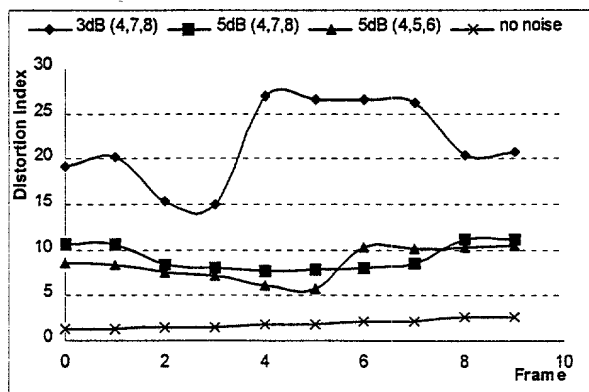


Figure 9 Distortion  $\Delta_G$  of the Decoded Frames over  
Noisy Channel with object distortion index 1

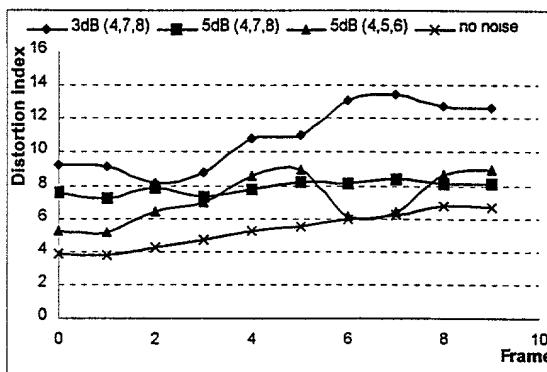


Figure 10 Distortion  $\Delta_G$  of the Decoded Frames  
over Noisy Channel with object distortion index 5

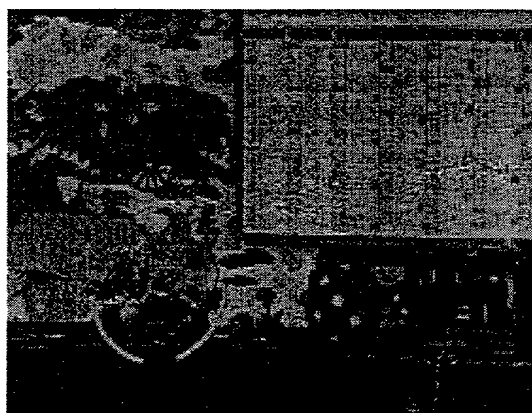


Figure 11 Decoded Frame 3 of "Calendar-Train"  
with  $E_b / N_o = 3dB$  (4,7,8) protection

The following references are hereby incorporated  
by reference:

1. Taubman and A. Zakhor, "Multirate 3-D subband coding of video", *IEEE Trans. Image Processing*, vol. 3, pp572-588, Sept. 1994.
2. C. Podilchuk, N. Jayant and N. Farvardin, "Three-dimensional subband coding of video", *IEEE Trans. Image Processing*, vol. 4, pp125-139, Feb. 1995.
3. D. Taubman, "Directionality and scalability in image and video compression", *Ph.D. dissertation*, Univ. California, Berkeley, 1994.
4. J. Y. Tham, S. Ranganath and A. Kassim, "Scalable low bit rate video compression using motion compensated 3-D wavelet decomposition",



University of Maryland  
IS-99-030

- in *IEEE ICCS/ISPACS 1996*, vol. 3, pp39.7.1-39.7.5, Nov. 1996.
5. N. Jayant, J. Johnston and R. Safranek, "Signal compression based on models of human perception", *Proc. IEEE*, vol. 81, pp1385-1422, Oct. 1993.
6. 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, pp467-476, Dec. 1995.
7. 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, pp143-156, April 1996.
8. S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity" in *Measurement and Prediction of Visual Quality*.
9. X. Ran and N. Farvardin, "A perceptually motivated three-component image model - Part I: Description of the model", *IEEE Trans. Image Processing*, vol. 4, pp401-415, April, 1995.
10. M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform", *IEEE Trans. Image Processing*, vol. 1, pp205-220, April 1992.
11. D. H. Kelly, "Motion and vision II. Stabilized spatio-temporal surface", *J. Opt. Soc. Amer.*, vol. 69, pp1340-1349, Oct. 1979.
12. N. Jayant, "Signal compression: Technology targets and research directions", *IEEE JSAIC*, vol. 10, pp796-818, June 1992.
13. N. Farvardin and J. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless source", *IEEE Trans. Information Theory*, vol. 30, pp485-497, May 1984.
14. I. Witten, R. M. Neal and J. G. Cleary, "Arithmetic coding for data compression", *Comm. of the ACM*, Vol. 30, pp520-540, June 1987.
15. 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.
16. J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and Their Application ", *IEEE Trans. Comm.*, April 1988, vol. 36, no.4, pp. 389-400.
17. A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems ", *IEEE Trans. Comm.*, Oct., 1971, vol. COM-19, no. 5, pp. 751- 772.

# A Perceptually Optimized Joint Source/Channel Coding Scheme for Video Transmission over Satellite Broadcast Channel

Yimin Jiang \*, Junfeng Gu +, John S. Baras + Fellow, IEEE

\* Hughes Network Systems Inc., 11717 Exploration Lane, Germantown, MD 20876

Tel: +1-301-601-6494, Fax: +1-301-428-7177, e-mail: [yjiang@hns.com](mailto:yjiang@hns.com)

+ Institute for Systems Research, University of Maryland, College Park, MD 20742

Tel: +1-301-405-6606, Fax: +1-301-314-9218, e-mail: [fengqing@isr.umd.edu](mailto:fengqing@isr.umd.edu), [baras@isr.umd.edu](mailto:baras@isr.umd.edu)

## ABSTRACT

This paper presents a perceptually optimized joint source/channel coding scheme for video transmission over satellite broadcast channel. We express the end-to-end expected distortion in terms of the source coder distortion and the channel distortion. Perceptual distortion model just-noticeable-distortion (JND) is applied to improve the subjective quality of compressed video. 3-D wavelet video codec that is jointly optimized with channel codec can remove spatial and temporal redundancy and provide the scalability of video quality for broadcast channel. In order to conceal the errors under bad channel conditions and prevent the error propagation of arithmetic decoding, slicing method and a channel coding scenario that combines RCPC, Reed-Solomon codes and CRC are proposed. Two perceptual distortion indices based on JND are discussed. These two indices are combined in our joint source/channel coding scenario to achieve minimal bandwidth consumption subject to acceptable visual quality under normal and bad channel conditions. Computer simulation is conducted at difference signal-to-noise (SNR) ratios. Our system uses offset QPSK (OQPSK) modulation scheme to combat the nonlinear effect of satellite channels.

## I. INTRODUCTION

With the rapid advances in video compression techniques and digital communications, high quality digital video broadcasting via satellites is of great commercial interests nowadays. Compared with digital data communications, video transmission has some salient attributes [20]. First, the performance metric is the delivered visual quality of the source due to both quantization and channel distortion under constraints of fixed system resource like bandwidth and transmission energy. This metric contrast with commonly used performance index like bit error rate (BER). Second, the existence of unequal error sensitivities in a typical video bit stream emphasizes the desirability of a layered approach to both source and channel coding. Third, due to the stringent delay requirements of synchronous video applications, there is a need to include finite buffer constraints. These requirements will affect the choice of error control coding schemes like forward error correction (FEC).

Because the ultimate objective of video transmission systems is to maintain the subjective visual quality of images, performance metrics (other than mean-square-error (MSE) or peak-signal-noise-ratio (PSNR)) that take the psychovisual properties of human visual system (HVS) into account are proposed [5]. Several modern human visual models are developed, such as just-noticeable-distortion (JND) [5][12], visible difference predictor (VDP) [8] and three-component image model [9]. The JND model provides each pixel with 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][3][4] 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][2]. The JND model and 3-D wavelet decomposition are applied in our video codec. The quantizer is based on the JND model 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

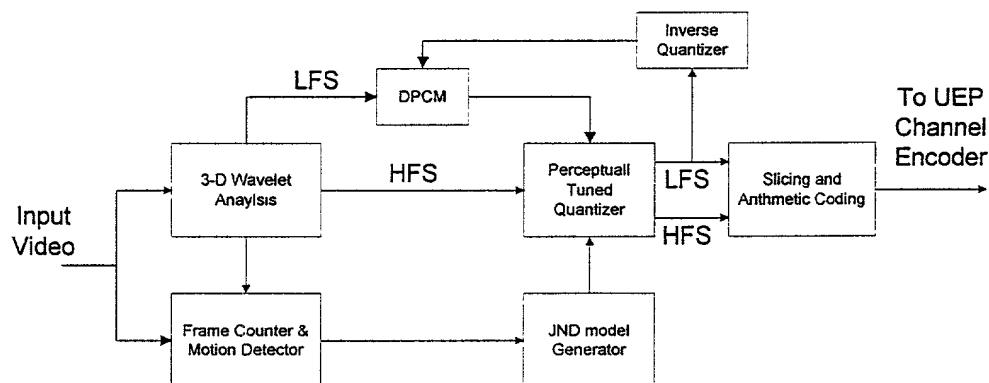
system with finite delay and finite complexity in source and channel coders there are advantages in joint source-channel coding. [15][19][20] discuss the recent progress on it.

As indicated in [19][20], digital broadcast differs from digital point-to-point transmission in that different receivers have different channel capacities that depend on channel conditions such as location, weather. Cover [22] showed that one could trade off channel capacity from the poor channel to the better ones, and that this trade off can in theory be worthwhile. These ideas point out the efficiency of using multiresolution approach to digital broadcast. In satellite broadcast case feedback channel is not available, thus the transmitter has no information about the receivers and their channel conditions. It is difficult to guarantee the average video qualities under diversified channel conditions without large channel coding overhead. In this work we adopt multiresolution source codec and unequal error protection (UEP) channel coders and optimize the overall bandwidth consumption subject to acceptable end-to-end perceptual distortion based on the joint design of video quantizer and the selection of channel codes.

The outline of the paper is as follows. Section II presents the JND model based video codec. Section III discusses the error concealment scenario. We first describe the rate compatible punctured convolutional (RCPC) codes [16] roughly. A slicing method that truncates the data stream from each subband into small packets to prevent the error propagation of arithmetic decoding is adopted. Reed-Solomon codes 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 the error detection capability. Section IV proposes two subjective performance indices based on the JND model. The overall bandwidth is optimized based on these two indices. The first index is used as a constraint in our source coding scheme to assure the video quality under normal channel conditions, the second index that combines channel distortion is used as a constraint at bad channel conditions. Performance is analyzed in details. Section V shows computer simulation results. Concluding remarks are provided in Section VI.

## II. THE JND MODEL BASED VIDEO CODEC

Our video codec structure is based on the work in [7]. New quantization scenario (jointly optimized with channel encoder), slicer and entropy coding module are new features. In order to make this paper self-complete, the video codec is roughly described. Figure 1 and Figure 2 show the JND model based video encoder and decoder respectively. In 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. The JND model generators estimate the spatio-temporal JND profile from analyzing local video signals, the distortion allocation algorithm determines the JND profile for each subband. The perceptually tuned quantizer quantizes the wavelet coefficients in each subband according to their JND profiles. The spatial LLLL temporal L subband is encoded by DPCM module. The data stream from each subband goes through the slicer which break the data into small packets to prevent the error propagation of arithmetic decoding before entropy coding. Because the main purpose of this paper is to present joint source/channel coding scheme, more powerful compression schemes (e.g. zero-tree, motion estimation and run-length coding) are not applied in our video codec. These modules are presented subsequently.



**Figure 1 JND Based Video Encoder**

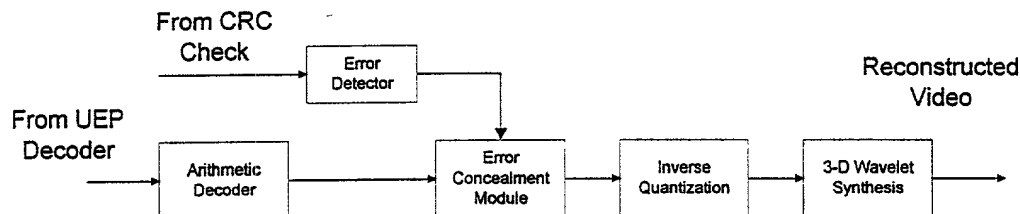


Figure 2 JND Based Video Decoder

### 1. 3-D Wavelet Analysis

A two-tap Haar wavelet is adopted to proceed temporal analysis, Antonini (7,9) wavelet [10] 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.

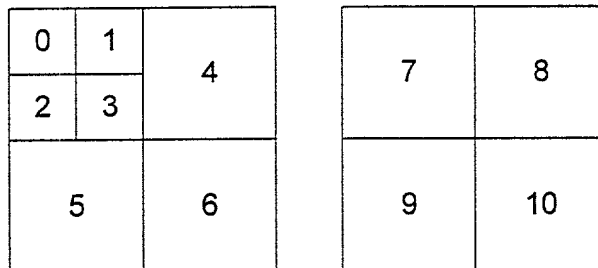


Figure 3 Subbands after 3-D Wavelet Decomposition

### 2. Frame Counter & Motion Detector

Because the calculation of the JND profiles is computational resource consuming, the frame counter & motion Detector is designed to control the renew process of the JND profiles. 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 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.

### 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 are as [6][7]. The same syntax is used, please refer to [6][7] for explanation:

$$\begin{aligned} JND_{s-t}(x, y, n) &\equiv f_3(\text{ild}(x, y, n)) \cdot JND_s(x, y, n) \\ JND_s(x, y, n) &\equiv \max\{f_1(\text{mg}(x, y, n)), f_2(\text{mg}(x, y, n))\} \quad 0 \leq x < W, 0 \leq y < H \end{aligned} \quad (1)$$

where  $f_1$  is the error visibility threshold due to texture masking,  $f_2$  is the visibility threshold due to average background luminance,  $W$  and  $H$  are the width and height of the image respectively,  $\text{mg}(x, y)$  is the maximal weighted average of the luminance gradients around the pixel at  $(x, y)$ ,  $\text{bg}(x, y)$  is the average background luminance. From spatio-temporal JND profiles that measure the perceptual redundancy inherent in full-band video signals quantitatively, the JND profiles for each subband are set up with certain distortion allocation algorithm [11] for different subbands. As a result, the JND profile for subband  $i$  is the weighted sum of spatio-temporal JND values at the corresponding locations. The weights indicate the perceptual importance of this subband  $i$  [11].

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 [12]. The perceptual quality of the reconstructed image is accordingly expected to degrade evenly if bit-rate is reduced. The MND of pixel  $(x, y)$  is expressed as:

$$MND(x, y) \equiv JND(x, y) \cdot d(x, y) \quad (2)$$

where  $0 \leq x < W$ ,  $0 \leq y < H$ ,  $d(x, y)$  is the distortion index at point  $(x, y)$ , which is given as a system parameter typically. The energy of MND of a small block indexed by  $(i, j)$  is defined as:

$$E_{MND}(i, j) \equiv \sum_{(x, y) \in r_{ij}} MND^2(x, y) = \sum_{(x, y) \in r_{ij}} JND^2(x, y) \cdot \delta(i, j) \quad (3)$$

where  $r_{ij}$  is a small block (typically 8 by 8 pixels),  $\delta(i, j)$  is the distortion index for this

block.  $\delta(i, j)$  in reconstructed image can be measured by calculating 
$$\frac{\sum_{x \in r_{ij}} \|x - \hat{x}\|^2}{\sum_{x \in r_{ij}} JND(x)^2},$$

where  $x$  is the luminance of a pixel in block  $r_{ij}$ ,  $\hat{x}$  is its reconstructed value.

#### 4. Perceptually 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 [7] [13].

If the system cost function is bandwidth as explained in section IV,  $\delta(i, j)$  (typically a constant ranging from 1.0 to 10.0 for all blocks in all subbands) can be carefully selected and each subband is partitioned into non-overlapped 8 by 8 blocks. For each 8 by 8 block, the step size of the quantizer is maximized subject to the condition that the quantization error energy is less or equal to the MND energy (3) in this block. A quantization table that leads uniform error energy over all subband is setup for each subband. The table is packaged in the header of the bit stream of this subband.

The quantizer design can be jointly optimized with channel coding rate allocation. We will address this problem in Section IV.

#### 5. Arithmetic Coding And Slicing

Arithmetic coding [14] 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 error from spreading, slicing data streams into small packets before arithmetic coding is necessary.



### III. ERROR CONCEALMENT

#### 1. Channel Model

The satellite channel can be well modeled as an additive white Gaussian noise (AWGN) channel; the received base-band time-domain signal is as follows:

$$y(t) = \sum_{n=0}^{N-1} [(a_n g(t-nT) + ja_{Qn} g(t-nT-\tau T)) \exp[j(2\pi ft + \theta)]] + n(t) \quad (4)$$

where  $g(t) = g_T(t) \otimes c(t) \otimes f(t)$ ,  $g_T(t)$  is transmitter shaping function,  $c(t)$  is channel response,  $f(t)$  is prefilter,  $n(t)$  is AWGN with two-sided power spectral density  $N_o/2$ ,  $a_n \equiv a_m + ja_{Qn}$  is data symbol from complex plane ( $a_n = A(\pm 1 \pm j)$  for OQPSK signaling,  $|a_n|^2 = E_s$ ).  $\tau$  is delay factor, which is 0.5 for OQPSK. OQPSK is good at combating the nonlinear effect of the amplifiers on satellites. In our system simulation, we use square root raise cosine shaping filter with rolloff factor equal to 0.25.

#### 1. RCPC Codes and Its Performance Bounds

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 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 [16] 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 [16], we order the information bits from 11 subbands according to their source significance information (SSI) as  $S_{10} \dots S_0$ . The ordered information bits are shifted into the shift registers of a  $1/N$ , memory  $M$  convolutional encoder. During the  $n_{10}$  information bits from subband 10 the puncturing matrix  $a(l_{10})$  (we borrow the syntax used in [16]) 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 an integer of

the puncturing period  $P$ . The bit error rate upper bound of a rate  $P/(P+1)$  convolutional code is given by [17]:

$$P_B < \frac{1}{P} \sum_{k=d}^{\infty} C_k P_k \quad (5)$$

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

$$\left. \frac{dT(D, N)}{dN} \right|_{N=1} = \sum_{k=d}^{\infty} C_k D^k, \quad P_k \text{ is probability that the wrong path with distance } k \text{ to the}$$

correct path is selected. A more relax upper bound for binary symmetric channel (BSC) is given by [17]:

$$P_B < \frac{1}{P} \left. \frac{dT(D, N)}{dN} \right|_{N=1, D=\sqrt{p(1-p)}} \quad (6)$$

where  $p$  is the transition probability of BSC. The BSC is a good model for OQPSK

signal over AWGN channel,  $p$  is lower bounded by  $Q\left(\sqrt{\frac{E_s}{N_o}}\right)$ . It is easy to achieve this

bit error rate with the cutting-edge modem techniques even at low SNR.

## 2. UEP Channel Encoder and Decoder

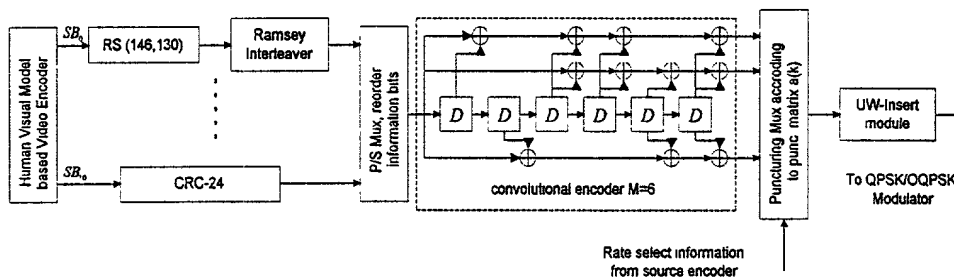
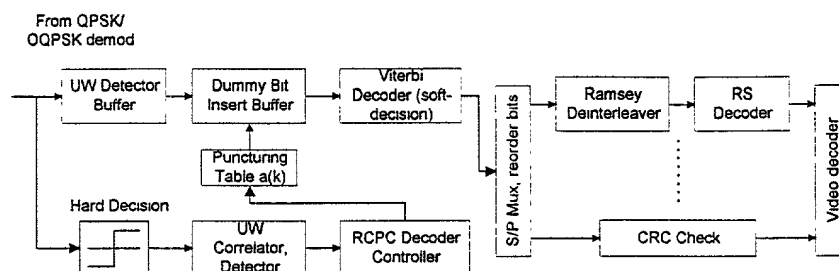


Figure 4 UEP Channel Encoder



**Figure 5 UEP Channel Decoder**

Figure 4 shows the 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 more protection. We select a family of RCPC codes (Table II in [16], 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  $S_i$  before arithmetic decoding, we discard this block. 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.

#### IV. JOINT SOURCE/CHANNEL CODING

##### 1. End-to-End Distortion

To assess the perceptual quality of reconstructed image and to rate the performance of the proposed joint source/channel coding algorithm, an effective fidelity criterion is needed. The peak signal-to-noise ratio

$$PSNR \equiv 20 \log_{10} \frac{255}{\sqrt{MSE}} \quad (7)$$

is widely used to measure the image quality. It however cannot accurately reflect the real perceptual quality of reconstructed image. Consequently a quantitative fidelity measure that is conformable to the subjective evaluation of image quality is desired. In this paper, we discuss two computable fidelity measures. One distortion measure is proposed in [23], which is based on evaluating  $\delta(i, j)$  (3) of reconstructed image:

$$\Delta_G \equiv \frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L \varepsilon(k, l) \quad (8)$$

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:

$$\varepsilon(k, l) \equiv \text{median}(\delta(i, j) | r_{ij}(k, l) \in R_{kl}, 1 \leq i \leq M, 1 \leq j \leq N) \quad (9)$$

$\Delta_G$  ranges from 1.0 to 10.0, 1.0 means that the distortion is not perceptible. The larger  $\Delta_G$  is, the larger the subjective perceptual distortion is.  $\Delta_G$  is set to equal  $\delta(i, j)$  as a system design parameter before the quantizer is designed. Simulation shows that if  $\delta(i, j) = \Delta_G^*$  ( $\Delta_G^*$  is the target distortion) for each small block before quantization and there is no noise in channel,  $\Delta_G$  of reconstructed image is very close to  $\Delta_G^*$ . However  $\Delta_G$  is intractable to analyze mathematically.

In this paper we also propose a distortion measure called “perceptible mean square error” that is similar to Chou’s measure in [7] and easy to analyze, it is defined as follows:

$$D_{tot} = E \left[ \max(\|x - \hat{x}\|^2 - JND(x)^2, 0) \right] \quad (10.a)$$

where  $x$  is the luminance value of a pixel in input image (typically with zero mean) and  $\hat{x}$  is the reconstructed luminance value,  $JND(x)$  is the JND value of the pixel. Equation (10) reflects the perceptible distortion energy and the way we design the quantizer. Since a distortion measure must be a non-negative number, we have to either

assure  $\|x - \hat{x}\|^2 \geq JND(x)^2$  which means that the average distortion energy should be larger than the average JND energy or set it to zero. In order to simplify the presentation, it is reasonable to assume (for tight bit budget) that  $\|x - \hat{x}\|^2$  is larger than  $JND(x)^2$  with high probability under bad channel conditions, therefore,  $D_{tot}$  is approximated at low SNR as

$$D_{tot} = E[\|x - \hat{x}\|^2] - E[JND(x)^2] \quad (10.b)$$

According to the orthogonality of Karhunen-Loeve Transform [21],

$$E[\|x - \hat{x}\|^2] = \sum_{i=0}^{N-1} E[\|x_i - \hat{x}_i\|^2] \quad (11)$$

where  $x_i$  are the wavelet coefficients of subband  $i$ ,  $\hat{x}_i$  are the reconstructed wavelet coefficients,  $N$  is the total number of subbands. In nonorthogonal case, the equation above does not hold, however well designed linear phase FIR filter banks (that it, with good out-of-band rejection) are often close to being orthogonal and thus satisfy the equation above approximately. For each subband  $i$ , the data stream is broken into small packets as described in section III. The following assumptions are made: the length of each packet after arithmetic coding and channel encoding roughly equals to a constant  $L_i$ , CRC codes can detect all the errors in each packet. We can get

$$E[\|x_i - \hat{x}_i\|^2] = P_{ei} E[\|x_i\|^2] + (1 - P_{ei}) D_{qi} \quad (12)$$

where  $P_{ei}$  is the probability that a packet of subband  $i$  is corrupted by channel errors,  $D_{qi}$  is the distortion caused by perceptually tuned quantization,  $D_{qi} = E[\|x_i - x_{qi}\|^2]$ ,  $x_{qi}$  is quantized value of  $x_i$ . The first part of equation (12) is the channel distortion, the second part is the source coder distortion. It is reasonable to model the channel with RCPC codes and Reed Solomon codes as a memoryless BSC channel with bit error probability  $p(r_{ci})$  if the packet length is long enough, where  $r_{ci}$  is the channel coding rate of subband  $i$ .  $p(r_{ci})$  can be approximated by computer simulation for specific codes. We

define that a packet is corrupted by channel errors if one or more errors occur in this packet, therefore  $P_{ei}$  is

$$P_{ei} = 1 - [1 - p(r_{ci})]^L \quad (13)$$

According to the work in [7], the JND profile for each subband is given by

$$E[JND(x_i)^2] = E[JND(x)^2] w_i, \quad i = 0, \dots, 10 \quad (14)$$

where  $w_i$  are the weighting functions that can be derived as the relative desensitivity of the HVS to each subband.

## 2. Joint Optimization

In order to assure that the video quality under most circumstances is acceptable and the coding cost is moderate, these two performance indices  $\Delta_G$  (8) and  $D_{tot}$  (10) are combined to provide the following scenario.  $\Delta_G^*$  is given as subjective quality requirement for normal channel conditions,  $D_{tot}$  is constrained within some threshold  $D_r$  at the lowest operating SNR, i.e., subject to the following conditions:

1. For each subband  $i$ ,  $D_{qi} \leq \delta_i E[JND(x_i)^2]$ ,  $\delta_i$  (equals to  $\Delta_G^*$ ) and  $p(r_{ci})$  at lowest operating SNR are set
2. The end-to-end distortion  $D_{tot} \leq D_r$ , where  $D_r$  is the tolerable average end-to-end distortion under worst channel condition, ordinarily  $D_r = [(\Delta_G^* - 1) + K_{ch}] E[JND(x)^2]$

we try to minimize the overall bandwidth  $\sum_{i=0}^{N-1} \frac{r_{si}}{r_{ci}}$ , where  $r_{si}$  is the source coding rate for subband  $i$ .

For mid-rising uniform quantizer, the number of quantization step  $M_i$  and the step size  $\Delta_i$  for subband  $i$  are closely related to  $r_{si}$ , which is given by

$$r_{si} = \lfloor \log_2 M_i \rfloor \quad (15)$$

where  $\lfloor x \rfloor$  is the smallest integer larger than  $x$ .

Let us assume that  $f_i(x)$  is the probability density function (pdf) of the wavelet coefficient of subband  $i$ , and  $f_i(x)$  is symmetric about 0.  $D_{qi}$  and  $E[\|x_i\|^2]$  can be calculated as follows,

$$E[\|x_i\|^2] = \int_{-\infty}^{\infty} x^2 f_i(x) dx \quad (16)$$

$$D_{qi} = 2 \left[ \int_0^{a_i} x^2 f_i(x) dx + \sum_{l=2}^{M_i/2-1} \int_{(l-1)\Delta_i}^{l\Delta_i} (x - (l-1/2)\Delta)^2 f_i(x) dx + \int_{(M_i/2-1)\Delta_i}^{M_i/2\Delta_i} (x - (M_i/2-1/2)\Delta)^2 f_i(x) dx \right] \quad (17)$$

Equation (17) shows that  $M_i$ ,  $\Delta_i$  and  $D_{qi}$  are closely related to the pdf of  $x_i$ . This problem will be addressed for different pdfs. We first start from uniform distribution, then expand the solution to more general cases, e.g.,  $x_i$  satisfies so-called General Gaussian distribution.

i. Uniform Distribution

$$f_i(x) = \frac{1}{2a_i}, \quad -a_i \leq x \leq a_i \quad (18)$$

after some simple arithmetic, we can get

$$E[\|x_i\|^2] = \frac{a_i^2}{3} \quad (19)$$

$$D_{qi} = \frac{a_i^2}{3M_i} \quad (20)$$

$$\text{and } \Delta = \frac{2a_i}{M_i}.$$

ii. General Gaussian Distribution

According to the work done by Mallat [18], the wavelet coefficients follow a distribution called General Gaussian distribution in stead of uniform distribution:

$$f_i(x) = K e^{-(|x|/\alpha)^\beta} \quad (21)$$

$$K = \frac{W\beta}{2\alpha\Gamma(\frac{1}{\beta})}, \Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$$

where  $W$  is the total number of pixels in the given detail picture. Typically the wavelet coefficients of low pass subbands satisfy Gaussian distribution, i.e.,  $\beta \approx 2$ , and the wavelet coefficients of high pass subbands satisfy Laplacian distribution, i.e.,  $\beta \approx 1$ . We address each case separately in the following:

- Gaussian Distribution

Let us assume

$$f_i(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (22)$$

then

$$E[\|x_i\|^2] = \sigma^2 \quad (23)$$

By applying the following arithmetic,

$$(x - (l-1/2)\Delta_i)^2 = x^2 - (2l-1)\Delta_i x + (l-1/2)^2 \Delta_i^2,$$

$$\int_a^b x f_i(x) dx = \frac{\sigma}{\sqrt{2\pi}} \left( e^{-\frac{a^2}{2\sigma^2}} - e^{-\frac{b^2}{2\sigma^2}} \right), \quad \text{and} \quad \int_a^b f_i(x) dx = Q\left(\frac{a}{\sigma}\right) - Q\left(\frac{b}{\sigma}\right), \quad \text{where}$$

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \approx \left[ \frac{1}{(1-c)x + c\sqrt{x^2 + d}} \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad c = \frac{1}{\pi}, d = 2\pi$$

we can get  $D_{qi}$  as follows,

$$\begin{aligned} D_{qi} = & \sigma^2 - 2 \sum_{l=2}^{M_i/2-1} (2l-1) \frac{\Delta_i \sigma}{\sqrt{2\pi}} \left( e^{-\frac{(l-1)^2 \Delta_i^2}{2\sigma^2}} - e^{-\frac{l^2 \Delta_i^2}{2\sigma^2}} \right) - 2(M_i-1) \frac{\Delta_i \sigma}{\sqrt{2\pi}} e^{-\frac{(M_i/2-1)^2 \Delta_i^2}{2\sigma^2}} \\ & + 2 \sum_{l=2}^{M_i/2-1} (l-1/2)^2 \Delta_i^2 \left[ Q\left(\frac{(l-1)\Delta_i}{\sigma}\right) - Q\left(\frac{l\Delta_i}{\sigma}\right) \right] + 2\left(\frac{M_i}{2} - \frac{1}{2}\right)^2 \Delta_i^2 Q\left(\frac{(M_i/2-1/2)\Delta_i}{\sigma}\right) \end{aligned}$$

After some further simplification, we can get

$$D_{qi} = \sigma^2 + 2 \left[ \frac{9}{4} Q\left(\frac{\Delta_i}{\sigma}\right) + \sum_{l=2}^{M_i/2-1} 2l Q\left(\frac{l\Delta_i}{\sigma}\right) \right] \Delta_i^2 - \frac{2\Delta_i \sigma}{\sqrt{2\pi}} \left( 3e^{-\frac{\Delta_i^2}{2\sigma^2}} + 2 \sum_{l=2}^{M_i/2-1} e^{-\frac{l^2 \Delta_i^2}{2\sigma^2}} \right) \quad (24)$$



Adding the conditions that  $M_i$  is the least positive integer and  $\Delta_i > 0$ , we can get  $M_i$  and  $\Delta_i$  given desired  $D_{qi}$  by solving the equation above.

- Laplacian Distribution

Let us assume that

$$f_i(x) = \frac{1}{2\alpha} e^{-\frac{|x|}{\alpha}} \quad (25)$$

then we can get

$$E[\|x_i\|^2] = 2\alpha^2 \quad (26)$$

after some arithmetic, we can derive the following  $D_{qi}$

$$D_{qi} = 2\alpha^2 - e^{-\frac{\Delta_i}{\alpha}} \left( \frac{3\Delta_i^2}{4} + 3\Delta_i\alpha \right) - 2 \sum_{l=2}^{M_i/2-1} \Delta_i \alpha e^{-\frac{l\Delta_i}{\alpha}} \quad (27)$$

We can apply the same conditions used in Gaussian case, then we can calculate  $M_i$  and  $\Delta_i$  given desired  $D_{qi}$ .

In the calculation above, the analytical solution of  $M_i$  and  $\Delta_i$  is intractable. Numerical result can be obtained via the help of computer. For different expected  $D_{qi}$ , different  $M_i$  and  $\Delta_i$  can be obtained from numerical solution, then a lookup table can be made for each subband  $i$  according to its statistical properties and different distortion requirement. The lookup table is used to decide the quantization parameters for each 8 by 8 block in each subband  $i$  with different JND profiles. After finishing every step described above, we can get the end-to-end distortion  $D_{tot}$  as a function of  $r_{si}$  and  $r_{ci}$ .

The combination of  $\Delta_G$  (that equals to  $\delta_i$ )  $D_{tot}$  achieves perceptual optimum and is beneficial for broadcast channel.  $\Delta_G$  considers the effect of normal channel conditions ( $P_{ei} \approx 0$ ),  $D_{tot}$  considers the worst case. Subject to the conditions explained before, the overall bandwidth  $\sum_{i=0}^{N-1} \frac{r_{si}}{r_{ci}}$  is minimized. This rate allocation problem can be solved by optimization methods such as dynamic programming.

## Reference:

The following references are hereby incorporated by reference:

1. D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video", *IEEE Trans. Image Processing*, vol. 3, pp. 572-588, Sept. 1994.
2. C. Podilchuk, N. Jayant and N. Farvardin, "Three-dimensional subband coding of video", *IEEE Trans. Image Processing*, vol. 4, pp. 125-139, Feb. 1995.
3. D. Taubman, "Directionality and scalability in image and video compression", *Ph.D. dissertation*, Univ. California, Berkeley, 1994.
4. J. Y. Tham, S. Ranganath and A. Kassim, "Scalable low bit rate video compression using motion compensated 3-D wavelet decomposition", in *IEEE ICCS/ISPACS 1996*, vol. 3, pp. 39.7.1-39.7.5, Nov. 1996.
5. N. Jayant, J. Johnston and R. Safranek, "Signal compression based on models of human perception", *Proc. IEEE*, vol. 81, pp. 1385-1422, Oct. 1993.
6. 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.
7. 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, April 1996.
8. S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity" in *Measurement and Prediction of Visual Quality*.
9. 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.
10. 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.
11. D. H. Kelly, "Motion and vision II. Stabilized spatio-temporal surface", *J. Opt. Soc. Amer.*, vol. 69, pp. 1340-1349, Oct. 1979.
12. N. Jayant, "Signal compression: Technology targets and research directions", *IEEE JSAC*, vol. 10, pp. 796-818, June 1992.

13. 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.
14. 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.
15. 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.
16. J. Hagenaur, "Rate-compatible punctured convolutional codes (RCPC codes) and their application ", *IEEE Trans. Comm.*, vol. 36, no.4, pp. 389-400, April 1988.
17. A. J. Viterbi, "Convolutional codes and their performance in communication systems", *IEEE Trans. Comm.*, vol. COM-19, no. 5, pp. 751- 772, Oct., 1971.
18. S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, July 1989.
19. K. Ramchandran, A. Ortega, K. Metin Uz, M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding", *IEEE JSAC*, vol. 11, No. 1, pp. 6-23, Jan. 1993.
20. H. Poor, G. Wornell, *Wireless Communications, Signal Processing Perspectives*, Upper Saddle River, NJ: Prentice-Hall, 1998.
21. M. Vetterli, J. Kovacevic, *Wavelet and Subband Coding*, Upper Saddle River, NJ: Prentice Hall, 1995.
22. T. Cover, "Broadcast channels", *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2-14, Jan. 1972.
23. J. Gu, Y. Jiang, J. Baras, "A Video Transmission System Based on the Human Visual Model for Satellite Channels", ISR Technical Report TR 99-14, University of Maryland, College Park, Mar. 1999.

## ABSTRACT

Title of Thesis: 3D WAVELET BASED VIDEO CODEC WITH  
HUMAN PERCEPTUAL MODEL

Degree Candidate: Junfeng Gu

Degree and Year: Master of Science, 1999

Thesis Directed by: Professor John S. Baras  
Institute of Systems Research

This thesis explores the utilization of a human perceptual model in video compression, channel coding, error concealment and subjective image quality measurement. The perceptual distortion model just-noticeable-distortion (JND) is investigated. A video encoding/decoding scheme based on 3D wavelet decomposition and the human perceptual model is implemented. It provides a prior compression quality control which is distinct from the conventional video coding system. JND is applied in quantizer design to improve the subjective quality of compressed video. The 3D wavelet decomposition helps to remove spatial and temporal redundancy and provides scalability of video quality. In order to conceal the errors that may occur under bad wireless channel conditions, a 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 proposed. A new subjective quality index based

on JND is proposed and used to evaluate the overall performance at different signal to noise ratios (SNR) and at different compression ratios.

Due to the wide use of arithmetic coding (AC) in data compression, we consider it as a readily available unit in the video codec system for broadcasting. A new scheme for conditional access (CA) sub-system is designed based on the cryptographic property of arithmetic coding. Its performance is analyzed along with its application in a multi-resolution video compression system. This scheme simplifies the conditional access sub-system and provides satisfactory system reliability.

60230-6646F03

3D WAVELET BASED VIDEO CODEC WITH  
HUMAN PERCEPTUAL MODEL

by

Junfeng Gu

Thesis submitted to the Faculty of the Graduate School of the  
University of Maryland at College Park in partial fulfillment  
Of the requirements for the degree of  
Master of Science  
1999

Advisory Committee:

Professor John S. Baras, Chair/Advisor  
Professor Rama Chellappa  
Professor Nariman Farvardin  
Professor Haralabos Papadopoulos

## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and thanks to my advisor Dr. John S. Baras for his guidance, patience, and support throughout my study and thesis research at University of Maryland at College Park for these two years.

I would thank Mr. Yimin Jiang and Mr. Li Liu. I got a lot from the discussion with them. Several times I came to the "no outlet" of research, their brilliant suggestions helped me find the right ways. I also owe gratitude to Professor Farvardin, Professor Chellappa, Professor Papadopoulos and other faculties at Department of Electrical and Computer Engineering, for their excellent lectures, instructions, helpful advice, and for serving on my committee.

I would like to give me special thanks to my wife. Her support and love accompany me all the time. Thanks for the understandings.

This work was supported by NASA contracts under the cooperative agreement NASA-NCC5227 and NASA-NCC3528, and by a contract from Texas Instruments.

## TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION.....	1
1.1 Backgrounds .....	1
1.2 Contribution of This Thesis .....	5
CHAPTER 2 THE HUMAN PERCEPTUAL MODEL OF JND .....	7
2.1 Human Perceptual Models and Perceptual Coding.....	7
2.1.1 Human Perceptual Models .....	7
2.1.2 Perceptual Coding .....	10
2.2 Just-noticeable-distortion (JND) Profile .....	11
2.3 A Novel Human Perceptual Distortion Measure Based on JND.....	17
CHAPTER 3 VIDEO CODING/DECODING SYSTEM.....	19
3.1 3-D Wavelet Analysis.....	21
3.2 Frame Counting & Motion Detection.....	26
3.3 JND Model Generation.....	27
3.4 Perceptually Tuned Quantization .....	27
3.4.1 Uniform Quantization.....	28
3.4.2 Lloyd-Max Quantizer .....	30
3.5 Arithmetic Coding and Slicing.....	32
3.6 Perceptual Channel Coding.....	33
3.7 Video Decoder and Error Concealment .....	34
CHAPTER 4 SIMULATION RESULTS .....	36
4.1 Spatio-temporal JND Profiles and JND Profiles for Subbands .....	37



4.2 Human Perceptual Distortion Measure vs. PSNR.....	39
4.3 Video Transmission over Satellite Channels .....	40
4.4 Comparison with MPEG.....	43
4.5 Quantization Schemes Comparison.....	49
CHAPTER 5 CONDITIONAL ACCESS WITH ARITHMETIC CODING.....	50
5.1 Introduction to Arithmetic Coding.....	51
5.2 Dependency of Arithmetic Coding.....	52
5.3 Conditional Access with Arithmetic Coding.....	55
5.4 Summary .....	60
CHAPTER 6 CONCLUSIONS AND OPEN ISSUES.....	62
6.1 Video Codec with JND.....	62
6.2 Conditional Access using Arithmetic Coding.....	63
6.3 Video Codec with Motion Estimation .....	64
6.4 Joint Source-Channel Coding Based on Human Visual Model.....	65
BIBLIOGRAPHY .....	68

Table 1 Average Distortion  $D_l$  for Each Subband.....34  
Table 2 Rate Index of RCPC.....40

## LIST OF FIGURES

Figure 1 Error visibility threshold in the spatio-temporal domain.....	15
Figure 2 Subbands after 3D Wavelet Decomposition.....	16
Figure 3 JND Based Video Encoder.....	20
Figure 4 JND Based Video Decoder.....	21
Figure 5 Heterogeneous 3D Wavelet Decomposition.....	24
Figure 6 Subbands after 3D Wavelet Decomposition.....	25
Figure 7 Frame #1 of "Calendar-Train".....	37
Figure 8 JND Subband Profile for Subband 0 to 6.....	38
Figure 9 Decoded Frame of "Claire", $\Delta_G=2.38$ , PSNR=30.80dB.....	39
Figure 10 Decoded Frame of "Claire", $\Delta_G=3.07$ , PSNR=30.15dB.....	39
Figure 11 Distortion of the Decoded Frames over Noisy Channel with Object Distortion Index=1.....	41
Figure 12 Distortion of the Decoded Frames over Noisy Channel with Object Distortion Index=5.....	42

Figure 13 Decoded Frame #3 of "Calendar-Train" with  $E_b/N_o = 3dB$  (4,7, 8)

Protection..... 43

Figure 14 Performance Comparison between the MPEG Coder (MPEG A) and the

JND Based Coder..... 44

Figure 15 "Claire" from the JND based encoder,  $\Delta_G=0.85$ , PSNR=36.3dB,

compression ratio=27.0:1..... 45

Figure 16 "Claire" I frame from the MPEG-1 coder,  $\Delta_G=1.3$ , PSNR=37.5dB,

compression ratio=26.7:1..... 45

Figure 17 "Claire" from the JND based encoder, PSNR=34.5dB, CR is 35.0:1..... 47

Figure 18 "Claire" from the MPEG-1 encoder, PSNR=37.6dB, CR is 35.5:1..... 47

Figure 19 "Calendar-Train" from the JND based encoder,  $\Delta_G=3.03$ ,

PSNR=32.6dB, compression ratio 8.93:1..... 48

Figure 20 "Calendar\_Train" from the MPEG-1 encoder (quantization scale=7),

$\Delta_G=3.11$ , PSNR=34.0dB, compression ratio 9.06:1..... 48

Figure 21 Performance Comparison between Uniform Quantization and

Mixed Optimum Quantization..... 49

Figure 22 Arithmetic Coding based Conditional Access Sub-system.....	56
Figure 23 Combined Video Coding and CA System.....	57
Figure 24 Slicing of Subband.....	58

CONFIDENTIAL

## Chapter 1 Introduction

### 1.1 Backgrounds

The scale and power of computing and communication systems is keeping on increasing. Meanwhile, the data required to represent the image and video signal in digital form would continue to overwhelm the capacity of many communication and storage systems. In particular, the growth of data-intensive digital video and image applications and the increasing use of bandwidth-limited media such as radio and satellite links have not only sustained the need for more efficient ways to encode analog signals, but have made signal compression central to digital communication and signal-storage technology. Furthermore, for some applications such as image/video database browsing and multipoint video distribution over heterogeneous networks and display devices, there is a growing need for other useful features such as video scalability. Highly scalable video compression schemes [5][14][15] allow selective transmission of different sub-bitstreams to different destinations, depending on their respective needs. In this manner, each receiver can have the best possible quality session according to its bandwidth.

The ultimate object of a video/image compression system is to minimize the average number of bits used to represent the digital video/image signal while maintaining subjective video/image quality as good as possible. Since the traditional measures of image signal quality, mean square error (MSE), and signal-to-noise ratio (PSNR), do not provide a satisfactory reflection of the human's subjective perception on the video/image quality, more satisfactory metrics should take advantage of the

human perception model implicitly or explicitly. They give the encoder certain fidelity to allocate more bits to signals that are more meaningful to the human visual system. And they lead to better quantitative evaluation methods. A variety of schemes have been proposed to incorporate certain psychovisual properties of the human visual system (HVS) into image/video coding algorithms [16][17][18][20][42]. The frequency sensitivity, brightness sensitivity, texture sensitivity and color sensitivity are considered for the distortion sensitivity profiles, which leads to modern human visual models, such as just-noticeable-distortion (JND) [16], visible differences predictor (VDP) [19] and Ran's perceptually motivated three-component image model [20]. Jayant's JND model provides each signal being coded with a threshold level 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.

To approach the goal of high compression ratio and scalability, subband coding is widely explored in recent years. The signal is decomposed into frequency subbands and these subbands are encoded independently or dependently. The structures in the high frequency subbands usually appear as sparse edges and impulses corresponding to the localized discontinuities in spatial or temporal domains. After carefully designed quantization, these subbands lead to a large quantity of compression, provide spatial and bitstream scalability naturally, and require less error protection in channel coding. Earlier work on subband coding for image compression applies to the still image [1][2]. Tanabe and Farvardin [8] suggest a scheme using entropy-coded quantization to get the optimal quantizer performance. Kwon and Chellappa [9] use adaptive entropy-

constrained quantizers for different regions in images. The early work on subband coding of video includes [3] and [4]. In recent years, the subband decomposition has been extended to three dimensions. The codec of Taubman and Zakhor [5], which employs a global motion compensation scheme accounting for camera panning motion, generates a single embedded bit stream supporting a wide range of bit rates. Podichuk, Jayant and Farvardin [6] combine a 3-D subband coder with geometric vector quantization and obtain good compression performance at low bit rates. There are novel wavelet-based video coding systems that take advantage of good features of other components such as overlapped motion estimation. The video codec developed by David Sarnoff Research Center [10] uses the technology of overlapped block motion compensation and zero-tree entropy coding (ZTE), which is the extension of Shapiro's EZW [11] and Said & Pearlman's SPIHT [12]. It outperforms the VM of MPEG-4 and provides scalability. Cinkler [13] uses an edge-sensitive subband coding (ESSBC) method and overlapped motion compensation. From an edge map combined with motion vectors, the ESSBC technology generates areas of significance. These areas are processed by a modified wavelet transform to concentrate the energy.

Even if a very high percentage of total signal energy is contained in the lowest frequency subband, the truncation or undercoding of high-band signals will result in the perception of distortion due to aliasing effects. On the other hand, unless the significant signals are cautiously encoded, the overcoding of high-band signals is the price to pay for gaining good image quality. Consequently, the problem to be solved for optimizing the subband coding scheme is how to locate perceptually important signals in each



frequency subband, and how to encode these signals with the lowest possible bit rate without exceeding the error visibility threshold.

As a critical (and often controlling) technology in the video broadcasting industry, a conditional access sub-system comprises a combination of scrambling and encryption to prevent unauthorized reception. Encryption is the process of protecting the secret keys that are transmitted with a scrambled signal to enable the descrambler to work. Way back in 1988, an attempt was made by France Telecom and others to develop a standard encryption system for Europe. The result was Eurocrypt.

Unfortunately, in its early manifestations it was not particularly secure and multiplex operators went their own way. Thus, in 1992 when the DVB started their consideration of CA systems, they recognized that the time had passed when a single standard could realistically be agreed upon and settled for the still difficult task of seeking a common framework within which different systems could exist and compete. They therefore defined an interface structure, the Common Interface, which would allow the set top box (STB) to receive signals from several service providers operating different CA systems. The common interface module contains the CA system, rather than the STB itself, if necessary allowing multiple modules to be plugged into a single STB.

However, there were serious objections to the common interface module from many CA suppliers on the grounds that the extra cost would be unacceptable. As a result, the DVB stopped short of mandating the Common Interface, instead recommending it, along with simulcrypt, which is one of the DVB recommended approaches for conditional access. These all bring up a diversified market of conditional access system, which makes the exploration in this field so valuable.

## 1.2 Contribution of This Thesis

We have implemented a video encoding/decoding scheme based on 3-D wavelet decomposition and a human perceptual model. Jayant's just-noticeable-distortion (JND) model is adopted. The quantizers in different subbands are designed to approach perceptual optimum. The source encoder has the global control on subjective distortion of the compressed video quality ahead of time, which is distinct from the conventional compression schemes. A new subjective distortion index for video is proposed and used to evaluate the overall performance. Its fidelity is compared with the traditional quality metric PSNR. From the result of simulation, we conclude that our distortion index based on JND profile is more accurate than PSNR in the sense of measuring the human subjective distortion. Using the new distortion index, the performance of our video codec is compared with the coding of I frames in MPEG. At the same bit rate, our encoder has performance comparable with MPEG encoder for I frames. Our simulation shows that our encoder assigns more error energy to the perceptually less important pixels in the frames. But due to the lack of motion estimation and run-length coding technologies, the overall compression performance of our encoder is worse than MPEG. To present its application more concretely, a practical transmission system over a satellite channel using unequal error protection is discussed. Since in the satellite broadcasting case, a feedback channel is not available, 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 to confine the propagation of bit

errors. Rate compatible punctured convolutional (RCPC) codes [27] are adopted in our system to provide unequal error protection for different subbands. The bit rates of RCPC for these subbands are finely chosen following the JND model to make the unequal error protection perceptually sub-optimal. Simulations are done for different combinations of RCPC coding and channel SNR, showing some characteristics of our coding scheme.

Following the rapid expansion of the commercial broadcasting industry, a conditional access sub-system is always included in the broadcasting system. It is used to control which customer can get particular program services. Particular programs are only accessible to customers who have satisfied the required payments. In this paper, a brand new conditional access sub-system which takes advantage of the cryptographic property of arithmetic coding is suggested. And a video broadcasting system based on subband coding is described to present the application of this new condition access sub-system. The performance analysis is provided. Compared to the traditional structures, our scheme is quite simple and of low cost while provides reliable security.

## Chapter 2 The Human Perceptual Model of JND

### 2.1 Human Perceptual Models and Perceptual Coding

#### 2.1.1 Human Perceptual Models

A common model of vision incorporates a low-pass filter, a logarithmic nonlinear transformer, and a multi-channel signal-sharpening high-pass filter [26]. A biologically correct and complete model of the human perceptual system would incorporate descriptions of several physical phenomena including peripheral as well as higher level effects, feedback from higher to lower levels in perception, interactions between audio and visual channels, as well as elaborate descriptions of time-frequency processing and nonlinear behavior. Some of the above effects are reflected in existing coder algorithms, either by design or by accident. For example, certain forms of adaptive quantization and prediction provide efficient performance in spite of inadequate response time because of temporal noise masking. The basic time-frequency analyzers in the human perceptual chain are described as bandpass filters. Bandpass filters in perception are sometimes reflected in coder design and telecommunication practice in the forms of "rules of thumb".

A particularly interesting aspect of the signal processing model of the human system is non-uniform frequency processing. The critical bands in vision are non-uniform. It is necessary to use masking models with a non-uniform frequency support to incorporate this in coder design. Here masking refers to the ability of one signal to hinder the perception of another within a certain time or frequency range. It is also necessary to recognize that high-frequency signals in visual information tend to have a

short time or space support, while low-frequency signals tend to last longer. An efficient perceptual coder therefore needs to not only exploit properties of distortion masking in time and frequency, but also have a time-frequency analysis module that is sufficiently flexible to incorporate the complex phenomena of distortion masking by non-stationary input signals. All of this is in contrast to the classical redundancy-removing coder, driven purely by considerations of minimum mean square error (MMSE), MMSE bit allocation, or MMSE noise shaping matched to the input spectrum.

Distortion sensitivity profiles of human perception are driven as functions of frequency, brightness, texture, and temporal parameters. These four kinds of sensitivity are under consideration for gray scale video/image [16][18][26]:

(1) Brightness sensitivity:

It was found that human visual perception is sensitive to luminance contrast rather than absolute luminance values. The ability of human eyes to detect the magnitude difference between an object and its background is dependent on the average value of background luminance. According to Web's Law [28], if the luminance of a test stimulus is just noticeable from the surrounding luminance, the ratio of just noticeable luminance difference to stimulus's luminance (Weber fraction) is almost constant. However, due to the ambient illumination falling on the display, the noise in dark areas tends to be less perceptible than that occurring in regions of high luminance. In general, high visibility thresholds will occur in either very dark or very bright regions, and lower thresholds will occur in regions of gray levels close to the mid-gray luminance, which is 127 for 8 bit sampling.

## (2) Texture sensitivity:

The reduction in the visibility of stimuli due to the increase in spatial nonuniformity of the background luminance is known as texture masking. Several efforts have been made to utilize some forms of texture making to improve coding efficiency. In many approaches, visibility thresholds are defined as functions of the amplitude of luminance edge in which perturbation is increased until it becomes just discernible. The visibility threshold in this approach is associated with the masking function defined at each pixel as the maximum prediction error from the four neighboring pixels.

## (3) Temporal sensitivity:

The masking of temporally changing stimuli is extremely important in interframe coding. However, temporal masking is complicated by many factors, and its application to video coding is still in its infancy. Many researches have attempted to evaluate the losses of spatial resolution and magnitude resolution as an object moves in a scene. If movement is drastic, such as scene change, the perceived spatial and intensity resolution is significantly reduced immediately after the scene change. It was found that the eye is noticeably more sensitive to flicker at high luminance than at low luminance [28].

## (4) Frequency sensitivity:

Many psychovisual studies have shown that the human perception of distortion depends on its frequency distribution. The response of the HVS to sinewave gratings of different frequencies has been experimentally measured as the so-called contrast sensitivity function (CSF). Many models of spatial-domain CSF have been proposed,

which indicate general bandpass characteristics. The spatial-domain CSF has been widely used to improve the quality of the coded still images. There are only a few models of spatio-temporal CSF reported in the literature, among which the most well known model is proposed by Kelly [22]. The spatio-temporal CSF provides relative sensitivities of the HVS to different spatio-temporal frequencies, or relative tolerance of noises at different spatio-temporal frequencies. It can be used to allocate coding bits, or distortion, by adjusting the quantizer stepsize of the target signal as inversely proportional to the sensitivity of the corresponding frequency.

### 2.1.2 Perceptual Coding

There are two intrinsic operations to signal coding: removal of redundancy and reduction of irrelevancy [16]. The removal of redundancy is the effect of predictive coding or transform coding. Almost all sampled signals in coding are redundant because Nyquist sampling typically tends to preserve some degree of inter-sample correlation. This is reflected in the form of a nonflat power spectrum. Greater degrees of nonflatness, as resulting from a low-pass function for signal energy versus frequency, or from periodicities, lead to greater gains in redundancy removal. These gains are also referred to as prediction gains or transform coding gains, depending on whether the redundancy is processed in the spatial domain or in the frequency (or transform) domain.

The reduction of irrelevancy is the result of amplitude quantization. In a signal compression algorithm, the inputs of the quantizing system are typically sequences of prediction errors or transform coefficients. The idea is to quantize the prediction error, or the transform coefficients, just finely enough to render the resulting distortion

imperceptible, although not mathematically zero. If the available bit rate is not sufficient to realize this kind of perceptual transparency, the intent is to minimize the perceptibility of the distortion by shaping it advantageously in space or frequency, so that as many of its components as possible are masked by the input signal itself. The term perceptual coding is used to signify the matching of the quantizer to the human visual system, with the goal of either minimizing perceived distortion, or driving it to zero where possible. These goals do not correspond to the maximization of signal-to-noise ratio or the minimization of mean square error.

## 2.2 Just-noticeable-distortion (JND) Profile

To remove the redundancy due to spatial and temporal correlation and the irrelevancy of perceptually insignificant components from video signals, the concept of just-noticeable distortion profile introduced by Jayant [21] has been successfully applied to perceptual coding of video and image. JND provides each signal to be coded with a visibility threshold of distortion, below which reconstruction errors are rendered imperceptible. The JND profile of a still image is a function of local signal properties, such as the background luminance and the activity of luminance changes in the spatial domain. For video sequences, the derivation of JND profiles must take both spatial and temporal masking effects into consideration. For a successful estimation of JND profiles, the subject should not be able to discern the difference between a video sequence and its JND-contaminated version. Once JND profiles of video signals are obtained, the perceptual redundancy can be quantitatively measured, and the perceptual significance of each target signal can be evaluated.



Whenever transparent coding cannot be attained due to a tight bit-rate budget, the minimally-noticeable-distortion (MND) profile rather than the JND profile is required. In MND the increased distortion can be uniformly distributed over the reconstructed video signals and thus there is minimally perceptible. The perceptual quality of the reconstructed video signals is expected to degrade gracefully if the available bit rate is reduced.

The generation of a JND model consists of several steps [18]. First, the perceptual redundancy inherent in the spatial domain is quantitatively measured as a 2D profile by a perceptual model that incorporates the visibility thresholds due to average background luminance and texture masking [17]. It is described by the following expression [17]:

$$JND_s(x, y) = \max\{f_1(bg(x, y), mg(x, y)), f_2(bg(x, y))\}, \quad (1)$$

$$\text{for } 0 \leq x < W, 0 \leq y < H$$

where  $f_1$  represents the error visibility threshold due to texture masking,  $f_2$  the visibility threshold due to average background luminance;  $H$  and  $W$  denote respectively the height and width of the image;  $mg(x, y)$  denotes the maximal weighted average of luminance gradients around the pixel at  $(x, y)$ ;  $bg(x, y)$  is the average background luminance.

$$f_1(bg(x, y), mg(x, y)) = mg(x, y)\alpha(bg(x, y)) + \beta(bg(x, y)) \quad (2)$$

$$f_2(bg(x, y)) = \begin{cases} T_0 \cdot (1 - (bg(x, y)/127)^{1/2}) + 3 & \text{for } bg(x, y) \leq 127 \\ \gamma \cdot (bg(x, y) - 127) + 3 & \text{for } bg(x, y) > 127 \end{cases} \quad (3)$$

$$\alpha(bg(x, y)) = bg(x, y) \cdot 0.0001 + 0.115$$

$$\beta(bg(x, y)) = \lambda - bg(x, y) \cdot 0.01 \quad \text{for } 0 \leq x < H \quad 0 \leq y < W$$

where  $T_0$ ,  $\gamma$  and  $\lambda$  are found to be 17,  $3/128$  and  $1/2$  through experiments [17].

The value of  $mg(x, y)$  across the pixel at  $(x, y)$  is determined by calculating the weighted average of luminance changes around the pixel in four directions. Four operators,  $G_k(i, j)$ , for  $k=1, \dots, 4$ , and  $i, j=1, \dots, 5$ , are employed to perform the calculation, where the weighting coefficient decreases as the distance away from the central pixel increases.

$$mg(x, y) = \max_{k=1,2,3,4} \{ |grad_k(x, y)| \} \quad (4)$$

$$grad_k(x, y) = \frac{1}{16} \sum_{i=1}^5 \sum_{j=1}^5 p(x-3+i, y-3+j) \cdot G_k(i, j)$$

for  $0 \leq x < H, \quad 0 \leq y < W,$

where  $p(x, y)$  denotes the pixel at  $(x, y)$ . The operators  $G_k(i, j)$  are

$$G_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 0 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & -8 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix},$$

$$G_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 8 & 0 \\ -1 & -3 & 0 & 3 & 1 \\ 0 & -8 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}, \quad G_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 8 & 0 & -8 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

The average background luminance,  $bg(x,y)$ , is calculated by a weighted lowpass operator,  $B(i,j)$ ,  $i, j=1, \dots, 5$ .

$$bg(x,y) = \frac{1}{32} \sum_{i=1}^5 \sum_{j=1}^5 p(x-3+i, y-3+j) \cdot B(i,j), \quad (5)$$

$$\text{where } B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

At the second step of JND model generation, the JND profile representing the error visibility threshold in the spatio-temporal domain is expressed as

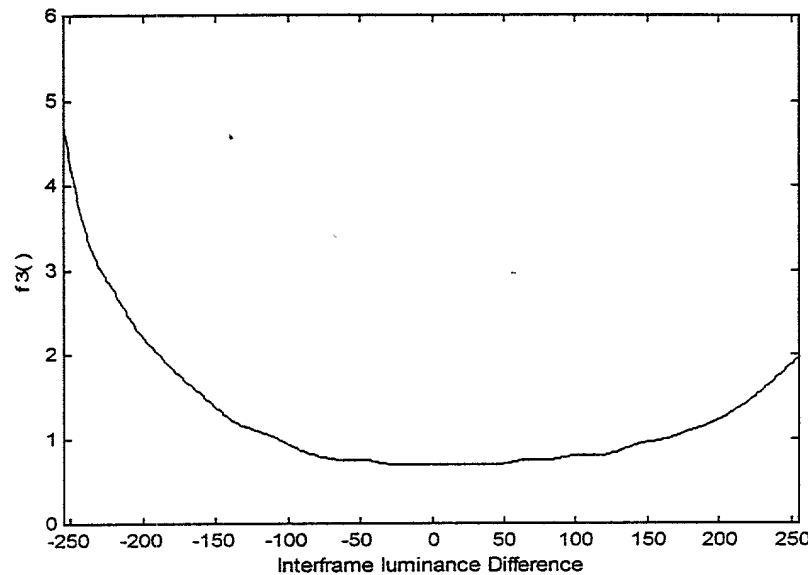
$$JND_{S-T}(x,y,n) = f_3(ild(x,y,n)) \cdot JND_S(x,y,n) \quad (6)$$

where  $ild(x,y,n)$  denotes the average interframe luminance difference between the  $n$ th and  $(n-1)$ th frame. Thus, to calculate the spatio-temporal JND profile for each frame in a video sequence, the spatio JND profile of itself and its previous reference frame are required.

$$ild(x,y,n) = 0.5 \cdot [p(x,y,n) - p(x,y,n-1) + bg(x,y,n) - bg(x,y,n-1)]$$

$f_3$  represents the error visibility threshold due to motion. The empirical results of measuring  $f_3$  for all possible interframe luminance differences are shown in **Figure 1**. To purposely minimize the allowable distortion in the nonmoving area, the scale factor is switched to 0.8 as  $|ild(x,y,n)| < 5$ . It can be noted that the error visibility threshold is increased with the increasing interframe luminance difference. This conforms the research findings, that after a rapid scene change or large temporal difference, the

sensitivity of the HVS to spatial details is decreased. Moreover, it can be found that temporal masking due to high-to-low luminance changes is more prominent than that due to low-to-high luminance changes.



**Figure 1 Error visibility threshold in the spatio-temporal domain**

From the spatio-temporal JND profile that quantitatively measures the perceptual redundancy inherent in full-band video signals, the JND profiles for each subband are set up with certain distortion allocation [22] for different subbands. The JND for subband  $q$  is a function of spatio-temporal JND values at corresponding locations multiplied by a weight that indicates the perceptual importance of this subband. When each pair of video frames is decomposed into 11 spatio-temporal subbands as in **Figure 2** (the details of such a 3D wavelet decomposition is described in Chapter 3), the relationship between the full-band JND profile and the component JND profiles can then be obtained by the following equations:

$$JND_q(x, y) = \left[ \left[ \sum_{i=0}^3 \sum_{j=0}^3 \sum_{t=0}^1 JND_{S-T}^2(i+x \cdot 4, j+y \cdot 4, t) \right] \cdot \omega_q \right]^{1/2}, \quad (7)$$

for  $0 \leq q \leq 3, \quad 0 \leq x < W/4, \quad 0 \leq y < H/4$

and

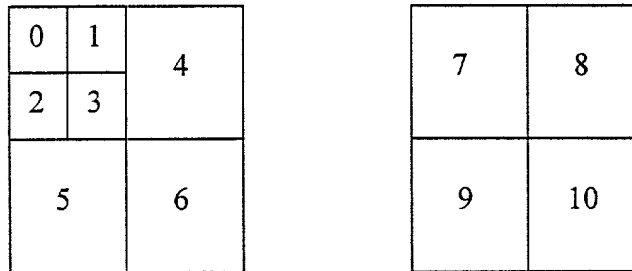
$$JND_q(x, y) = \left[ \left[ \sum_{i=0}^2 \sum_{j=0}^1 \sum_{t=0}^1 JND_{S-T}^2(i+x \cdot 2, j+y \cdot 2, t) \right] \cdot \omega_q \right]^{1/2} \quad (8)$$

for  $4 \leq q \leq 10$ ,  $0 \leq x < W/2$ ,  $0 \leq y < H/2$

The weighting function for distributing the full-band JND energy to a subband can be derived as the relative sensitivity of the HVS to the frequency subband. For 11 spatio-temporal subbands, the weighting function of the  $q$ th subband is obtained as

$$\omega_q = \frac{S_q^{-1}}{\sum_{i=0}^{10} S_i^{-1}}, \quad \text{for } 0 \leq q \leq 10 \quad (9)$$

where  $S_q$  represents the average sensitivity of the HVS to the  $q$ th subband.  $S_q$  is obtained from the spatio-temporal CSF presented by Kelly [22].



**Figure 2 Subbands after 3D Wavelet Decomposition**

### 2.3 A Novel Human Perceptual Distortion Measure Based on JND

Based on the basic concept of the JND, the idea of minimally-noticeable-distortion (MND) is developed for situations where the 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[26]. The perceptual quality of the reconstructed image is accordingly expected to degrade evenly if bit-rate is reduced. MND is expressed as:

$$MND(x, y) \equiv JND(x, y) \cdot d(x, y) \quad (10)$$

where  $0 \leq x < W$ ,  $0 \leq y < H$ ,  $W$  and  $H$  are the width and height of an image respectively,  $d(x, y)$  is the distortion index at point  $(x, y)$ .

The energy of JND can be understood as the minimum energy of quantization error that will cause conceivable loss to the human visual system. So if the error is so small in a small area that the error energy here is less than the JND energy, the compression will be perceptually lossless. We define the energy of the MND of a small area indexed by  $(i, j)$  as:

$$\sum_{(x, y) \in r_{ij}} MND^2(x, y) \equiv \sum_{(x, y) \in r_{ij}} JND^2(x, y) \cdot \delta(i, j) \quad (11)$$

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:

$$\Delta_G \equiv \frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L \varepsilon(k, l) \quad (12)$$

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:

$$\varepsilon(k,l) \equiv \text{median} \left( \delta(i,j) \mid r_{ij}(k,l) \in R_{kl}, 1 \leq i \leq M, 1 \leq j \leq N \right) \quad (13)$$

The larger  $\Delta_G$  is, the larger the subjective perceptual distortion is. Compared with PSNR or MSE,  $\Delta_G$  has the same convenience to describe the picture quality with one quantitative value. However,  $\Delta_G$  takes the human visual model into account, therefore it can reflect subjective visual quality better than PSNR or MSE. It is well accepted that the value of MSE or PSNR is meaningless to video/image subjective evaluation and when two images are compressed, the comparison of their MSE or PSNR values cannot give a creditable psychovisual conclusion. On the other hand, the distortion  $\Delta_G$  can be explained as "this image is compressed at a scale of  $\Delta_G$  times the perceptually noticeable distortion". Generally speaking, if one image is coded with  $\Delta_G$  larger than another one, the former's subjective quality is higher. Due to these considerations, we will use  $\Delta_G$  as our index of performance for our video compression system. PSNR will be calculated at the same time as reference.

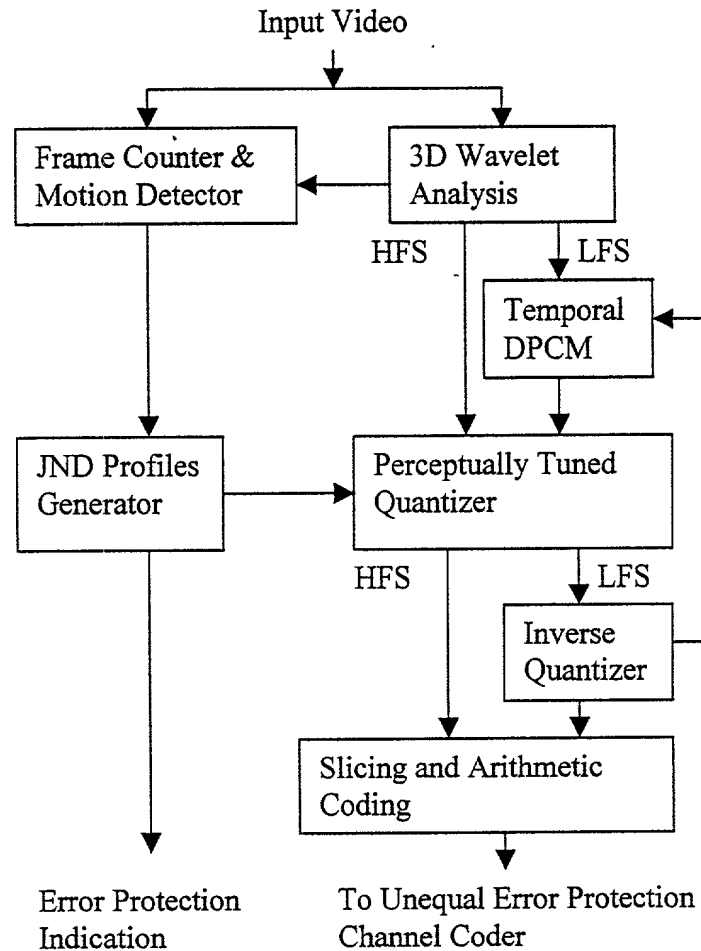
## Chapter 3 Video Coding/Decoding System

Figure 3 and Figure 4 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 the 3-D wavelet analysis module. The Frame Counter & Motion Detector controls the renewing of the JND profiles from frame counter and drastic movement 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 Perceptually Tuned Quantizer implements quantization for the wavelet coefficients in each subbands 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 signals in 11 bit streams. These bit streams are fed into the Unequal Error Protection Channel Coder and an error protection indication from JND Model Generator is also given to the channel coder. These modules will be discussed subsequently.

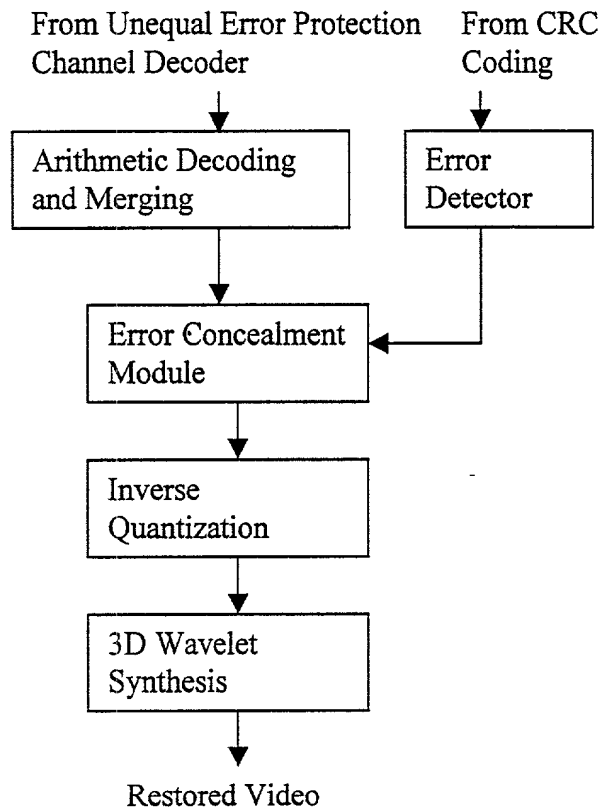
In the decoder, there are Arithmetic Decoding and Merging, Error Detection, Error Concealment, Inverse Quantization and 3D Wavelet Synthesis modules. The Arithmetic Decoding and Merging part decodes the bit streams back to slices of wavelet coefficients, and gives them appropriate arrangements in the corresponding subbands. CRC coding alarms the Error Detector for bit errors. The latter's output information about the location of error helps the Error Concealment Module to diminish the effects of bad data from fault-causing noisy channel. Inverse Quantization and 3D Wavelet



Synthesis parts implement the inverse operation of the quantization and 3D wavelet decomposition in the video encoder. Finally the video signal is restored.



**Figure 3 JND Based Video Encoder**



**Figure 4 JND Based Video Decoder**

### 3.1 3-D Wavelet Analysis

Wavelet multiresolution analysis techniques have been applied primarily to 1D and 2D signals. These techniques project the signal onto a chain of embedded approximation and detail spaces designed to represent the signal and its details at various levels of resolution [29]. For practical purposes, the projection coefficients are obtained using a discrete subband transform that employs a quadrature mirror filter pair related to the type of wavelet used in the analysis. In conventional 2D wavelet multiresolution analysis, the separable 2D approximation spaces are formed from the tensor product of identical 1D approximation spaces [30]. This restriction generates analyzing filters with homogeneous spectral characteristics in 2D frequency space.

When extended to three dimensions, the multiresolution analysis is constructed from a separable 3D analyzing, or “scaling”, function formed from the product of three nonidentical 1D scaling functions, or two identical 1D spatial scaling functions and one different 1D temporal scaling function. This brings a much richer set of orthonormal basis vectors with which to represent 3D signals, and it produces filters that can be easily tailored to more closely match the spatial and temporal frequency characteristics of the 3D signal.

An  $L_2(\mathbb{R})$  multiresolution analysis consists of a chain of closed, linear “approximation” spaces  $V_j$  and a scaling function  $\phi$  which satisfy the following properties for all  $f \in L_2(\mathbb{R})$ .

$$1) \dots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots$$

$$2) \overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}); \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\}$$

$$3) \begin{aligned} f(x) \in V_j &\Leftrightarrow f(2x) \in V_{j+1}; \quad j \in \mathbb{Z} \\ f(x) \in V_j &\Rightarrow f(x + n/2^j) \in V_j; \quad n \in \mathbb{Z} \end{aligned}$$

$$4) \text{ The set of functions } \{2^{j/2} \phi(2^j x - n) | j \in \mathbb{Z}, n \in \mathbb{Z}\} \text{ forms an orthonormal basis for the approximation space } V_j.$$

As presented by Mallat, the purpose of multiresolution analysis is to create a mathematical framework that facilitates the construction of a wavelet orthonormal basis for the space for all finite energy signals  $L_2(\mathbb{R})$ . To this end, denote the orthogonal complement of  $V_j$  in  $V_{j+1}$  by  $W_j$  where

$$V_{j+1} = V_j \oplus W_j$$

and the symbol  $\oplus$  indicates the direct sum.  $W_j$  is typically referred to as the  $j$ th detail space, because it captures the difference in signal information between the approximation spaces  $V_{j+1}$  and  $V_j$ .

Mallat has shown that one can create a mother wavelet  $\psi(x)$  such that the set of functions  $\{2^{j/2} \psi(2^j x - n) | n \in Z\}$  forms an orthonormal basis for  $W_j$ . The spaces  $W_j$ , where  $j \in Z$ , are mutually orthogonal; thus, by the denseness property of the multiresolution analysis, the set of scaled and dilated wavelets

$\{2^{j/2} \psi(2^j x - n) | j \in Z, n \in Z\}$  forms an orthonormal basis for  $L_2(\mathbb{R})$ . The scaling functions and the mother wavelet are related by the “two-scale” recursion relations

$$\begin{aligned}\phi(x) &= \sum_{n=-\infty}^{\infty} h_n \sqrt{2} \phi(2x - n) \\ \psi(x) &= \sum_{n=-\infty}^{\infty} g_n \sqrt{2} \phi(2x - n)\end{aligned}\tag{14}$$

where  $h_n$  and  $g_n$  are the coefficients of the QMF pair which is used to compute the approximation and detail projection associated with  $V_j$  and  $W_j$  from the approximation at the next higher scale  $V_{j+1}$ .

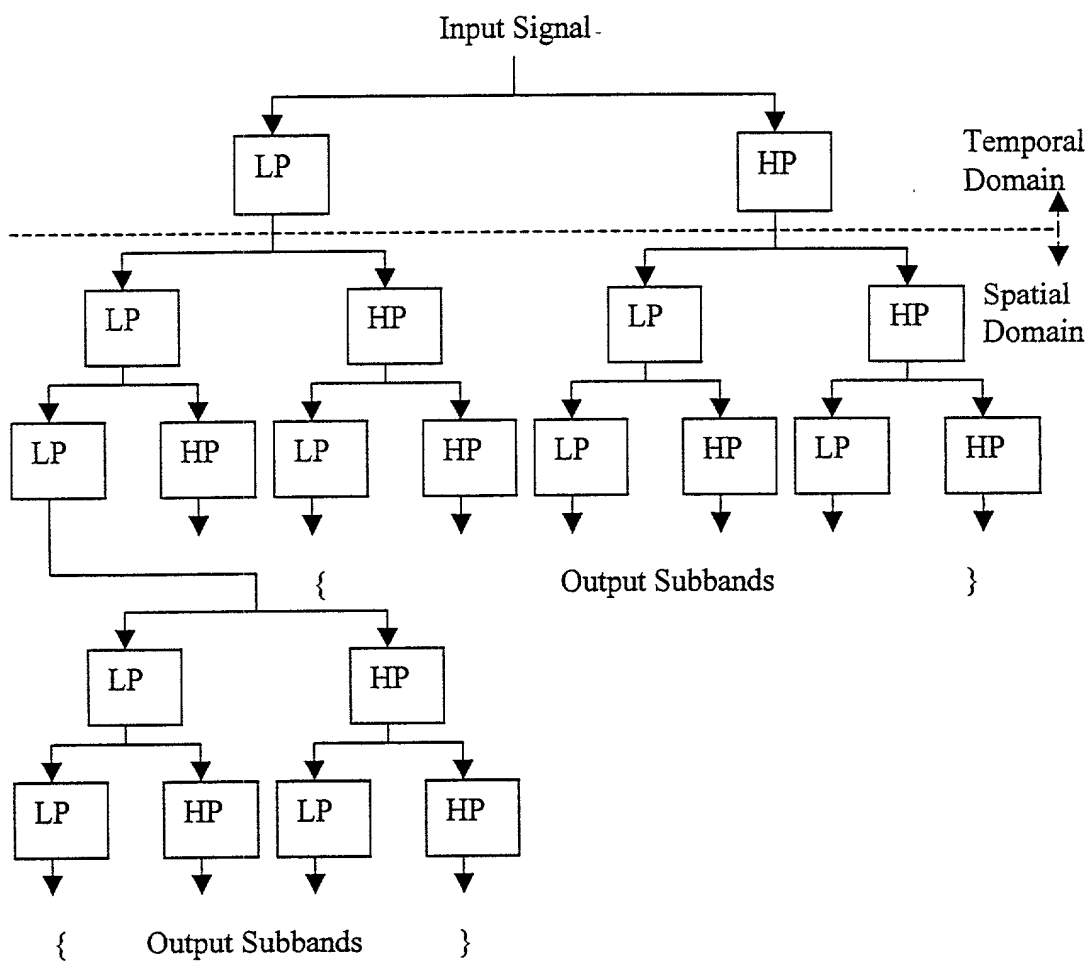
Approximation and detail signals are created by orthogonally projecting the input signal  $f$  onto the appropriate approximation or detail space. Since each space is spanned by an orthonormal basis set, the signal projection onto a given approximation or detail space at the  $j$ th resolution, is equivalent to the sequence of projection coefficients obtained by the inner product operations

$$a_{j,n} = \int_{-\infty}^{\infty} f(x) 2^{j/2} \phi(2^j - n) dx$$

$$d_{j,n} = \int_{-\infty}^{\infty} f(x) 2^{j/2} \psi(2^j - n) dx$$
(15)

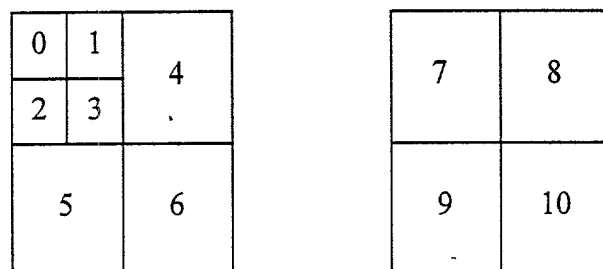
where  $a_{j,n}$  and  $d_{j,n}$  represent the  $j$ th approximation and detail coefficients respectively.

**Figure 5** shows the block diagram of the heterogeneous 3D wavelet decomposition.



**Figure 5 Heterogeneous 3D Wavelet Decomposition**

In our video codec system, the wavelet transform performs decomposition of video frames into a multi-resolution subband representation. Each pair of incoming video frames is decomposed into 11 subbands as shown in **Figure 6**.



**Figure 6 Subbands after 3D Wavelet Decomposition**

We use the simple two-tap Haar filter, which essentially separates the signal into temporal low frequency and high frequency parts. Then, the low frequency part is decomposed for two levels with the spatial 2-D wavelet decomposition, while the high frequency part is decomposed for one level. The Antonini (7,9)-filter [21] is used here. The Haar and Antonini filters work together to achieve a satisfying balance between complexity and performance.

The coefficients for the Antonini (7,9) wavelet filter are:

For the LPF, the  $h_n$  are given by {0.0378284555, 0.0238494650, 0.1106244044, 0.3774028556, 0.852698679, 0.3774028556, 0.1106244044, 0.0238494650, 0.0378284555}.

For the HPF, the  $g_n$  are given by {0.0645388826, 0.0406894176, 0.4180922732, 0.788485616, 0.4180922732, 0.0406894176, 0.0645388826}.

### 3.2 Frame Counting & Motion Detection

The computation of the JND model is resource consuming for the real-time video system. Even if it is implemented with dedicated hardware, the reconstruction of the JND profile for each pair of incoming video frames is expensive and unnecessary. The Frame Counter & Motion Detector is designed to control the renew process of the JND. The frame counter is used to count the number of frames being coded. After a certain number of frames (typically 10 or 20 frames) have been coded with the current JND model, this model is refreshed with the updated signal. Our assumption is that the scene in frames remains almost the same so the original JND model becomes effective before the update.

If, however, drastic movement or scene cut happens, the scene changes greatly. So the original JND has to be refreshed right away to follow the change. This is why a motion detector is necessary.

In our system, a simple motion detection scheme is adopted at the current time and stage of development. It considers two factors relative to picture contents change, which are scene cut and drastic movement. Initially, after the 3D wavelet decomposition, the energy in the spatial-LLLL temporal-L subband (i.e. subband 0 in **Figure 6**) is calculated for each group of two frames and stored as  $E_{old}$ . It is then compared with the new generated energy  $E_{new}$  for subband 0. If their difference exceeds the threshold, we assume a scene cut has occurred. That threshold can be obtained from experiments. As for detecting drastic movement, the energy in the spatial-LL temporal-H subband (i.e. subband 7 in **Figure 6**) is also calculated after the wavelet

decomposition. If it exceeds a given threshold, a drastic motion is assumed to be happening.

With each detection event from the Frame Counter & Motion Detector, the JND model is reconstructed.

### 3.3 JND Model Generation

The JND provides each signal a threshold of visible distortion, below which reconstruction errors are rendered imperceptible. In this part, the spatial-temporal JND profile for each group of two frames in a video sequence and the JND profiles for subbands are generated sequentially. The details are as described in Chapter 2.

Since the generation of the spatial-temporal JND profile for each frame requires one previous frame as reference, when a video sequence is being coded, the encoder has to assign the first frame a reference frame. In our implementation, the first frame in the video sequence will use itself as reference to generate each renewed JND profile.

### 3.4 Perceptually Tuned Quantization

The perceptually tuned quantizer is the core of the source encoder. With the JND profiles for each subband at hand, the most important task is to allocate the available bits to certain coefficients obtained from the wavelet decomposition as efficiently as possible.

The beauty of JND is that it provides a quantitative measure of the error sensitivity threshold with spatial and frequency localization. In schemes using DCT like



MPEG-2, the DCT coefficients are quantized using a quantization table that assigns more bits to the more important lower frequency coefficients in one 8 by 8 block. Although such a table is designed according to the HVS response based on psychovisual theory and experiments, its wide use all over the whole picture brings shortcomings, because different parts in a picture can have different visual importance at different scenes. On the other hand of the quantization table based on JND, which can adapt itself to local scenes results in less perceptual distortion.

At the beginning of our video coding, a parameter is requested for by the encoder. It is the object global distortion index  $\Delta_G$  that will control the overall video quality during the coding procedure. The same symbol  $\Delta_G$  is used here as the symbol we used in Chapter 2 to indicate our new human perceptual distortion measure based on JND, since these two have the same psychological meaning. Ideally, the coding procedure controlled by the object global distortion index  $\Delta_G = d$  will produce a compressed video sequence whose average perceptual distortion index  $\Delta_G$  value is  $d' = f(d)$ . This powerful functionality to control the compressed video quality ahead of time makes our scheme so distinct from conventional video coding systems. In addition the coding distortion is perceptually evenly distributed across the whole image, as our scheme will show.

### 3.4.1 Uniform Quantization

A mid-rising uniform quantizer is used as our basic quantizer due to its simplicity of error analysis and its sound performance under certain conditions for

optimality [24]. First, a global distortion index  $\Delta_G$  is given for the quantization procedure. It usually ranges from 1.0 to 10.0, where 1.0 stands for just noticeable distortion. Second, each subband is partitioned into non-overlapped blocks  $(r_{ij}(k,l))$ . These blocks are set up with the size of 8x8 or 16x16. For each block  $r_{ij}(k,l)$ , the step size of the mid-rising uniform quantizer is maximized under the condition that quantization error energy is less than or equal to the MND energy in this block that has the distortion index  $\delta(i,j)$  equal to  $\Delta_G$ , i.e.,

$$\max_{\tau_{ij}(k,l)} \arg \left| \begin{array}{l} \sum_{(x,y) \in r_{ij}(k,l)} MND^2(x,y) \geq \sum_{(x,y) \in r_{ij}(k,l), |w(x,y)| < \tau_{ij}(k,l)} w(x,y)^2 + \\ \sum_{(x,y) \in r_{ij}(k,l), |w(x,y)| \geq \tau_{ij}(k,l)} [w(x,y) - \hat{w}(x,y)]^2 \end{array} \right. \quad (16)$$

where the energy of MND defined as (11),  $w(x,y)$  is the wavelet coefficient,  $\hat{w}(x,y)$  is the quantized wavelet coefficient,  $\tau_{ij}(k,l)$  is the quantization step size of  $r_{ij}(k,l)$ .

Therefore, one quantization table that leads to the uniform error energy distribution over all subbands is set up for each subband. It is transmitted in the header of the bit stream for this subband. If after quantization the proportion of zero signals in a block is larger than 7/8 or 15/16, this block is assumed to be unimportant. Its stepsize is recorded as 0, and in the decoder, all values in this block are recovered as 0's. So the coefficients in such an unimportant block need not be transmitted.

When the bandwidth is dynamically assigned to this source encoder,  $\Delta_G$  can be kept constant to maintain the video quality at the same level. If the bandwidth is fixed, i.e., the bit rate is limited for this source encoder, a bunch of  $\Delta_G$  values should be tried and the corresponding bit rates are compared with the available channel capacity. One  $\Delta_G$  value that provides proper bit rate will be chosen finally. And this procedure of refreshing of the quantization table and  $\Delta_G$  value choice is repeated when the Frame Counter reaches a certain number or a drastic movement happens resulting to the need to update the JND model.

### 3.4.2 Lloyd-Max Quantizer

An optimum mean square error quantizer is also tried in our scheme. This quantizer minimizes the mean square error for a given number of quantization levels. For a random variable  $u$ , the reconstruction levels  $a_k$  are calculated with [31]

$$a_k = \frac{\int_k^{k+1} u p_u(u) du}{\int_k^{k+1} p_u(u) du} \quad (17)$$

where  $p_u(u)$  is the continuous probability density function of the random variable  $u$ , and  $t_k$  is the transition level

$$t_k = \frac{(a_k + a_{k+1})}{2}. \quad (18)$$

To design an appropriate quantizer, the distribution of the subband samples needs to be known. It has been suggested that the probability density function of the

subband values and their prediction errors is Laplacian [32]. Although more accurate probability distribution of the subband values, i.e., generalized Gaussian distribution, has been suggested and its shape parameters estimation method is explored [33], the resulting overhead for the calculation and transmission of the parameters is too large. So we still use the Laplacian distribution as a reasonable approximation.

In the procedure of quantization, first, a global object distortion index  $\Delta_G$  is selected. Second, the variance of the wavelet coefficients for each subband is calculated. And all the coefficients are unified. Third, each subband is partitioned into blocks  $(r_{ij}(k,l))$  with size of 8 by 8 or 16 by 16. For each block  $r_{ij}(k,l)$ , the levels of Lloyd-Max quantizer are minimized under the condition that the quantization error energy is less than or equal to the MND energy in this block that has the distortion index  $\delta(i,j)$  equal to  $\Delta_G$ , i.e.,

$$\min_n \arg \left| \begin{array}{l} \sum_{(x,y) \in r_{ij}(k,l)} MND^2(x,y) \geq \sum_{(x,y) \in r_{ij}(k,l), |w(x,y)| < t_1^{(n)}} w(x,y)^2 + \\ \sum_{(x,y) \in r_{ij}(k,l), t_m^{(n)} \leq |w(x,y)| < t_{m+1}^{(n)}} [w(x,y) - a_m^{(n)}(x,y)]^2, m=1,2,\dots,N \end{array} \right| \quad (19)$$

where the energy of MND is defined as in equation (11),  $w(x,y)$  is the wavelet coefficient,  $n$  is the number of quantizer levels with  $n=3,5,\dots,2N+1$ ,  $2N+1$  is the maximal number of quantization levels,  $m$  indicates the interval in which  $w(x,y)$  is located, and  $a_m^{(n)}(x,y)$  is the quantized wavelet coefficient for a quantizer with  $n$  levels.

Here a look-up table is set up for the  $t_m^{(n)}$  and  $a_m^{(n)}$ , since the wavelet coefficients for

quantization have been unified. The index of levels for the optimum quantizer is transmitted in the header of the bit stream of this subband.

### 3.5 Arithmetic Coding and Slicing

The arithmetic coding scheme developed by Witten, Neal and Cleary [25] is widely used in video codecs due to its superior performance. It easily accommodates adaptive models and is computationally very efficient. The arithmetic coding scheme provides efficient compression, however the decoding result of one coefficient depends on the decoding result of the previous one because of the adaptive coding procedure employed.

In the environment of a noisy channel, in order to prevent decoding errors from spreading, a slicing algorithm is derived to truncate the whole subband into short bit streams before arithmetic coding. The idea is to make each such small bit stream carry the same amount of "distortion sensitivity". If we want to segment the subband  $S_I$  into  $I$  short bit streams, we can define  $I$  sets  $G_i$  of the point  $(x,y)$ , such that for each set  $G_i$

$$(i = 1, \dots, I): S_I = \bigcup_i G_i \text{ and}$$

$$G_i = \left\{ (x,y) \left| \begin{array}{l} \sum_{(x,y) \in G_i} \frac{1}{JND^2(x,y)} = \frac{1}{I} \sum_{(x,y) \in S_I} \frac{1}{JND^2(x,y)}, \\ (x,y) \in S_I, (x,y) \notin G_j, j \neq i \end{array} \right. \right\} \quad (20)$$

So every time when such a short bit stream is being encoded, a new adaptive statistical model is set up for it. Before the arithmetic encoded output data of these short bit streams are merged into one bit stream, header and trailer symbols are added so they

can be selected out from the received bit stream at the decoder side. The slicing information is transmitted along with the data stream.

### 3.6 Perceptual Channel Coding

In our practical video transmission system over a satellite channel, rate compatible punctured convolutional (RCPC) codes [27] are adopted. 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 Ramsay 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. [34]

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

$$D_l = \frac{1}{H_l W_l} \sum_{(x,y) \in S_l} JND(x,y)^2 \quad (21)$$

where  $S_l$  is the set of pixels of subband  $l$ ,  $H_l$  and  $W_l$  are height and weight of it separately.  $D_l$  is an indication of the robustness of  $S_l$  to errors. The larger  $D_l$  is, the more robust it is to errors, the higher coding rate we choose. Table 1 shows the  $D_l$  for the video sequence "Calendar-Train".

$L$	$D_l$	$l$	$D_l$	$L$	$D_l$
0	4.5	4	7.5	8	27.6
1	7.6	5	7.5	9	27.6
2	7.6	6	8.7	10	42.8
3	8.8	7	7.4		

**Table 1 Average Distortion  $D_l$  for Each Subband**

From simulations we can see that  $D_l$  divides  $S_l$  into four categories,  $\{S_0\}$ ,

$\{S_1, S_2, S_3, S_4, S_5, S_7\}$ ,  $\{S_8, S_9\}$ ,  $\{S_{10}\}$ , which is intuitive for RCPC unequal error protection. According to their different importance, the subbands are assigned different RCPC coding rates correspondingly.

### 3.7 Video Decoder and Error Concealment

In the video decoder, the functionality of most modules is straightforward. In the error concealment, when errors are detected via CRC decoding in the slices in subband  $S_l$ , these slices are discarded. If no error is detected, but there are some errors in the received slice, the arithmetic decoder can detect some conflicts during decoding sometimes, therefore can find some errors and discard this slice. In this case, coefficients in this slice are retrieved from its DPCM reference if it belongs to the spatio-LLLL temporal-L subband. And the coefficients are set to zero's if the slice belongs to other subbands of higher frequency. The error effect is trivial and will be confined within the slice. In the worst case where errors are not detected, they will not spread to the whole subband due to slicing.

[illegible]



## Chapter 4 Simulation Results

A C/C++ program was written to simulate the performance of our video codec system. For simulation of a video system, C code is better than Matlab due to the large volume of data and the consideration of speed. It is easy to be transplanted between PCs and Sun Workstations to match the requirements of other application software.

After the input video is decomposed into 11 subbands, each of them is treated as one object variable of class Subband. So they can be operated separately for fine adjustment of parameters if necessary. The class Subband provides corresponding methods to deal with the data in subbands, such as:

```
void Quantize( );           //uniform quantization

void adjustStepSize( );     //stepsize adjustment for each block from JND profile

void LQuantize( );         //optimum quantization for Laplacian distribution

void adjustLapQIndex( );   //index of levels adjustment for each block from JND

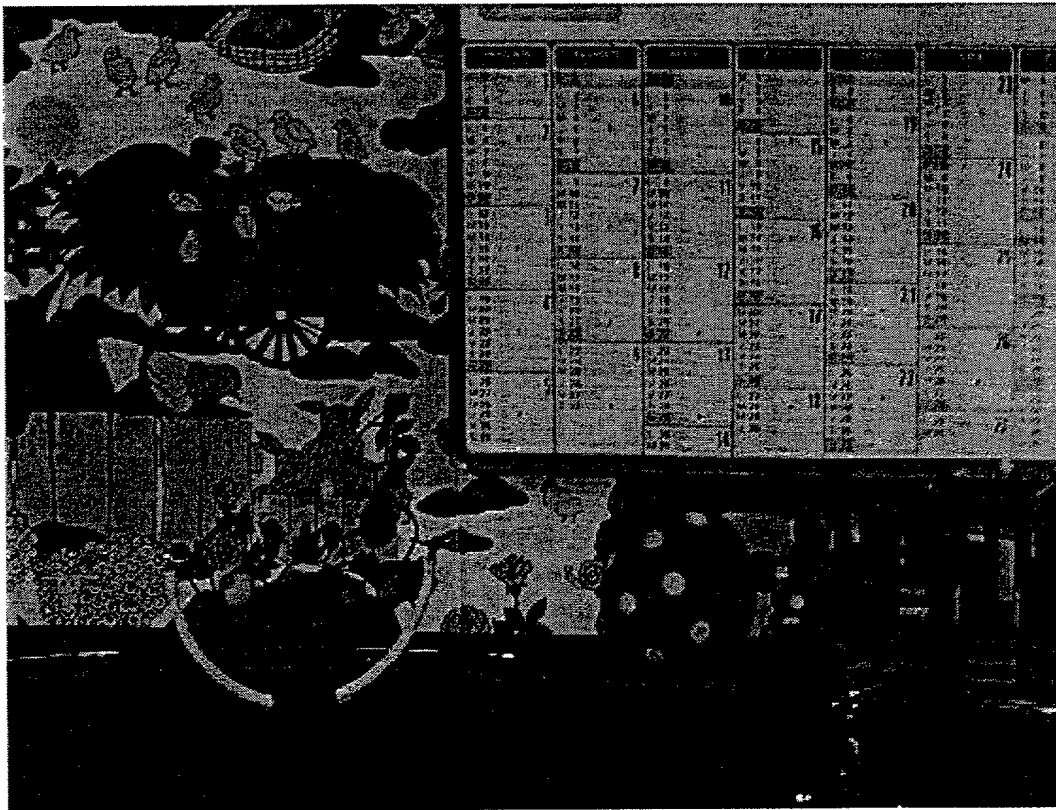
void AC_put_Qvalue( );     //arithmetic encoding

void AC_get_Qvalue( );     //arithmetic decoding
```

We use the video sequences of "Calendar-Train" and "Claire" downloaded from <http://www.ipl.rpi.edu>. They are resized to 512x384 and 352x288 for easy manipulation.

#### 4.1 Spatio-temporal JND Profiles and JND Profiles for Subbands

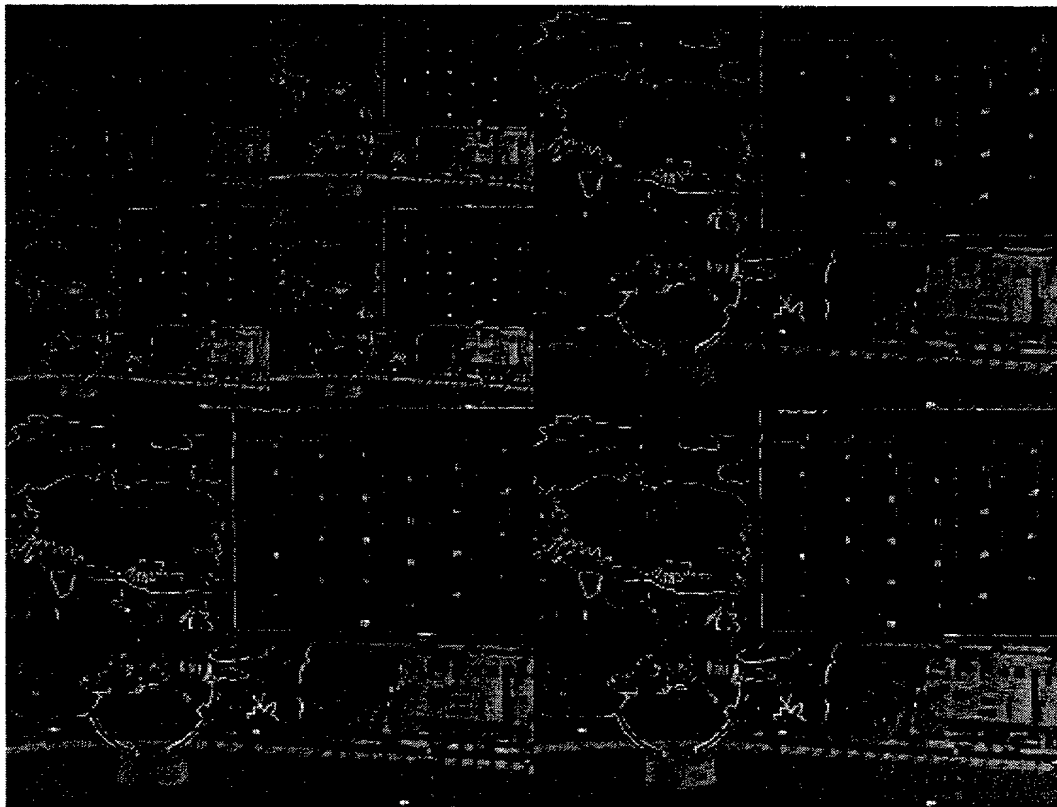
Original frame #1 of "Calendar-Train" and its JND subband profiles are shown in **Figure 7** and **Figure 8**. This JND model is calculated from the signal of this frame and its reference frame. It reflects the visual just-noticeable distortion for video at a speed of 30 frames/sec.



**Figure 7 Frame #1 of "Calendar-Train"**

From the JND subband profile we can tell the difference of the perceptual importance of different subbands. The JND subband profile for spatial LLLL temporal L subband (subband 0) has smaller values of JND (displayed as smaller gray scale values in **Figure 8**) than other subbands. It means that the threshold of just noticeable

distortion is lower, i.e., a small distortion in this subband will be captured by the human visual system while it won't be noticed if it happens in other subbands. We can also see the perceptual importance of different parts in this frame. For example, the calendar part, while comprised of many fine lines and digits, has smaller values of JND, therefore the distortion there is more sensible to human eyes. However, human eyes have more tolerance of the distortion at the moving train, whose body is almost evenly dark.



**Figure 8 JND Subband Profile for Subband 0 to 6**

#### 4.2 Human Perceptual Distortion Measure vs. PSNR



Figure 9 Decoded Frame of "Claire",  $\Delta_G=2.38$ , PSNR=30.80dB

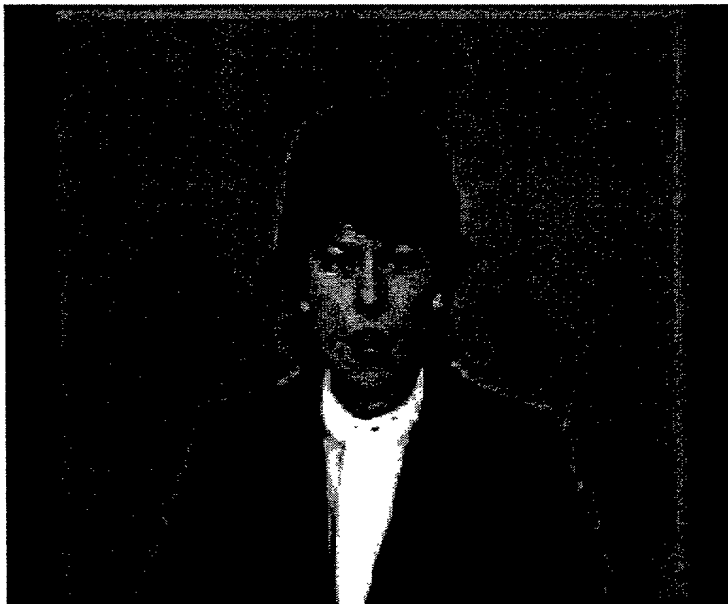


Figure 10 Decoded Frame of "Claire",  $\Delta_G=3.07$ , PSNR=30.15dB

**Figure 9** and **Figure 10** show that our distortion measure (equation (12)) is better than PSNR in the sense that it reflects the subjective visual quality of image/video better. **Figure 9** and **Figure 10** show frame #1 in the decoded sequence of “Claire”. The PSNR of these two frames are almost the same, but the  $\Delta_G$  values indicate that the distortion of **Figure 9** is smaller than that of **Figure 10** as we can tell from direct observation (e.g. shoulder, hair and cheek).

### 4.3 Video Transmission over Satellite Channels

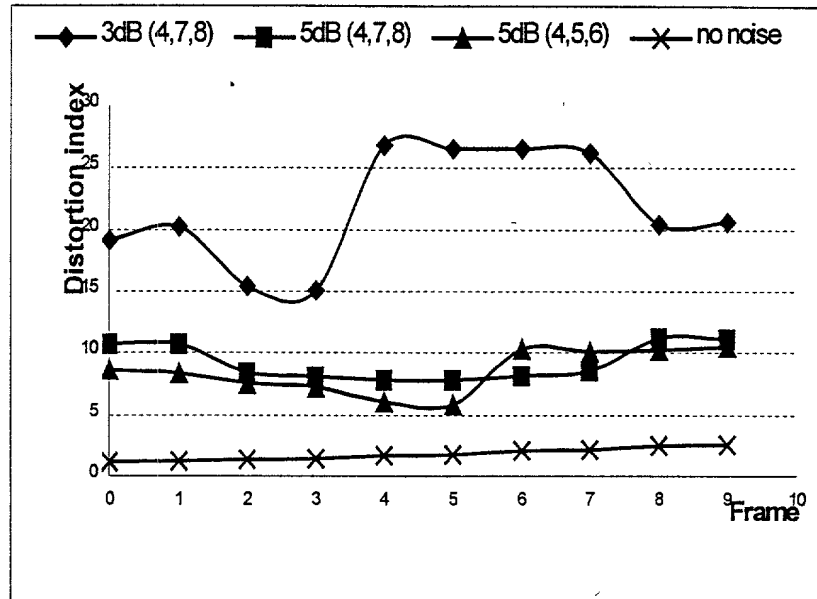
In Chapter 3 we discussed the design of a channel coding scheme using RCPC code, Reed-Solomon Coding, CRC Code and Ramsay interleaver. OQPSK modulation is in our simulation. The human vision model based rate allocation for RCPC is also described. The average JND energy for each subband is calculated as shown in **Table 2**. Because the difference between  $S_{10}$  and  $S_8, S_9$  is relatively small, we finally assign 11 subbands into three error protection groups:  $\{S_0\}$ ,  $\{S_1, S_2, S_3, S_4, S_5, S_7\}$ ,  $\{S_8, S_9, S_{10}\}$ .

Index	4	5	6	7	8	9
Rate	8/18	8/16	8/14	8/12	8/10	8/9

**Table 2 Rate Index of RCPC**

**Table 2** shows the coding rate index. The sequence of “Calendar-Train” is coded and transmitted over AWGN channel at different SNR. **Figure 11** and **Figure 12** show the distortion index of the first 10 decoded frames with different protection

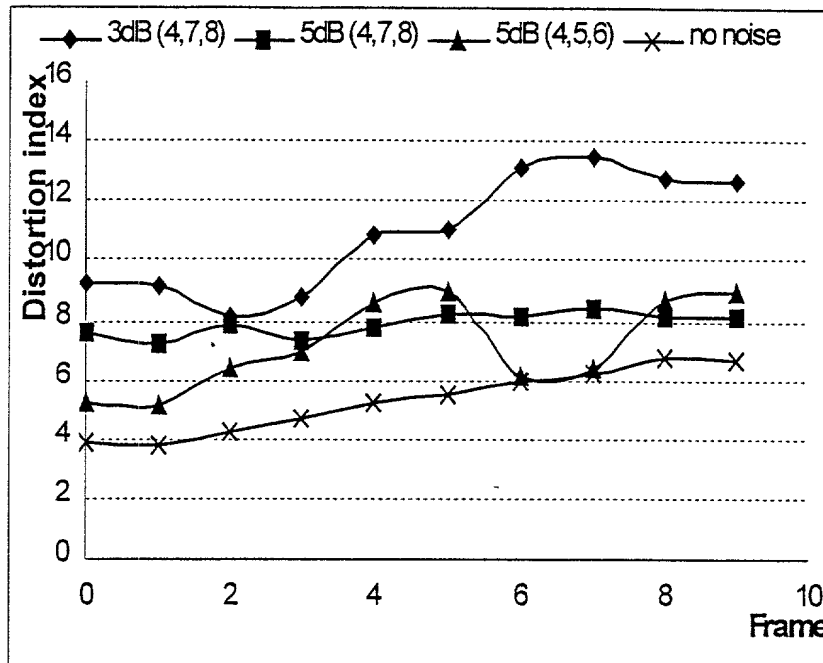
schemes at different SNR (In Figure 11 and Figure 12, the legend 3dB (4,7,8) means that the  $E_b / N_o$  is 3dB and we use channel coding rate 8/18 for subband 0, rate 8/12 for subband 1 to 7, and rate 8/10 for subband 8 to 10).



**Figure 11 Distortion of the Decoded Frames over Noisy Channel with Object Distortion Index=1**

In Figure 11 the original frames are encoded with the object distortion measure  $\Delta_G=1$ , which means the compression brings just noticeable distortion in the pictures.

Even if the channel is ideal, the distortion  $\Delta_G$  goes larger gradually ( $\Delta_G=1.19$  for Frame 0 and 1.77 for Frame 5) in the following frames. The reason is that the JND model was initially constructed from the first two frames and is not renewed for the subsequent frames, which brings bias in the encoding. The larger the frame number is, the more the bias accumulates.

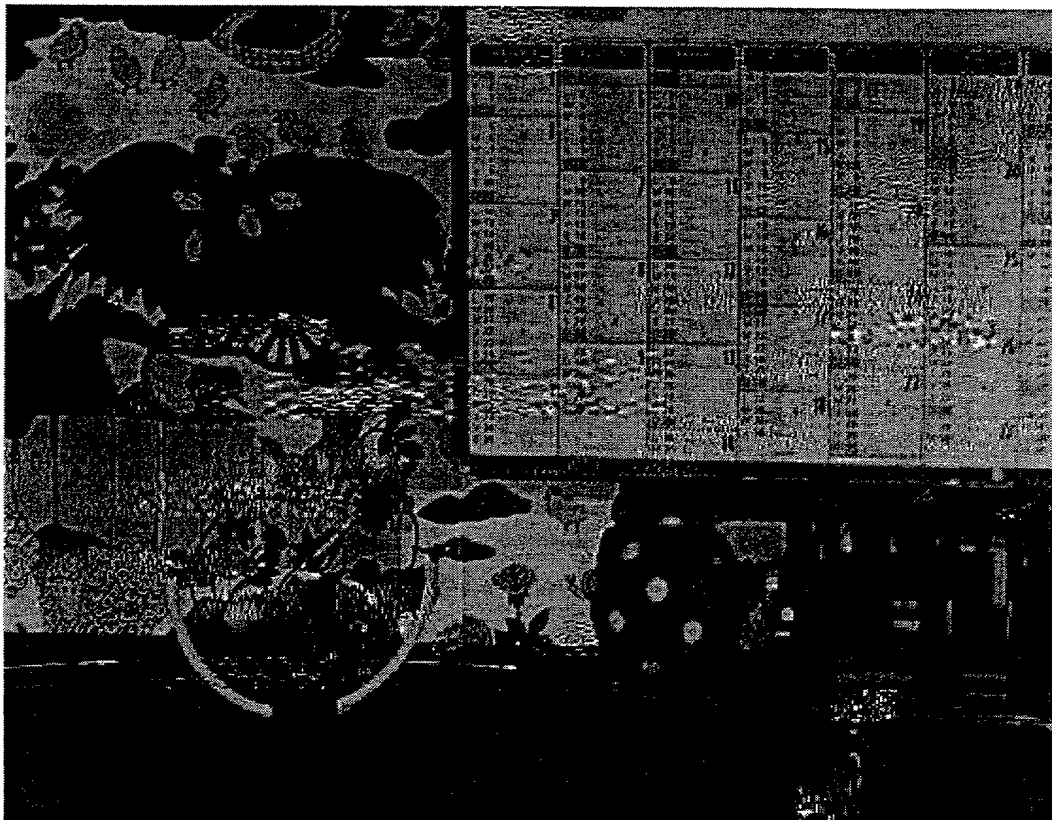


**Figure 12 Distortion of the Decoded Frames over Noisy Channel with Object Distortion Index=5**

In Figure 12 the original frames are encoded with the object distortion measure  $\Delta_G=5$ , which means the perceptual distortion is 5 times of the just noticeable distortion. Intuitively, one would accept that the distortion in Figure 11 should be less than that in Figure 12; however this is not correct. The reason is that the slices generated by the source encoder are longer when  $\Delta_G=1$  (finer step size is chosen and more data is transmitted), and therefore the probability that a slice is corrupted by errors increases as the number of bits in this slice increases at the same bit error rate (BER). So if a video service requires better quality, the corresponding better channel protection scheme should be chosen.

Figure 13 shows frame No. 3 in a recovered sequence of "Calendar-Train".

Some areas corrupted by the channel noise can be observed.



**Figure 13 Decoded Frame #3 of "Calendar-Train" with  $E_b/N_o = 3dB$  (4,7,8) Protection**

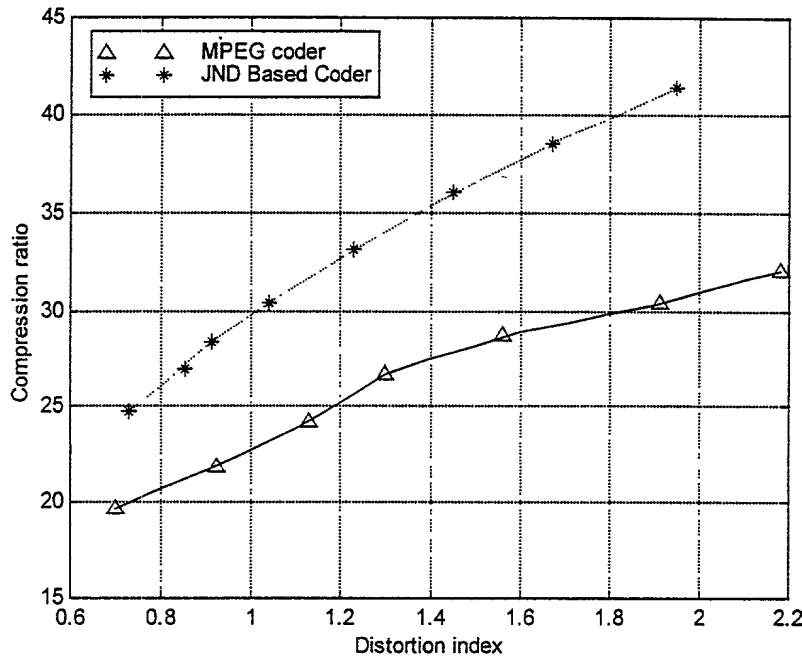
#### 4.4 Comparison with MPEG

The performance of our JND based video codec is compared with the MPEG-1 system. Due to the fidelity of our JND distortion index  $\Delta_G$ , it will be used as the major performance indicator. Because the focus of our system is the subjective quality, more powerful compression schemes (e.g. zero-tree, motion estimation and run-length coding) are not applied in our system. Therefore, our system will be compared with the performance of I frames of MPEG.

There are two MPEG-1 encoders used in our experiments. The performance of these two encoders, say, MPEG A and MPEG B, is a little bit different. MPEG-A has a



higher compression ratio and MPEG-B has a higher PSNR at the same quantization scale. But basically, their output bit streams all follow the MPEG-1 standard and the key features of MPEG like DCT, zig-zag scanning, Huffman coding and run-length coding are implemented in the encoder.



**Figure 14 Performance Comparison between the MPEG Coder (MPEG A) and the JND Based Coder**

In Figure 14 the original sequence “Claire” is encoded according to MPEG-1 at different scales. It gives different compression ratios for the I frames. The distortion indexes of these I frames are calculated using spatial-temporal JND profiles. At the same perceptual distortion level, the JND based encoder provides higher compression ratio than the MPEG-1 encoder for I frames. It also turns out that the subjective video quality of JND based compression is better than the subjective quality of I frames from MPEG-1 at the same compression ratio.

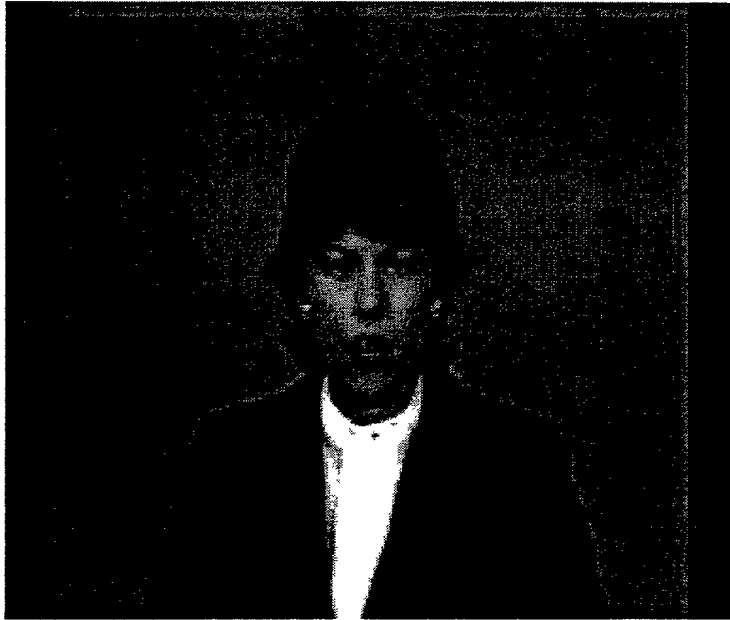


Figure 15 "Claire" from the JND based coder,  $\Delta_G=0.85$ , PSNR=36.3dB, compression ratio=27.0:1

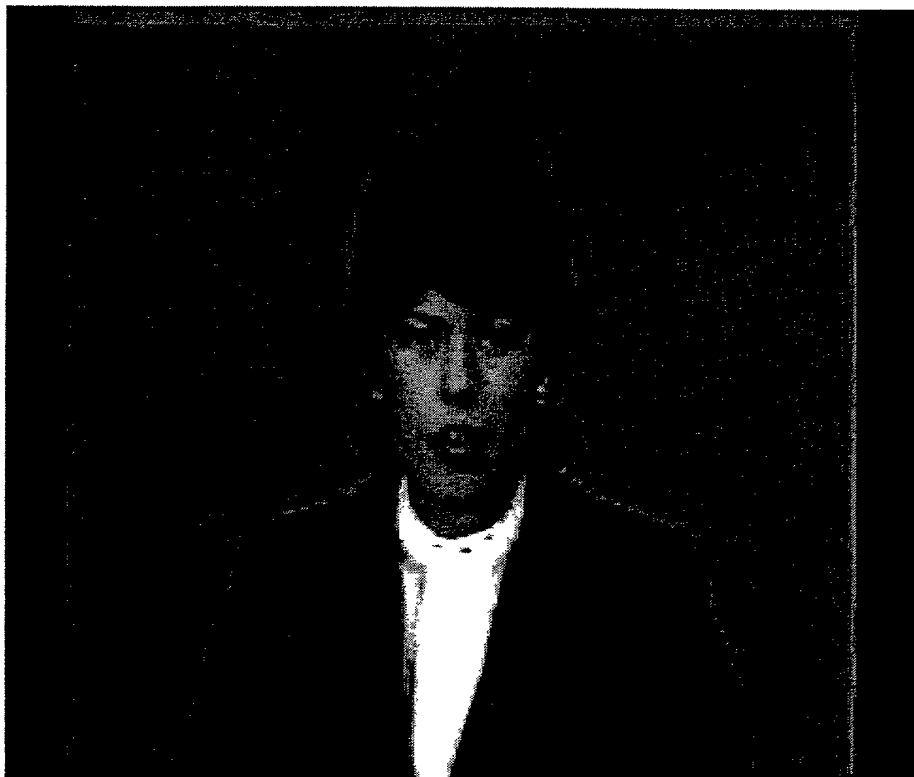


Figure 16 "Claire" I frame from the MPEG-1 coder,  $\Delta_G=1.3$ , PSNR=37.5dB, compression ratio=26.7:1

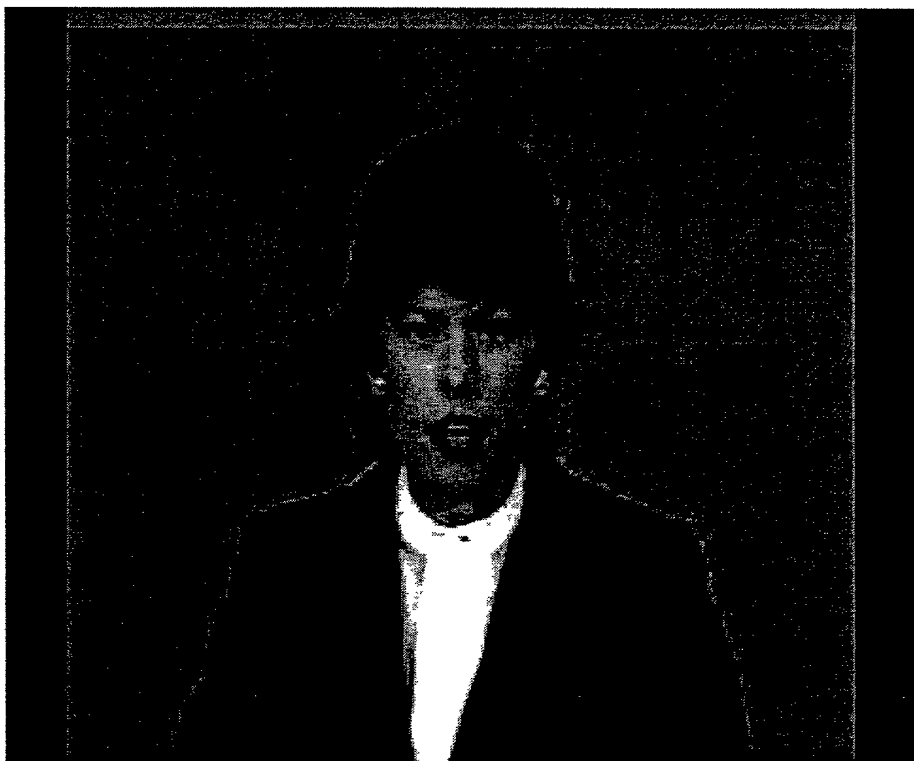
Generally, the PSNR of decompressed frames from the MPEG-1 coder is 1-2 dB larger than that from the JND based coder at the same compression ratio. This is a natural result for the JND based compression scheme. Since with the consideration of JND profile, the irrelevant information for perception is removed, it renders more compression at the same perceptual quality. Ant it renders more numerical error which means less PSNR at the same compression ratio, while keeping the subjective quality superior.

**Figure 15** and **Figure 16** show the decompressed "Claire" of the JND based coder and the MPEG-1 coder (MPEG A). The compression ratio is almost the same. MPEG-1 provides better PSNR and the JND based coder provides better  $\Delta_G$ . From direct observation we can tell that the subjective quality of **Figure 15** is better than that of **Figure 16**.

**Figure 17** and **Figure 18** show the decompressed "Claire" of the JND base coder and MPEG-1 coder (MPEG B). At a higher compression ratio (35:1), their distortion can be seen clearly. But it is proper for us to describe the characteristics of our JND based encoder. The MPEG-1 encoder keeps the quality of the contour of her body, the texture of her hair, and even the glare on her ear ring, while it brings smear to the lady's face, which is the focus of a viewer. The JND based coder assigns the distortion mainly to the moving edges of the body, the moving mouth, and the texture of hair, while it makes the face seem more comfortable. That is the objective of our perceptual optimal coding.

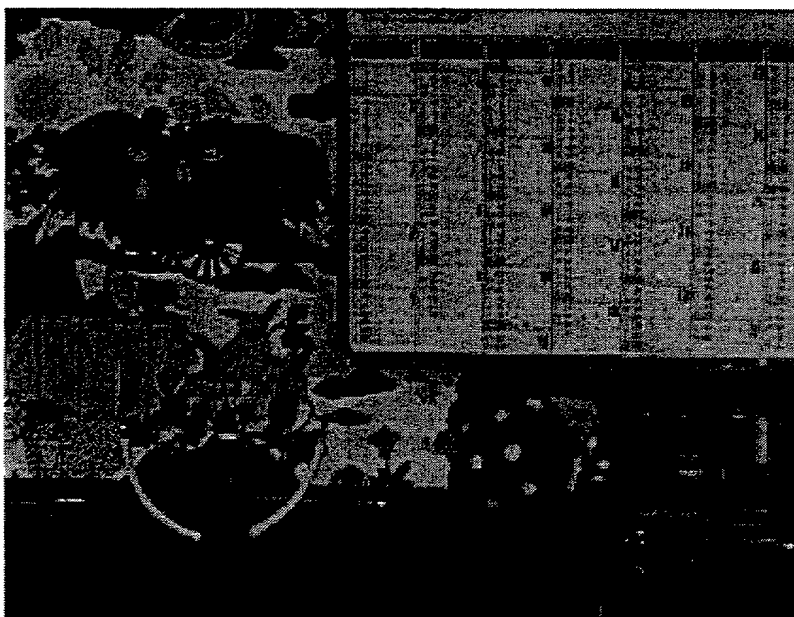


**Figure 17 "Claire" from the JND based coder, PSNR=34.5dB, CR is 35.0:1**

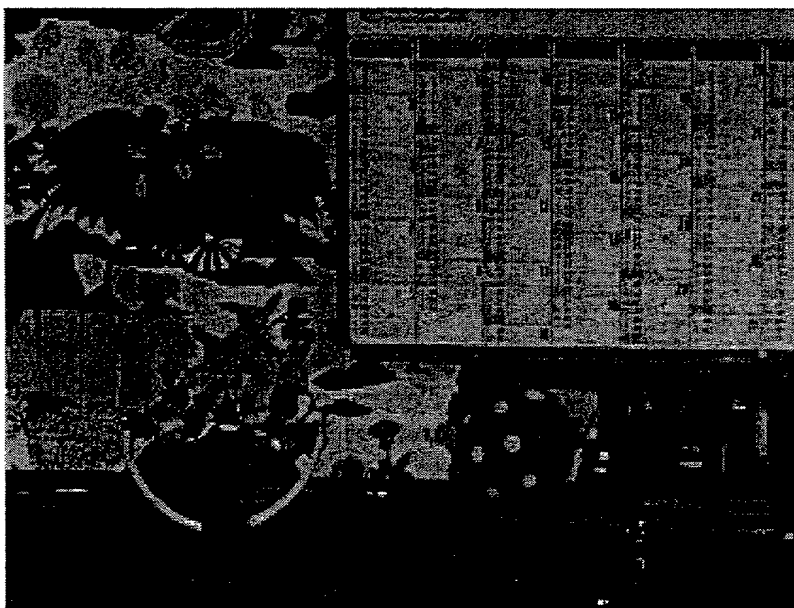


**Figure 18 "Claire" from the MPEG-1 coder, PSNR=37.6dB, CR is 35.5:1**

The video sequence "Calendar-Train" is more difficult to compress due to the complex contents of this sequence. The performance of the JND based coder is very close to that of the MPEG-1 coder on I frames. **Figure 19** and **Figure 20** show the result.



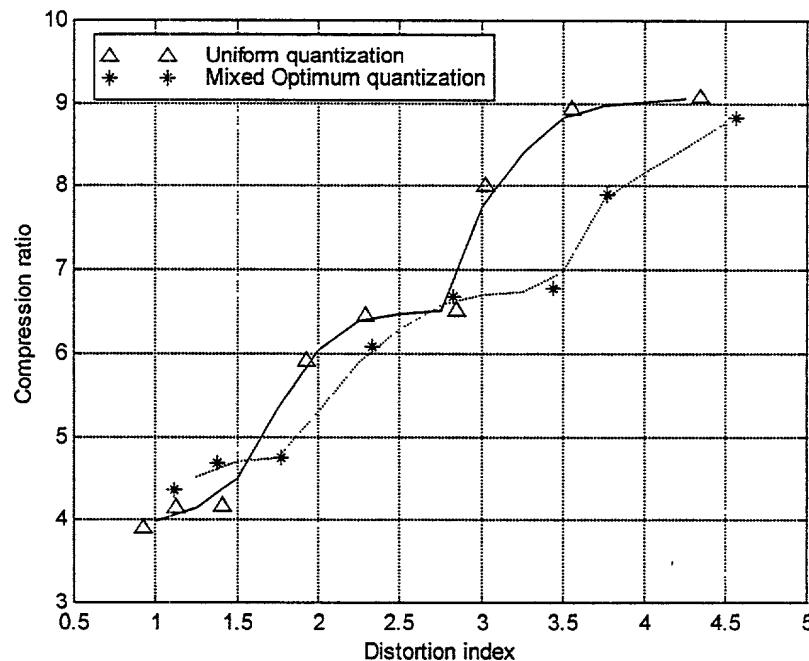
**Figure 19**  
"Calendar-Train"  
from the JND based  
encoder,  $\Delta_G=3.03$ ,  
PSNR=32.6dB,  
compression ratio  
8.93:1



**Figure 20**  
"Calendar-Train"  
from the MPEG-1  
encoder  
(quantization  
scale=7),  $\Delta_G=3.11$ ,  
PSNR=34.0dB,  
compression ratio  
9.06:1

#### 4.5 Quantization Schemes Comparison

The performance of uniform quantization and mixed optimum quantization for Laplacian distribution is shown in **Figure 21**. Here the mixed optimum quantizer implements optimum quantization for subbands 4,5,6,7 and uniform quantization for subbands 0,1,2,3,8,9,10. This scheme is based on experimental adjustment for getting the best performance. But as shown in **Figure 21**, the optimum quantization is not better than the uniform quantization, which is contrary to our expectation. It can be explained by the observation that the application of entropy coding following quantization diminishes the benefits of bit-constrained optimal quantization. Further research leads to the topic of entropy-constrained quantization. In our other simulation experiments, actually uniform quantization is adopted.



**Figure 21 Performance Comparison between Uniform Quantization and Mixed Optimum Quantization**

## Chapter 5 CONDITIONAL ACCESS WITH ARITHMETIC CODING

The conditional access sub-system is being deployed extensively to follow the rapid expansion of the commercial broadcasting industry. It is used to control which customer can get particular program services. Particular programs are only accessible to customers who have satisfied prepayment required.

Current digital broadcasting systems, e.g., DVB, uses a scrambler unit to implement encryption [35]. The bit stream out of the MPEG-2 encoder/multiplexer is fed into the scrambler unit along with the scrambler key, called the Control Word (CW). The CWs are sent to the receiver in encrypted form as an entitlement control message (ECM). The CA subsystem in the receiver will decrypt the control word only when authorized to do so; that authority is sent to the receiver in the form of an entitlement management message (EMM). The subscribers' information necessary for authority is maintained in the RSMS-TAM (Resource and Subscriber Management System – Transmitter Access Management) unit. Scrambling is performed with an ASIC due to the high bit rate.

The benefits of applying data compression prior to transmission are well known and widely exploited. In situations where both data compression and data encryption are desired, such as in commercial video broadcasting, the system will be quite simplified if we combine these two techniques efficiently. The idea of using data compression schemes for encryption dates back to 13<sup>th</sup> century. After the suggestion of Huffman coding, Fraenkel and Klein [36] have shown that the problem of finding the encoding rule given both a sample of the source and the corresponding sample of the encoded file is NP-complete. Gillman et al. [37] have discussed the information-theoretic

impossibility of cryptanalyzing a Huffman code. Witten and Cleary [38] initially suggest the idea of combining data compression together with encryption using adaptive arithmetic coding. In their scheme, the key of the cryptosystem is adopted as the starting point of the model, i.e., taking the encryption key as a specific starting probability distribution for the symbols in the adaptive arithmetic coding model. Boyd [40] gives cryptanalysis for this scheme with the acclamation that there exists an order preserving property that allows a known plaintext attack in the case of a binary alphabet.

Since entropy coding is the indispensable part in all practical video compression systems, in this chapter, we investigate the problem of realizing scrambling by utilizing the cryptographic properties of arithmetic coding. A new scheme of low cost and reliable conditional access is the result.

## 5.1 Introduction to Arithmetic Coding

Theoretically, arithmetic coding encodes a message as a number in the unit interval  $[0,1]$ . Unlike most schemes, including Huffman Coding, there is no specific code word defined for any specific symbol, but how each symbol is encoded depends, in general, on the previous symbols of the string. For the source sequence  $x^n$ , let us denote its probability mass function as  $p(x^n)$  and its cumulative distribution function as  $F(x^n)$ . We can use a number in the interval  $(F(x^n)-p(x^n), F(x^n)]$  as the code for  $x^n$ . For example, expressing  $F(x^n)$  to an accuracy of  $\lceil -\log p(x^n) \rceil$  will give us a uniquely decodable code for the source [39]. However this is equal to the entropy of the message encoded, so that by Shannon's theory we have achieved the theoretical bound. There are two reasons why the theoretical bound cannot usually be achieved in practice [40]:



(1) For a message of unbounded length, arithmetic of unbounded precision is required for maintaining the value of the current interval. In practice this is overcome by scaling procedures which add to the average encoded word length by decreasing the size of the actual interval calculated.

(2) The decoding is not unique unless the length of the message is known, since a single point may represent any interval of which it is a member. Without knowledge of the message length, decoding can proceed indefinitely. This may be overcome in practice either by sending a length indicator or by using an end-of-message symbol. Both of these add overheads to the encoding.

In the sequential encoding procedure, each symbol in the message is assigned a half-open subinterval of the unit interval of length equal to its probability. We call this the coding interval for that symbol. As encoding proceeds a nesting of subintervals is defined. Each successive subinterval is defined by reducing the previous subinterval in proportion to the length of the current symbol's coding interval. This process continues until all symbols have been encoded.

In some cases that a precise probability distribution for the source is unavailable, a dynamical procedure of updating the symbol frequency model, which renders removal of the redundancy quite efficiently, will be used to adapt to the source. This is the basic idea of adaptive arithmetic coding.

## 5.2 Dependency of Arithmetic Coding

In arithmetic coding, the source sequence  $x^n$  is represented by its cumulative distribution function  $F(x^n)$ . The encoding output of the current symbol depends on the

previous one due to the tree-structured encoding procedure. So basically, it is a polyalphabetic cryptosystem. With the knowledge of a source symbol frequency model, the decoder restores the source symbol one by one with successive dependency. We will show this dependency with the following examples. Due to the popularity of the arithmetic coding implementation in C of Witten *et al.* [25], we will base our discussion on it. In this implementation, a binary bit stream is obtained as the encoding output.

Example 1:

We assume the source symbol frequency model is  $\begin{bmatrix} 0 & 1 & 2 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$ , which

indicates that Prob(symbol=0) is 0.3, Prob(symbol=1) is 0.3, Prob(symbol=2) is 0.4.

The source sequence for AC encoding is

1 1 1 2 0 2 0 1 0 1 0 1 2 2 1 0 0 2 1 0

The encoded binary bit stream is

10010000 10010001 01000110 11100100 11 (34 bits)

Note: Here the frequency model doesn't match the distribution of source sequence. But it only decreases the compression ratio and doesn't change the characteristics of AC.

If the third symbol 1 is replaced by 2 and 0 respectively, the encoder will give out the corresponding binary bit stream

10001000 00011000 10011011 01000101 1 (33 bits)

and

10010111 01111010 11010100 01100011 00 (34 bits)

Only a few bits in the front are the same in these three cases, and the numbers of output bits are also different, i.e., the change of the previous symbol changes the encoding of the following symbols completely.

The loss of source symbol also diversifies the encoding of the following symbols. If the sequence of source symbols is

1 1 2 0 2 0 1 0 1 0 1 2 2 1 0 0 2 1 0

(with the first symbol lost), it's encoded to

10001100 10010000 10000101 01101111 (32 bits)

Example 2:

We assume the same symbol frequency model as that in the previous case. The sequence of symbols for encoding is

1 2 2 0 2 0 1 0 1 0 1 2 2 1 0 0 2 1 0 1

The encoding output is

01110000 01010011 10110001 11110010 01 (34 bits)

Suppose there is one bit (7<sup>th</sup> bit) error in the bit stream, which turns it into

01110010 01010011 10110001 11110010 01.

Then the decoded symbol sequence is

1 2 2 0 0 1 0 0 2 2 1 0 2 0 2 2 2 0 2 0

The symbols after the 3<sup>rd</sup> one are totally different from the original symbols.

If the decoder erroneously locates the start bit position in the bit stream, the following symbols cannot be decoded correctly also, since in both cases the decoding path in the tree structure is misled.

Input symbol sequence:

1 2 2 0 2 0 1 0 1 0 1 2 2 1 0 0 2 1 0 1

Output bit stream after encoding:

01110000 01010011 10110001 11110010 01 (34 bits)

1 bit left shift:

11100000 10100111 01100011 11100100 1

Output symbol sequence after decoding:

0 1 1 0 2 0 1 2 2 2 2 1 1 2 0 1 2 1 2 1

1 bit right shift:

00111000 00101001 11011000 11111001 001

Output symbol sequence after decoding:

2 1 1 2 0 2 0 2 1 2 0 0 1 0 2 1 2 1 1 2

From these examples we can see that even with a fixed frequency model, the correctness of encoding/decoding of the previous symbol dominantly decides the correctness of the following one.

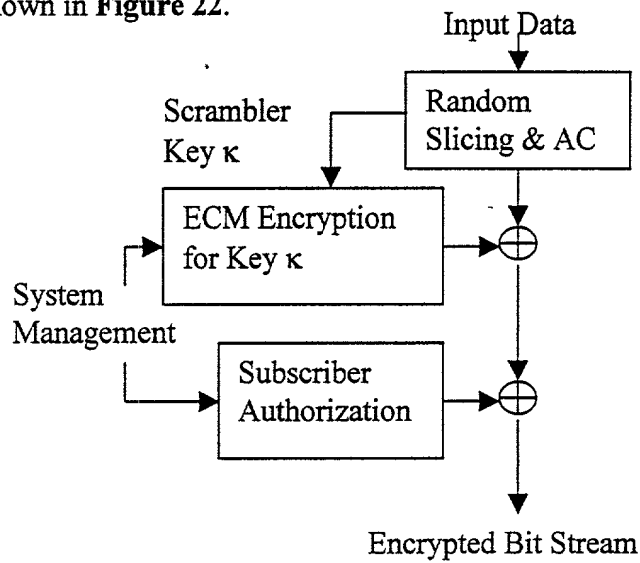
### 5.3 Conditional Access with Arithmetic Coding

Because the precise location of the start bit of an AC encoded bit stream is uniquely important, we propose a conditional access sub-system based on this property and we will basically identify the start bit locations of the bit streams to our scrambler key  $\kappa$ .

First, the data frame which will be transmitted is broken into slices  $\{s_1, s_2, \dots, s_M\}$ . The locations of breaking points are decided by a randomly generated vector  $\mathbf{v}$ . This vector  $\mathbf{v}$  is updated after a short time interval. Each slice is AC encoded respectively into bit stream  $\{b_1, b_2, \dots, b_M\}$ . Then, these bit streams are concatenated into one bit stream  $b_{\text{total}}$ . We assign function  $l(b)$  to represent the length of bit stream  $b$ .

So the value  $\sum_{k=1}^{i-1} l(\mathbf{b}_k)$  determines the start bit positions of  $\mathbf{b}_i$ ,  $i=1, 2, \dots, M$ , in  $\mathbf{b}_{\text{total}}$ . Only the values of  $l(\mathbf{b}_i)$  (i.e. the scrambler key  $\kappa$ ) are encrypted into ECM using any available encryption algorithm, and inserted into the header of this data frame and transmitted.

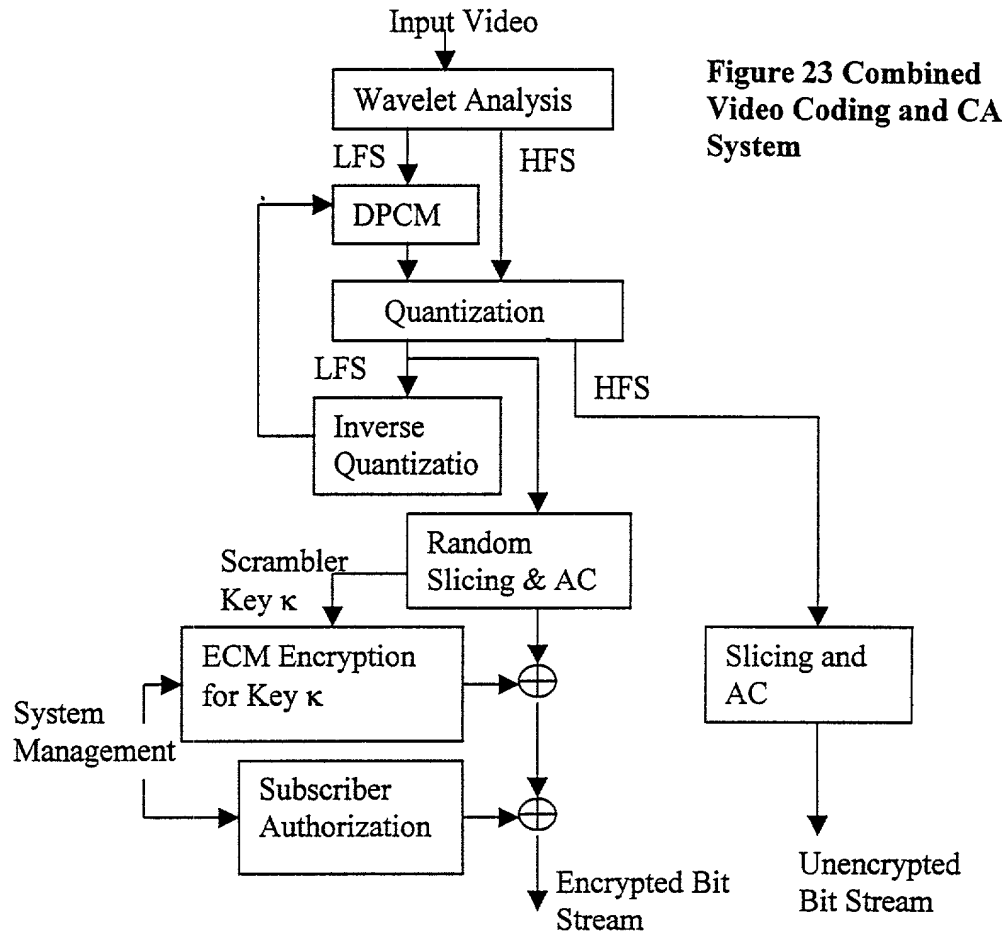
This is shown in **Figure 22**.



**Figure 22 Arithmetic Coding Based Conditional Access Sub-system**

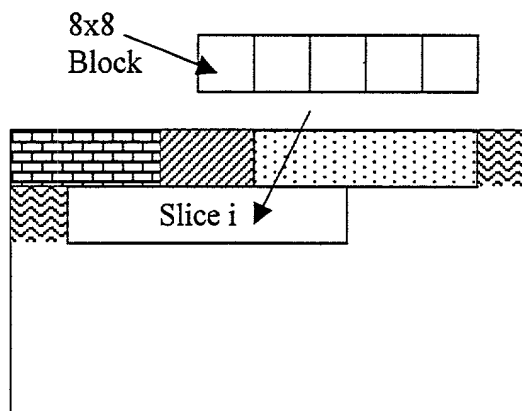
As an example of its application, let us implement this scheme in a video broadcasting system adopting subband coding. The system is shown in **Figure 23**. The input video is decomposed into several subbands using wavelets. The wavelet coefficients in the lowest frequency subband (LFS) are encoded with DPCM before quantization. Since this subband contains the most important information, the encryption of this part is enough for video broadcasting (only the sparse edges and impulses in the high frequency subbands have no meaning to the program viewers). So LFS is fed into our CA sub-system. And the other subbands of high frequency (HFS) are simply sliced at fixed locations and arithmetic encoded. Slicing here has two functions: (1) randomly breaking the LFS into slices to generate the scrambler key  $\kappa$ ;

(2) determinedly slicing the HFS and doing arithmetic coding for each slice to confine the propagation of error due to the noisy channel within short data units, i.e., slices.



For the illegal user, cryptanalysis of such a bit stream will be almost computationally impossible in the broadcasting case, even if the only thing unknown to him is the  $l(b_i)$  values. Let's assume that the frame size as 352x288. If each frame is wavelet decomposed in 2 levels, the most important subband LLLL (LFS), which actually needs encryption, has the size of 88x72. These coefficients are quantized and organized into 8x8 non-overlapped blocks. So we have 11x9 blocks totally. Suppose these 99 blocks are broken into 11 slices for AC randomly as shown in **Figure 24**.

There will be  $\binom{98}{10} = 1.4 \times 10^{13}$  schemes to choose. Only if the illegal user decides these slice lengths in the procedure of decoding, he can know when he has got enough symbols for the current slice, and he should stop the current AC decoding process and start a new procedure for the next slice. But the large amount of possibilities in deciding slice locations eliminates the feasibility of the method of “making mistakes and trying”, i.e., the key space is large enough to stop the real-time decryption of the data in commercial broadcasting.



**Figure 24 Slicing of Subband**

The real conditional access scheme will certainly be made more complicated for the illegal user. First and most important, an adaptive source symbol frequency model will be used. No fixed source symbol frequency model is available and the frequency model is dynamically updated along with the coding procedure. Thus the AC encoding / decoding of the symbol depends more tightly on the previous ones. Even if the illegal user gets the correct start position of  $\mathbf{b}_i$ ,  $i=1,2,\dots,M$ , he cannot decode it without knowing the NoS (number of symbols), which is vital for setting up the adaptive model and variable in different slices. And this NoS will also be encrypted and transmitted in the frame header. For the wavelet coefficients in the video coding system, the data

range differs frame by frame, so the illegal user cannot know it a priori. Example 3 shows the effect of NoS.

Example 3:

The symbol sequence for adaptive AC encoding is

1 2 2 0 2 0 1 0 1 0 1 2 2 1 0 0 2 1 0 1

If we adaptively encode with NoS of 3, and decode with NoS of 4, the decoded output symbols are

2 2 2 0 2 0 3 0 0 0 3 0 1 0 3 0 3 2 1 2

which turns out to be a total loss.

Second, the start position of  $b_1$  in  $b_{total}$  is variable. Stuffing bits with random length is added ahead of it, and the length of stuffing bits is encrypted along with the scrambler key  $\kappa$ .

The third trick played in a practical CA sub-system is that some slices will be picked out randomly for secondary encryption. The mingling of encrypted and unencrypted slices brings more difficulties to the breaking of the arithmetic code.

Fourth, because of the low computation load for scrambling, we can adapt more complicated encryption algorithms for scrambler keys at will and alternate them at different time intervals (with index given before the change).

Taking advantage of the cryptographic properties of arithmetic coding, we considerably decrease the amount of data that need to be processed for scrambling. To compare with the current CA sub-systems, we list the advantages of our scheme as follows: With relative small amount of information in the frame headers for encryption, the computational load of the encryption unit is light. A low cost processor can afford



the task. And the FIFO, timing unit, transfer stream mux/demux unit in the traditional conditional access system can be simplified considerably. The encryption algorithm used to protect the scrambler key information is independent from this scheme. So this system is flexible and more sophisticated algorithms can be easily designed and implemented. Furthermore, since arithmetic coding can be looked as the existing module for compression in a digital video broadcasting system, and breaking the data into slices is naturally necessary to constrain the propagation of error from noisy channels, there is no requirement for the cost of extra devices.

#### 5.4 Summary

Arithmetic coding is a powerful entropy coding scheme for data compression. It is a promising replacement to Huffman coding in video broadcasting systems because of its superior performance. Our scheme combines it with the conditional access subsystem to provide encryption solution at low cost. The illegal user who tries to break that AC based conditional access system faces these problems: (1) The positions of slices  $\{b_1, b_2, \dots, b_M\}$  in the concatenated bit stream are unknown. So they cannot be picked out for decoding. (2) The number of symbols in the adaptive source symbol frequency model is unknown. So it brings ambiguity in decoding. (3) Certain  $b_i$ ,  $i=1,2,\dots,M$ , in the concatenated bit stream are encrypted with probability  $p < 1$ . Such a mixture of cipherdata and plaindata brings more difficulties to cryptanalysis.

According to Shannon's theory on secrecy systems [41], the strength of a cryptosystem is inversely proportional to the source redundancy. Arithmetic coding

diminishes the information redundancy efficiently, so the random appearance of its output bit stream brings a good performance for encryption purposes.

61

## Chapter 6 CONCLUSIONS AND OPEN ISSUES

### 6.1 Video Codec with JND

In this thesis, we explored the application of a human visual model in video compression, channel coding, error concealment and subjective quality measurement. The 3-D subband video coding scheme is optimized by exploiting the psychovisual properties of the human visual system, so that high perceptual quality of the compressed video can be maintained at a reasonable bit rate. In order to reach the optimality where the coding error is invisible or minimally noticeable at a given viewing distance, the human visual model of JND/MND profiles in subbands are calculated. This model renders a perceptually optimal procedure to quantize the wavelet coefficients in these subbands. The estimation of JND profiles is performed in the spatial domain through analyzing local properties of the video signals. Then in the analysis, in the frequency domain, the JND profile is accordingly decomposed by the modulation transfer function of the HVS into subband profiles.

Taking advantage of the JND profiles, the source encoder has the global control for subjective distortion of the compressed video quality through deciding the coding parameter  $\Delta_G$  at the beginning of the coding process. It provides a powerful functionality to control the coding quality ahead of time, which is distinct from the conventional compression schemes.

We developed a new perceptual distortion index based on the JND profile to measure the subjective quality of the compressed video. With this numerical metric, we can have a more general and meaningful judgement on the perceptual distortion of

compression. Its performance is compared with the traditional quality measure PSNR. While PSNR can be used to compare the quality of several compression schemes for the same image sequence, heuristically, our distortion index can be used to describe the distortion extent for different image sequences.

In the scenario of a noisy channel, the functionality of error correction, detection and concealment is jointly realized by the channel coding and source coding. A slicing scheme based on JND is suggested to provide a reasonable information packaging according to its importance to human perception, which is the final destination of video information. The objective of the channel coding scheme selection is to approach the compromise of high error protection ability and low bit rate. The energy of JND is computed for each subband as guidance to the choice of RCPC coding rate. Those subbands of more perceptual importance get better error protection.

## 6.2 Conditional Access using Arithmetic Coding

Arithmetic coding is an efficient entropy coding method. It operates by dispensing with the requirement that each symbol translates into a fixed code. Instead, the code for one symbol in the message merges into that for the next symbol, with no identifiable boundary between. Arithmetic coding is guaranteed to transmit a message in a number of bits that can be made arbitrarily close to its entropy with respect to the model which is used. The probability distribution can change completely from one symbol to another without incurring any penalty in compression efficiency.

The cryptographic property of arithmetic coding is valuable in the design of conditional access sub-systems for commercial broadcasting. Without the knowledge of

start position in the transmitted bit stream for decoding, it is very difficult for the decoder to get synchronization. This contrasts with Huffman coding, where each symbol has its unique corresponding code. Our conditional access sub-system is constructed on the basis of arithmetic coding. The task of the scrambler unit is simplified to break the data into slices. The protection of the information in these slices is provided by the cryptographic feature of the arithmetic coding.

Due to the slicing method also used in our video source coding scheme for error protection, detection and concealment as described previously, to make it consistent with our conditional access sub-system, some variations are necessary. For example, the generation of slicing positions in scrambler cannot be totally random. Taking the JND profile into consideration, the slicing position is basically generated corresponding to the distribution of JND energy as in Chapter 3. Then, for the scrambler, it will modify the position to some extent according to its scrambling key. Since the JND profile is not fixed during the video sequence, it is not easy to figure it out to compute the basis of slicing position (The JND profile data is sliced also). This modification decreases the strength of encryption, but the cryptanalysis for it remains complicated.

### 6.3 Video Codec with Motion Estimation

Our current video codec is based on subband coding and no motion estimation scheme is used. Actually, lots of explorations have been done in this field. To improve the compression performance, overlapped or multi-scale motion estimation can be adopted along with subband coding. The JND profile can be used for the quantization of the residues after motion compensation. The perceptual distortion energy is still

assigned to proper coefficients in this way. Similar to the technology of P and B frame in MPEG, the distortion energy assigned to these motion predicted frames can be larger while the interframe coded frames have small perceptual distortion to provide a stable basis for motion estimation.

#### 6.4 Joint Source-Channel Coding Based on Human Visual Model

In Chapter 3, in order to constrain the error from noisy channels, we have developed a heuristic way (equation (19)) to segment the data of each subband into slices based on JND profiles. The length of the slice is proportional to its JND energy. Arithmetic coding is applied on each slice with a new adaptive statistical model. As a result, any corrupted slice carries almost the same amount of perceptual importance measured using JND.

Our future work is to develop a route to optimize this slicing process and combine it with the choice of RCPC coding rate for different subbands. Since the statistical expectation of wavelet coefficients in each subbands is close to zero in video coding, whenever a transmission error happens, the corresponding corrupted slice value will be restored as zeros if no other correlative information is available. So the possible error equals the wavelet coefficient value  $w(x,y)$ . The distortion index for block  $r_{ij}$  is defined as:

$$\delta(i, j) = \frac{\sum_{(x,y) \in r_{ij}} w^2(x, y)}{\sum_{(x,y) \in r_{ij}} JND^2(x, y)}, \quad (22)$$

where  $w(x,y)$  is the wavelet coefficient at  $(x,y)$ , and  $r_{ij}$  is the block (usually 8 by 8 pixels). The distortion index  $\delta_{ij}$  is also referred as relative distortion for block  $r_{ij}$ .

At the beginning of the slicing procedure for the  $k$ th subband ( $k=0,1,\dots,K-1$ , and  $K$  is the number of subbands), the amount of relative distortion in each slice is determined as  $\xi_k$ . The slice is composed of successive blocks, such that in each slice  $S_{k,l}$  ( $l=1,2,\dots,L_k$ , and  $L_k$  is the total number of slices in  $k$ th subband), heuristically, there holds

$$\sum_{\substack{(i,j) \\ r_{ij} \in S_{k,l}}} \delta(i,j) \cong \xi_k \quad (23)$$

As a prior, we have the knowledge about the noisy channel, channel coding and modulation in the form of bit error rate  $P_{m,n}$  which corresponds to  $m$ th RCPC coding scheme and channel SNR at  $n$ dB. We also assume each slice will be coded using arithmetic coding into a bit stream with the length of  $c$  bits with small variance. Because any slice with one bit error after channel decoding will not be correctly restored in arithmetic decoding, the probability of the corruption of each slice in the  $k$ th subband is calculated as:

$$P_k = 1 - (1 - P_{m,n})^c \quad (24)$$

The relative distortion for the  $k$ th subband is calculated as

$$\Xi_k = L_k P_k \xi_k \quad (25)$$

The relative distortion for the whole image is

$$\Xi^* = \sum_{k=0}^{K-1} \omega_k \Xi_k \quad (26)$$

where  $K$  is the number of subbands and  $\omega_k$  the perceptual weight of the  $k$ th subband, as defined in equation (9).

Equations (22 ) to (26) show the relationship between the global relative distortion of the video and the RCPC coding schemes. Through the adoption of different RCPC coding rate for different subbands, we get channel coding schemes corresponding to certain relative distortion indexes, which reflect the video quality perceptually. This leads to a reasonable control over the channel coding schemes to render the best available video quality, i.e. when the transmission system knows the channel SNR, it can configure the RCPC coding rates for different subbands through a look-up-table, and gets the least relative distortion on video quality. Since the distortion based on the human visual model is assigned to source coding through the selection of quantization stepsize and to the channel coding through the selection of RCPC coding rate, we can adjust the distribution jointly. Essentially, the whole system realizes the joint source-channel coding.



## Bibliography

The following references are hereby incorporated by reference:

1. J. Woods and S. O'Neil, "Subband coding of images", *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp1278-1288, Oct. 1986
2. Y. H. Kim and J. Modestino, "Adaptive entropy-coded subband coding of images", *IEEE Trans. Image Processing*, vol. 1, pp31-48, Jan. 1992
3. G. Karlsson and M. Vetterli, "Three dimensional subband coding of video", in *Proc. ICASSP*, 1988
4. H. Gharavi, "Subband image coding", in *Subband Coding of Video Signals* (J. W. Woods, Ed.). Boston: Kluwer, 1990
5. Taubman and A. Zakhori, "Multirate 3-D subband coding of video", *IEEE Trans. Image Processing*, vol. 3, pp572-588, Sept. 1994
6. C. Podilchuk, N. Jayant and N. Farvardin, "Three-dimensional subband coding of video", *IEEE Trans. Image Processing*, vol. 4, pp125-139, Feb. 1995
7. J. Y. Tham, S. Ranganath and A. Kassim, "Highly scalable wavelet-based video codec for very low bit-rate environment", *IEEE JSAIC*, vol. 16, pp12-27, Jan. 1998
8. N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels", *IEEE JSAIC*, vol. 10, pp926-943, June 1992
9. O. Kwon and R. Chellappa, "Region adaptive subband image coding", *IEEE Trans. Image Processing*, vol. 7, pp632-648, May 1998
10. S. Martucci, I. Sodagar, T. Chiang and Y. Zhang, "A zerotree wavelet video coder", *IEEE Trans. on Circuits and Systems for Video Tech.*, vol.7, pp109-118, Feb. 1997
11. J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients", *IEEE Trans. on Signal Processing*, vol. 41, pp3445-3462, Dec. 1993

12. A. Said and W. Pearlman, "A new fast and efficient image coder based on set partitioning in hierarchical trees", *IEEE Circuits and Systems for Video Tech.*, vol. 6, pp243-250, June 1996
13. K. Cinkler, "Very low bit-rate wavelet video coding", *IEEE JSAIC*, vol. 16, pp4-11, Jan. 1998
14. D. Taubman, "Directionality and scalability in image and video compression", *Ph.D. dissertation*, Univ. California, Berkeley, 1994
15. J. Y. Tham, S. Ranganath and A. Kassim, "Scalable low bit rate video compression using motion compensated 3-D wavelet decomposition", in *IEEE ICCS/ISPACS 1996*, vol. 3, pp39.7.1-39.7.5, Nov. 1996
16. N. Jayant, J. Johnston and R. Safranek, "Signal compression based on models of human perception", *Proc. IEEE*, vol. 81, pp1385-1422, Oct. 1993
17. 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, pp467-476, Dec. 1995
18. 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, pp143-156, April 1996
19. S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity", *Measurement and Prediction of Visual Quality*, 1991
20. X. Ran and N. Farvardin, "A perceptually motivated three-component image model - Part I: Description of the model", *IEEE Trans. Image Processing*, vol. 4, pp401-415, April, 1995

21. M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform", *IEEE Trans. Image Processing*, vol. 1, pp205-220, April 1992
22. D. H. Kelly, "Motion and vision II. Stabilized spatio-temporal surface", *J. Opt. Soc. Amer.*, vol. 69, pp1340-1349, Oct. 1979
23. N. Jayant, "Signal compression: Technology targets and research directions", *IEEE JSAIC*, vol. 10, pp314-323, June 1992
24. N. Farvardin and J. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless source", *IEEE Trans. Information Theory*, vol. 30, pp485-497, May 1984
25. I. Witten, R. M. Neal and J. G. Cleary, "Arithmetic coding for data compression", *Comm. of the ACM*, Vol. 30, pp520-540, June 1987
26. N. Jayant, "Signal compression: Technology targets and research directions", *IEEE JSAIC*, vol. 10, pp796-818, June 1992
27. J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and Their Application", *IEEE Trans. Comm.*, vol. 36, no.4, pp. 389-400, April 1988
28. A. N. Netravali and B. G. Haskell, "Digital Pictures: Representation and Compression", New York: Plenum, 1988
29. I. Daubechies, "Ten lectures on wavelets", *Society for Industrial and Applied Mathematics*, Philadelphia, 1992
30. S. G. Mallat, "A theory for multi-frequency signal decomposition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp674-693, July 1989
31. A. K Jain, "Fundamentals of Digital Image Processing", Prentice Hall

32. H. Gharavi and A. Tabatabai, "Subband coding of monochrome and color images",  
*IEEE Trans. Circuits Syst.*, vol. 35, pp207-214, Feb. 1998
33. K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized  
Gaussian distributions in subband decompositions of video", *IEEE Trans. on  
Circuits and Sys. For Video Tech.*, vol. 5, No. 1, Feb. 1995
34. J. Gu, Y. Jiang and J. Baras, "A Practical Transmission System Based on the  
Human Visual Model for Satellite Channels", *ISR Technical Report* TR 99-14, 1999
35. W. Kim, K. Chen and H. Cho, "Design and implementation of MPEG-2/DVB  
scrambler unit and VLSI chip", *IEEE Trans. Consumer Electronics*, vol. 43, pp980-  
985, Aug. 1997
36. A. S. Fraenkel and S. T. Klein, "Complexity aspects of guessing prefix codes",  
*Algorithmica*, vol. 12, pp. 409-419, 1994
37. D. W. Gillman, M. Mohtashemi and R; L. Rivest, "On breaking a Huffman code",  
*IEEE Trans. Information Theory*, vol. 42, pp972-976, May 1996
38. I. H. Witten and J. G. Cleary, "On the privacy afforded by adaptive text  
compression", *Computers and Security*, vol. 7, pp397-408, 1988
39. T. Cover and J. Thomas, "Elements of information theory", Prentice Hall, 1991
40. C. Boyd, "Applying compression coding in cryptography", *Cryptography and  
Coding*, vol.2, Clarendon Press, 1992
41. C. E. Shannon, "Communication theory of secrecy systems", *Bell Systems Technical  
Journal*, pp656-715, 1949
42. S. Dalbague, J. Baras and N. Sidiropoulos, "Compact image coding from multiscale  
edges", *ISR Technical Report* TR 98-61, 1998

Under the paperwork reduction act of 1995, no persons are required to respond to a collection of information unless it displays a valid CMB control number

**PROVISIONAL APPLICATION FOR PATENT COVER SHEET**

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (b)(2).

Docket Number

IS-99-030

Type a plus sign (+)  
inside this box →

+

**INVENTOR(s) APPLICANT(s)**

LAST NAME

FIRST NAME

MIDDLE NAME/INITIAL

RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)

Gu  
Jiang  
BarasJunfeng  
Yimin  
John

S.

Greenbelt MD  
Germantown MD  
Potomac MD**TITLE OF THE INVENTION (280 characters max)****Method and Apparatus for Perceptually Based  
Compression and Transmission of Video Streams****CORRESPONDENCE ADDRESS (including country if not United States)**Mr. Ernest R. Graf  
Office of Technology Liaison  
University of Maryland  
4312 Knox Road  
College Park, MD 20742**ENCLOSED APPLICATION PARTS (check all that apply)**

Specification

Number of Pages

108



Small Entity Statement



Drawing(s)

Number of Sheets



Other (specify)

ADVANCE NOTICE SERIAL  
NUMBER POSTCARD**METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (check one)**

A check or money order is enclosed to cover the filing fees

The commissioner is hereby authorized to charge  
filing fees and credit Deposit Account Number.

210683

FILING FEE  
AMOUNT (\$)

\$ 75.00

The invention was made by an agency of the United States Government or under contract with an agency of the United States Government.

☐ No.

Yes, the name of the Government agency and the Government contract number are: NASA - Lewis - NCC3528 &amp; NCC5227

Respectfully Submitted via Express Mail,

EL276682154US

SIGNATURE

Date

5/28/99

TYPED or PRINTED NAME

Ernest R. Graf

REGISTRATION NO.  
(if appropriate)

41,043



Additional inventors are being named on separately numbered sheets attached hereto

**USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT**

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending on the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

**VERIFIED STATEMENT CLAIMING SMALL ENTITY STATUS  
(37 CFR 1.9(f) & 1.27(d))--NONPROFIT ORGANIZATION**

Docket Number (Optional)

IS-99-030

Applicant or Patentee: Junfeng Gu, Yimin Jiang and John S. Baras

\*>Application< or Patent No. : \_\_\_\_\_

Filed or Issued: FILED HEREWITH

Title: **Method and Apparatus for Perceptually Based Compression and Transmission of Video Streams**

I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below:

NAME OF NONPROFIT ORGANIZATION OFFICE OF TECHNOLOGY LIAISON

ADDRESS OF NONPROFIT ORGANIZATION UNIVERSITY OF MARYLAND  
4312 KNOX RD, COLLEGE PARK, MD 20742

**TYPE OF NONPROFIT ORGANIZATION:**

☒ **UNIVERSITY OR OTHER INSTITUTION OF HIGHER EDUCATION**

☐ **TAX EXEMPT UNDER INTERNAL REVENUE SERVICE CODE (26 U.S.C. 501(a) and 501(c)(3))**

☐ **NONPROFIT SCIENTIFIC OR EDUCATIONAL UNDER STATUTE OF STATE OF THE UNITED STATES OF AMERICA**  
(NAME OF STATE \_\_\_\_\_)  
(CITATION OF STATUTE \_\_\_\_\_)

☐ **WOULD QUALIFY TAX EXEMPT UNDER INTERNAL REVENUE SERVICE CODE (26 U.S.C. 501(a) and 501(c)(3)) IF LOCATED IN THE UNITED STATES OF AMERICA**

☐ **WOULD QUALIFY AS NONPROFIT SCIENTIFIC OR EDUCATIONAL UNDER STATUTE OF STATE OF THE UNITED STATES OF AMERICA IF LOCATED IN THE UNITED STATES OF AMERICA**  
(NAME OF STATE \_\_\_\_\_)  
(CITATION OF STATUTE \_\_\_\_\_)

I hereby declare that the nonprofit organization identified above qualifies as a nonprofit organization as defined in 37 CFR 1.9(e) for purposes of paying reduced fees to the United States Patent and Trademark Offices regarding the invention described in:

- ☒ the specification filed herewith with title as listed above.  
☐ the application above.  
☐ the patent identified above.

I hereby declare that rights under contract or law have been conveyed to and remain with the nonprofit organization regarding the above identified invention. If the rights held by the nonprofit organization are not exclusive, each individual, concern or organization having rights in the invention must file separate verified statements averring to their status as small entities and that no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern or organization having any rights in the invention is listed below:

- ☐ no such person, concern or organization exists.  
☒ each such person, concern or organization is listed below

NASA - Lewis NCC 3528

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING

WAYNE E. SWANN

TITLE IN ORGANIZATION OF PERSON SIGNING

EXECUTIVE DIRECTOR, OFFICE OF TECHNOLOGY LIAISON

ADDRESS OF PERSON SIGNING

4312 KNOX RD, COLLEGE PARK, MD 20742

SIGNATURE

Wayne E. Swann

DATE

8-27-99

\*\*>Burden Hour Statement: This form is estimated to take .3 hours to complete. Time will vary depending on the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, D.C. 20231.

Rev 2, July 1996