A Perceptually Optimized 3-D Subband Codec for Video Communication over Wireless Channels

Chun-Hsien Chou and Chi-Wei Chen

Abstract— **Visual communication over wireless channels is becoming an important service in multimedia communications. Because of the limited bandwidth and high error rates of the wireless channel, the video codec should be designed to have high coding efficiency in maintaining acceptable visual quality at low bit rates and robustness to suppress the distortion due to transmission errors. In this paper, the coding efficiency of a 3-D subband video codec is optimized by removing not only the redundancy due to spatial and temporal correlation but also perceptually insignificant components from video signals. Unequal error protection is applied to the source code bits of different perceptual importance. An error concealment method is employed to hide the distortion due to erroneous transmission of perceptually important signals. The evaluation of each signal's perceptual importance is made first from measuring the just-noticeable distortion (JND) profile as the perceptual redundancy inherent in video signals, and then from allocating JND energy to signals of different subbands according to the sensitivity of human visual responses to spatio-temporal frequencies. Simulation results show that acceptable visual quality can be maintained in transmitting video sequences with low bit rates (< 64 kbps) over the wireless channel of high error rates (up to BER = 10^{-2}), and the distortion due to erroneous transmission of coded data can be effectively suppressed. In the simulation, the noisy channel is assumed to be corrupted by the random errors depending on the average strength of the received wave and the burst errors due to Rayleigh fading.**

I. Introduction

With the fast growing business in wireless access of multimedia information, the visual communication over wireless channels has become an important service in multimedia communications [1], [2]. However, due to the limitations of power and complexity, and high transmission error rates, wireless channels have narrower bandwidth than wired channels such as those used in ISDN [2]–[5]. In order to maintain acceptable visual quality of the low bit-rate video signals through noisy channels, a video codec of high coding efficiency has to be robustly designed. To design the video codec that meets the requirement, at least four issues must be taken into account. These are the effective exploitation of characteristics of human visual system (HVS) [3], [4], [6], the prevention of error propagation [3], [8], the concealment of erroneous data and the application of unequal error protection (UEP) to coded data of different significance [7]–[9].

In the past decade, the most widely used video coder structure is hybrid coding scheme in which 2-D transform is applied to the prediction error signal from the block-matching motion estimation along the temporal axis [10], [11]. In general, the hybrid coder is a recursive structure that maintains the state of the decoder within a predictive coding loop. As the transmission error occurs, the hybrid coder cannot track the state of the decoder and the distortion may propagate infinitely at the decoder [11]–[13]. To prevent error propagation, the complexity and bit rate cost of a more robust hybrid coder can be very high [2], [4]. On the other hand, at the blocks where motion compensation failed, the blocky distortion can be perceptually bothersome when prediction errors are coded with the discrete cosine transform (DCT) at low bit rates.

Several nice features inherent in the spatio-temporal frequency subband decomposition have been found useful in developing an efficient and robust video codec [12]–[16]. It provides the signal synthesis which avoids blocking effects at low bit rates. It provides an environment for adaptive coding that can incorporate the human visual system in bit allocation [15], [17] or distortion allocation [18]–[20]. It allows for producing a single embedded bit stream which supports multiresolution decoding and bit rate scalability. By replacing motion estimation with temporal filtering, any type of apparent motion can be systematically captured in different spatio-temporal frequency subbands, and computational complexity is lowered. By exploiting temporal correlation, the predictive coding loop can be eliminated, and the propagation of errors can be avoided. To attain high energy compaction in subband decomposition, it is favorable for the 3-D subband coding scheme to employ motion adaptation [12], [13]. However, considering coder complexity and assuming that only videophone or head-and-shoulder sequences are considered, the 3-D subband coding scheme without motion adaptation is adequate [12], [14], [15].

The perceptual coding that matches the compression algorithm to characteristics of human visual perception has been considered as one of the promising solutions to improve the video coding efficiency, and has recently become an important area of research [16]–[24]. Among numerous approaches to perceptual coding, the concept of just-noticeable distortion (JND) profile addressed by Jayant [6], [21] has been successfully applied to the coding of wideband audio [25] and achromatic images [17]–[20]. The 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

Manuscript received July 18, 1995; revised January 2, 1995. This paper was recommended by Associate Editor H. Gharavi.

The authors are with the Department of Electrical Engineering, Tatung Institute of Technology, Taipei, Taiwan, 10451, R.O.C.

Publisher Item Identifier S 1051-8215(96)03015-7.

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, while the average peak signal-to-noise ratio (PSNR) of the contaminated video signals should be as low as possible [18]–[20]. 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.

Suppose transparent coding cannot be attained due to a tight bit-rate budget, a minimally-noticeable-distortion (MND) profile rather than the JND profile is required, by which the increased distortion can be uniformly distributed over the reconstructed video signals and thus be minimally perceptible [6], [17], [18]. By incorporating the MND profile into coding process, the perceptual quality of the reconstructed video signals is expected to degrade gracefully if the available bit rate is reduced.

In subband coding, the errors occurring in different subbands lead to distortions of varying perceptibility in the reconstructed video signals. To maintain good visual quality of the transmitted video signals, the error probability of perceptually important (PI) source code bits should be lower than that of perceptually unimportant (PU) source code bits. Therefore, the source code bitstreams of different subbands should be protected in different degrees according to their perceptual importance in the reconstructed video sequence. With an effective UEP scheme, it is possible to robustly maintain the visual quality of the transmitted video signals without sacrificing the channel bandwidth. It has been found that BCH codes are very suitable for unequal error protection [3], [4], [9], [26]. One advantage of using BCH codes is that only a single encoder-decoder pair is needed for the design of different BCH codes of the same length, and the hardware can be easily implemented [26]. In order to gain high coding efficiency, source data are encoded with variable lengths, to which the UEP that exploits varying sensitivities of bits at fixed positions are difficult to apply [3], [4]. An expedient approach is the application of UEP to the bit segments of different lengths [8]. Thus, the error propagation due to erroneous bits can be confined within the segment and resynchronization is made possible. Once the corrupted segment is detected, local error concealment can be performed by replacing the distorted part by linear interpolation from neighboring data [8], [27]. In this paper, video signals are assumed to be transmitted through the wireless channels that are corrupted by two types of errors [3]. One is the stationary random errors that depends on the mean strength of the received wave, or on the distance between the station and the portable terminal. The other is the variable errors that are caused by the motion of the portable terminal, or the so-called Rayleigh fading [1]–[4]. The Rayleigh fading causes the wave strength dips that result in burst errors (Fig. 1) [3]. As the wave strength is 10 dB smaller than the mean wave strength, the BER can be 10 times worse. The average interval over which wave strength dips is dependent upon the Doppler frequency of the received wave. As the velocity of the portable terminal increases, the Doppler frequency increases, the average interval between dips becomes shorter, and the average BER worsens [3]. Therefore, the faster the portable terminal moves, the higher the frequency of distortion caused by burst errors becomes, and the more the visual quality deteriorates.

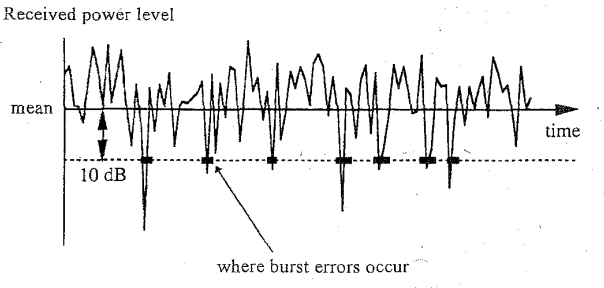


Fig. 1. Variations in received wave strength due to Rayleigh fading [3].

In this paper, a perceptually tuned 3-D subband video codec is presented, which can maintain acceptable visual quality of the low bit-rate video signals transmitted over error-prone wireless channels. The video coder is designed to remove both statistical and perceptual redundancies. The perceptual redundancy is quantified by the JND profile from a perceptual model that exploits both the spatial and temporal masking effects of the HVS. In the next section, basic characteristics of human visual perception are reviewed. The perceptual model for estimating the spatio-temporal JND profiles from video signals is introduced. The proposed subband video codec is described in Section III, where the sensitivity of the HVS to spatio-temporal frequencies is first reviewed. A distortion allocation algorithm for locating perceptually important signals in high-frequency subbands and for determining the associated quantizer stepsize is introduced. In Section IV, an UEP scheme that allocates check bits to source code segments of different perceptual importance is presented. An error concealment method applied to signals of the lowest subband is also described. Section V presents the results of computer simulation, where the performance of the proposed codec and the visual quality of the reconstructed sequence are investigated under different source and channel coding schemes. Conclusions are finally drawn in Section VI.

II. Measurement of Perceptual Redundancy

A. Basic Characteristics of Human Visual Perception

To support visual communications that demand high video quality at low bit rates, an efficient source coding algorithm is expected. Perceptual coding is one of the most promising solutions to attain such an effective video coder. An optimal perceptual coder should pay no bit in representing the perceptually redundant part of video signals. The perceptual redundancy inherent in video signals is basically due to the inconsistency in sensitivity of the HVS to stimuli of varying levels of contrast and luminance changes in the spatial and temporal domain. Masking refers to the perceptibility of one signal in the presence of another signal in its spatial, temporal or spectral vicinity [6], [28]. The phenomenon of distortion masking has been widely utilized to quantify the perceptual

redundancy as the visibility threshold of noises [28]–[31]. As considering grayscale images in the spatial domain, the primary factors affecting error visibility threshold of each pixel are the average background luminance behind the pixel and the spatial nonuniformity of background luminance. It has been found that a large portion of temporal redundancy is due to motion related blurring and resolution reduction [6], [28], [32]–[35]. Since video signals are displayed at a high frame rate, human eyes can hardly perceive the discontinuity of luminance on the time axis. Nevertheless, the sensitivity of human eyes in the temporal domain depends on the activity of luminance changes in the spatial domain. That is, the image activity of higher spatial frequency is less visible than that of lower spatial frequencies at high frame rates.

It was found that human visual perception is sensitive to luminance contrast rather than absolute luminance values [6], [28]. The ability of human eyes to detect the magnitude difference between an object and its background is dependent upon the average value of background luminance. According to Weber'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' luminance (Weber fraction) is approximately 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 [28]. In general, high visibility thresholds will occur in either very dark or very bright regions, and lower thresholds will occur in regions of grey levels close to the midgrey luminance, 127 [17], [18], [28].

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 masking to improve coding efficiency [29]–[31], [36]. 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 [28], [29]. Fig. 2 shows one of the experimental results that relates luminance difference threshold to luminance edge height [28]. In the experiment, both a perturbed edge and an unperturbed edge are shown to subjects. The perturbation is varied until it become just perceptible. In [29], a masking function of spatial activity is calculated at each pixel as the weighted sum of horizontal and vertical luminance gradients at neighboring pixels, where the weight decreases as the distance between the neighboring pixel and the central pixel increases. The approach of Musmann and Erdmann [37] can be considered as a simplified version of the above masking model, which keeps the quantization error below a visibility threshold. 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.

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 [6], [28]. Many researches have attempted to evaluate the losses of spatial resolution and magnitude resolution as an object moves

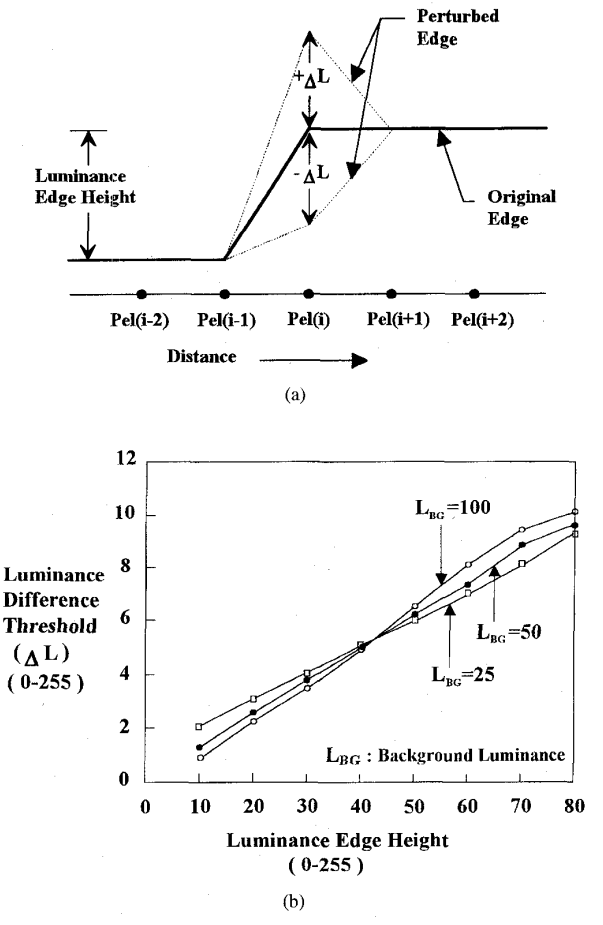


Fig. 2. (a) Stimuli for determining the visibility thresholds due to luminance changes in the spatial domain; (b) element visibility thresholds at different background luminance levels [28].

in a scene [33]–[35]. If movement is drastic, such as scene change, the perceived spatial resolution and intensity resolution are 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].

Many psychovisual studies have shown that the human perception of distortion depends on its frequency distribution [33], [34], [38], [39]. The response of the HVS to sine-wave 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 [38]–[40]. The spatial-domain CSF has been widely used to improve the quality of the coded still images [40], [41]. 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 [34]. 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.

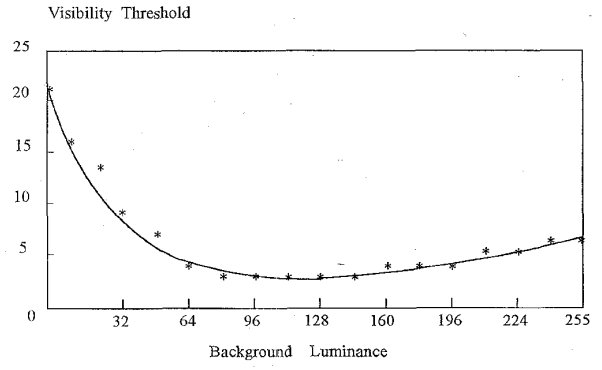


Fig. 3. Error visibility thresholds due to background luminance in the spatial domain [19].

B. Estimation of Spatio-Temporal JND Profiles

Based on the knowledge that spatial resolution and intensity resolution are reduced in moving objects and that scene motion involves the changes in luminance, the amount of temporal perceptual redundancy can be modeled as a function of interframe luminance difference and the amount of spatial perceptual redundancy. The perceptual redundancy inherent in the spatial domain has been quantitatively measured as a 2-D JND profile by a perceptual model that incorporates the visibility thresholds due to average background luminance and texture masking [19], [20]. The model assumes that the JND value of each pixel in the spatial domain is determined by the dominant of two masking effects

$$\mathrm{JND}_S(x,y) = \max\{f_1(mg(x,y)), f_2(bg(x,y))\}, \quad \text{for} \quad 0 \le x \le H, 0 \le y \le W \tag{2}$$

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. The linear relationship between visibility thresholds and luminance differences is approximated by the following equation [19], [28]

$$f_1(mg(x,y)) = mg(x,y) \cdot \beta, \quad \text{for} \quad 0 \le x \le H, 0 \le y \le W \tag{3}$$

where $mg(x,y)$ denotes the maximal weighted average of luminance gradients arround the pixel at (x,y). The parameter β is the slope of the linear function. The relationship between visibility threshold and the average background luminance, $bg(x,y)$, is given by $f_2(x,y)$ (Fig. 3). It is derived experimentally from a subjective test where a small square area (32×32 pixels) is located in the center of a large region of constant gray level [19]. For each possible gray level of the region, the noises of fixed amplitude are randomly added to or subtracted from the pixels within the square area. Through varying the amplitude of the noise, the error visibility threshold for each gray level of the background luminance is determined when the square region contaminated by the noises is just noticeable. Both of the above experiments were conducted in a dark room with a 19″ SPARC-workstation monitor, against which a viewing distance of approximately six times the image height is taken. Under these test conditions, the parameters β is found to be 2/17.

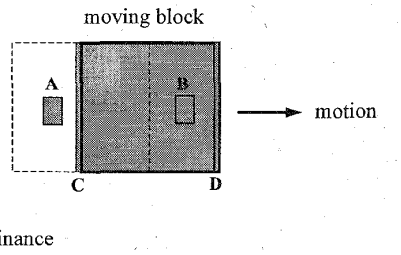


Fig. 4. The test sequence for measuring the error visibility threshold due to spatial and temporal masking.

To quantitatively measure the temporal perceptual redundancy inherent in video signals, the experiments based on the JND profile in the spatial domain and the interframe luminance difference are conducted. The experiments are conducted with a test sequence of 30 frames/s shown in Fig. 4, where a square block of a specified luminance moves horizontally across the screen of a different luminance. To measure the error visibility threshold due to interframe luminance difference and background luminance, the noises of a given amplitude are randomly added to or subtracted from a small region "A" in the uncovered area and the region "B" in the covered area. To measure the error visibility threshold due to interframe luminance difference and the luminance gradient in the spatial domain, the noises of a given amplitude are randomly added to or subtracted from the pixels within the four-pixel-wide square areas "C" and "D" adjacent to vertical edges of the moving block. For all possible values of the block's luminance and background luminance, the error visibility threshold is measured as the scaled amplitude of the spatial JND value by which the noise is just perceptible or imperceptible in the video sequence. Suppose the luminance of the moving block is larger than the background luminance. The error visibility threshold obtained in region "C" is due to a high-to-low temporal luminance change and a spatial luminance gradient. In the mean time, the error visibility threshold due to a low-to-high temporal luminance change and the same luminance gradient can be obtained in region "D." For simplicity, the scale factors obtained under the same interframe luminance difference are averaged to give a function $f_3(ild(x,y,n))$ which is used with the spatial JND value to obtain the error visibility threshold in the spatio-temporal domain. That is

$$\mathrm{JND}_{S-T}(x,y,n) = f_3(ild(x,y,n)) \cdot \mathrm{JND}_S(x,y,n) \tag{4}$$

where $ild(x,y,n)$ denotes the average interframe luminance difference between the nth and $(n-1)$th frame.

$$ild(x,y,n) = \frac{p(x,y,n) - p(x,y,n-1) + bg(x,y,n) - bg(x,y,n-1)}{2}. \tag{5}$$

The empirical results of f_3 for all possible interframe luminance differences are shown in Fig. 5. To purposely minimize the allowable distortion in the nonmoving area, the scale factor is reduced to 0.8 as $|ild(x,y,n)| < 5$. It can be noted that

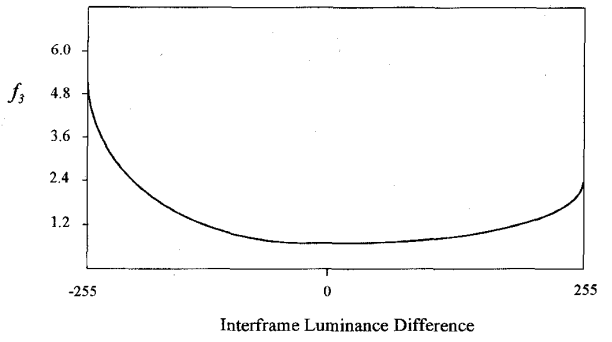
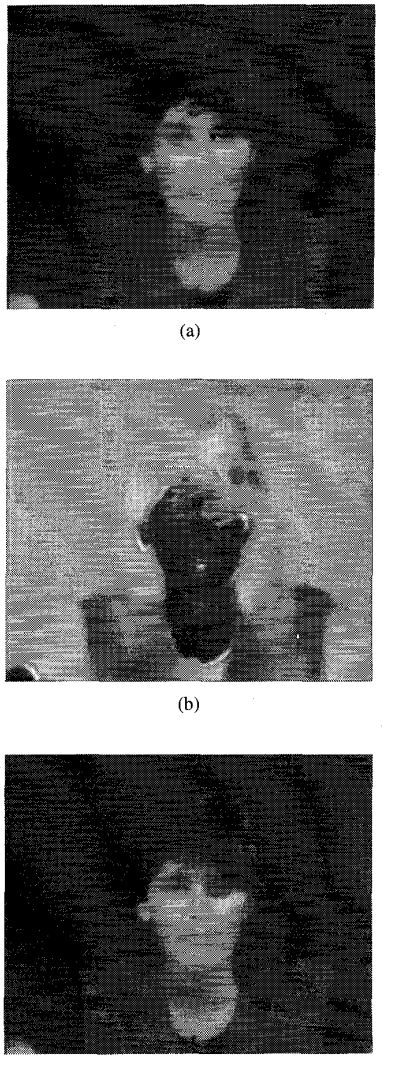


Fig. 5. Error visibility threshold in the spatio-temporal domain, which is modeled as a scale factor function of interframe luminance difference and the JND value in the spatial domain.

the error visibility threshold is increased with the increasing interframe luminance difference. This conforms to the previous research results that, after the rapid scene change or large temporal difference, the sensitivity of the HVS to spatial details is lowered. 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.

For an effective perceptual model, the subject should not be able to discern the difference between the original sequence and its JND-contaminated version at the lowest possible PSNR values. To justify the validity of the proposed perceptual model, a subjective test that compares the perceptual quality of the sequence contaminated by the noises of JND profile with the original sequence is conducted. The contaminated video frames are obtained by randomly adding to or subtracting from each pixel with the associated JND value. This is based on the assumption that quantization errors are randomly positive and negative. The frame #16 of a test sequence is shown in Fig. 6, where the associated spatio-temporal JND profile and its contaminated version are also shown. In the JND profile, the maximal value, 19, is represented by the brightest pixel while the minimal value, two, is represented by the darkest pixel. In this case, the PSNR of the contaminated frame is only 30.3 dB.



(a) (b) (c)

Fig. 6. (a) Frame #16 of a QCIF sequence "Miss America"; (b) the spatio-temporal JND profile of the frame which is used as the noises to contaminate the original frame. (the maximal value, 19, in the profile is represented by the brightest pixel and the minimal value, 2, by the darkest pixel); (c) the JND-contaminated version of the original frame.

III. A Perceptually Tuned 3-D Subband Video Coder

The functional block diagram of the proposed video coder is shown in Fig. 7, where the whole system is divided into four parts. The first part is the 3-D subband analysis that decomposes the input video sequence into a number of spatio-temporal frequency subbands. The second part is the prescribed perceptual model that estimates the spatio-temporal JND/MND profile from the analysis of local video signals and a distortion allocation algorithm that determines the component JND/MND profile of each frequency subband. The third part is the source coding algorithms that encode the signals of the lowest subband with error concealment and locate and encode the PI signals in high frequency subbands. The fourth part is a perceptually tuned channel coder that protects source code bits with BCH codes of different protection capabilities.

A. 3-D Subband Analysis

In the proposed coder, each pair of video frames are decomposed into 11 spatio-temporal subbands (Fig. 8) by a two-tap Haar filter for two-band temporal frequency analysis and the cascaded 16-tap QMF filterbanks [42] for spatial frequency analysis. The two-tap Haar filter is chosen for temporal filtering because of its acceptable frequency response, linear phase, and the merits of low delay and storage requirements. After the spatio-temporal subband analysis, more than 90% of total signal energy will be concentrated in the lowest frequency subband. The signal energy of high frequency subbands is also concentrated in lower spatio-temporal frequencies. In order to minimize the source bit rate, only subbands that possess a prespecified amount of PI signal energy are submitted to the coding process.

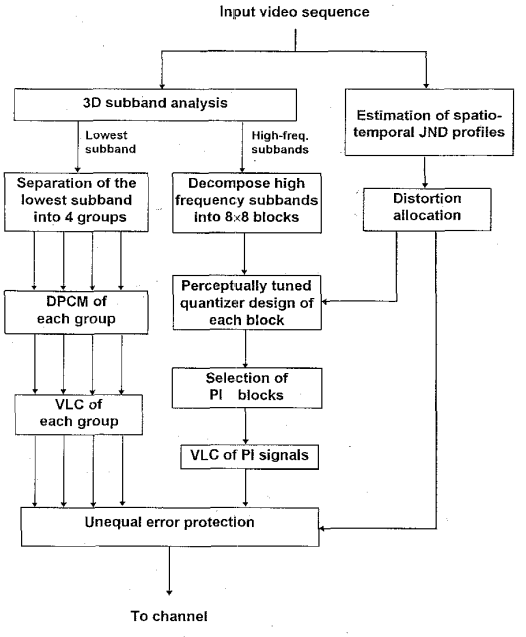


Fig. 7. Functional block diagram of the proposed 3-D subband coder.

B. Distortion Allocation

According to the bandpass characteristics of the spatio-temporal CSF (Fig. 9), the distortion of high spatio-temporal frequencies requires more energy to be perceptible than that of mid-range frequencies. The allocation of distortion energy to each subband is thus dependent upon the relative sensitivity of the HVS to the spatio-temporal frequencies of the subband. In this paper, the spatio-temporal CSF presented by Kelly [34] is used for distortion allocation. The empirical surface of this 3-D CSF is approximated by

$$C(\alpha,\mu) = (6.1 + 7.3|\log(\mu/3)|^3)\mu\alpha^2 \cdot \exp(-2\alpha(\mu+2)/45.9) \tag{6}$$

where

$$\alpha = 2f_s \quad \text{and} \quad \mu = 2\pi f_t/\alpha$$

with f_s being the spatial frequency in cycles per degree (cpd) and f_t being the temporal frequency in cycles per second (cps). μ, in this case, can be interpreted as the velocity of the moving object. As m video frames of size $H \times W$ are submitted to spatio-temporal filtering, the spatial and temporal frequencies can be derived as

$$f_s(u,v) = \sqrt{\left(\frac{27u}{H}\right)^2 + \left(\frac{27v}{W}\right)^2}, \quad \text{for} \quad 0 \le u \le H-1, \quad 0 \le v \le W-1. \tag{7}$$

$$f_t(k) = \frac{k}{2m} \cdot f_r, \quad \text{for} \quad 0 \le k \le m-1 \tag{8}$$

where f_r is the frame rate. The calculation of isotropic spatial frequency $f_s(u,v)$ is based on the viewing distance of six times the image height of 512 pixels. At this viewing distance, 54 pixels subtend approximately one degree of vision. The CSF in (6) can then be represented as $C(u,v,k)$ for $0 \le u \le H-1, 0 \le v \le W-1$ and $0 \le k \le m-1$.

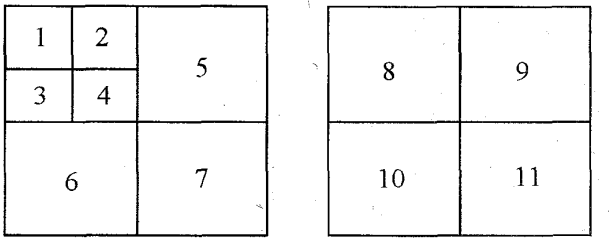


Fig. 8. Spatio-temporal frequency decomposition of two frames into 11 subbands.

To locate PI signals in each subband, the JND/MND profile of each subband must be acquired to assess the perceptual importance of each target signal. To find the JND/MND profile of each subband, the weighting function for distributing the full-band JND/MND energy to a subband can be derived as the relative desensitivity of the HVS to that frequency subband. For 11 spatio-temporal subbands, the weighting function of the qth subband is obtained as

$$w_q = \frac{S_q^{-1}}{\sum_{i=1}^{11} S_i^{-1}}, \quad \text{for} \quad 1 \le q \le 11 \tag{9}$$

where S_q represents the average sensitivity of the HVS to the qth subband. For $1 \le q \le 4$, S_q are calculated by

$$S_q = \frac{16}{H \cdot W} \sum_{u=i\cdot(H/4)}^{(i+1)\cdot(H/4)-1} \sum_{v=j\cdot(W/4)}^{(j+1)\cdot(W/4)-1} C(u,v,0),$$
$$i = \left\lfloor \frac{q-1}{2} \right\rfloor, \qquad j = (q+1) \bmod 2 \tag{10}$$

and, for $5 \le q \le 11$, S_q are calculated by

$$S_q = \frac{4}{H \cdot W} \sum_{u=i\cdot(H/4)}^{(i+1)\cdot(H/4)-1} \sum_{v=j\cdot(W/4)}^{(j+1)\cdot(W/4)-1} C(u,v,k),$$
$$k = \left\lfloor \frac{q}{8} \right\rfloor,$$
$$i = \left\lfloor \frac{q-4k-4}{2} \right\rfloor, \qquad j = (q-4k-2) \bmod 2. \tag{11}$$

The relationship between the full-band JND/MND profile and the component JND/MND profiles can then be obtained by the following equations:

$$\mathrm{JND}_q(x,y) = \left[\left[\sum_{i=0}^{3} \sum_{j=0}^{3} \sum_{t=0}^{1} \mathrm{JND}_{S-T}^2 \cdot (i + x\cdot 4, j + y\cdot 4, t) \right] \cdot w_q \right]^{1/2}, \quad \text{for} \quad 1 \le q \le 4, 0 \le x < W/4, \quad \text{and} \quad 0 \le y < H/4 \tag{12}$$

and

$$\mathrm{JND}_q(x,y) = \left[\left[\sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{t=0}^{1} \mathrm{JND}_{S-T}^2 \right.$$

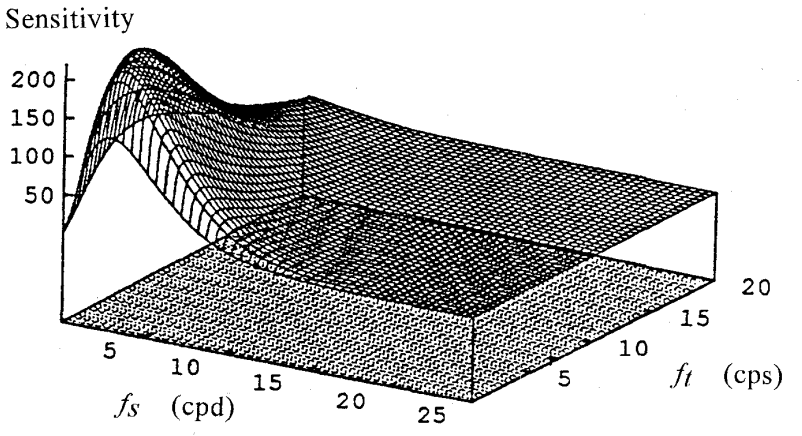


Fig. 9. The surface of a spatio-temporal CSF.

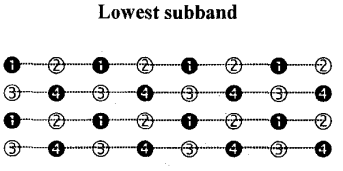


Fig. 10. The sampling pattern to separate the pixels in the lowest subband into four groups.

$$\cdot\ (i+x\cdot 2, j+y\cdot 2, t)\Big]\cdot w_q\Big]^{1/2},$$
$$\text{for}\quad 5\le q\le 11, 0\le x< W/2,$$
$$\text{and}\quad 0\le y< H/2 \tag{13}$$

where $\mathrm{JND}_q(x,y)$ denotes the JND value at location (x,y) of the qth subband.

C. Coding and Error Concealment of the Lowest Subband

The information contained in the lowest subband is the dominant part of video signals. The quality of the reconstructed lowest subband substantially affects the overall quality of the decoded video sequence. Therefore, the error-protection of the lowest subband's source code bitstream should be in the highest degree. However, the transmitted source bits can still be contaminated by the burst errors in the highly noisy channel, thus resulting in disastrous distortion in the reconstructed sequence. To solve this problem, the signals in the lowest subband are separated into four groups by subsampling the signals interleavingly in a pattern shown in Fig. 10. The sampled data in each group are then 1-D differential pulse code modulation (DPCM) encoded with a correlation coefficient of 0.85. That is, for a signal $s(x,y)$ in one of the four groups, it is predicted as $\hat{s}(x,y)=0.85\cdot s(x,y-2)$. The prediction error is variable-length encoded without losses. The source code bitstreams of four groups are error-protected and transmitted independently. Each bitstream will be properly segmented to receive error protection of different degrees. When some segments of a group's bitstream are detected to be erroneous and discarded in the decoder, the contaminated part can be replaced by linear interpolation from the neighboring decoded data in other groups.

D. Coding of High Spatio-Temporal Frequency Subbands

High spatio-temporal frequency subbands contain information of contour and motion. To encode signals in high frequency subbands, each subband is first partitioned into nonoverlapping 8×8 blocks. For each block, the stepsize or central dead-zone (twice the stepsize) of a uniform quantizer is determined by making the sum of truncation error energy and quantization error energy be equal to the block JND (or MND) energy. That is, for the block (f,g) at the qth subband, the associated quantizer stepsize $\tau_q(f,g)$ or the central dead-zone $(-\tau_q,\tau_q)$ is so determined such that

$$\sum_{i=0}^{7}\sum_{j=0}^{7}\mathrm{JND}_q^2(8f+i,8g+j)\cong\sum\sum_{(u,v)\in\Gamma_q(f,g)}p_q^2(8f+u,8g+v)+\sum\sum_{(s,t)\in\Re-\Gamma_q(f,g)}QE_q(f,g,s,t) \tag{14}$$

where

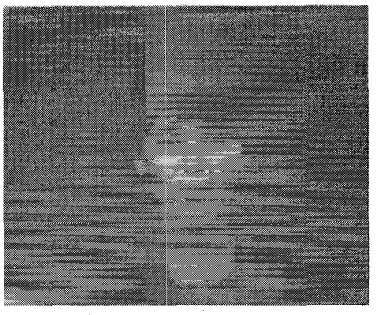
$$\Re=\{(u,v)\colon 0\le u\le 7, 0\le v\le 7\} \tag{15}$$

and

$$\Gamma_q(f,g)=\{(u,v)\colon |p_q(8f+u,8g+v)|<\tau_q(f,g),(u,v)\in\Re\} \tag{16}$$

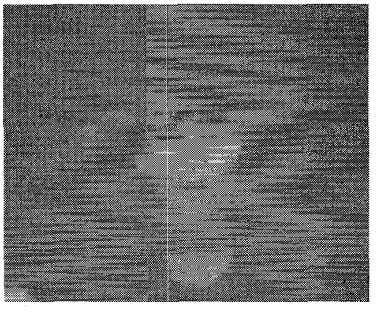
is the set of positions where signals have an amplitude less than the dead-zone threshold. The error energy resulting from quantizing the signal at (s,t) is calculated as

$$QE_q(f,g,s,t)=\left\{p_q(8f+s,8g+t)-\left(\left\lfloor\frac{p_q(8f+s,8g+t)}{\tau_q(f,g)}\right\rfloor+sgn(p_q(8f+s,8g+t))\cdot 0.5\right)\cdot\tau_q(f,g)\right\}^2. \tag{17}$$

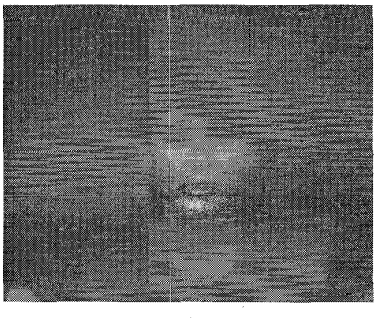


(a)

(b)



(c)



(d)

Fig. 11. (a) Frame #7 and (b) frame #10 of the test sequence; the reconstructed (c) frame #7 and (d) frame #10 which are encoded by the proposed coder with JND profiles and transmitted without error corruption.

The PI signal, in this case, is defined as the signal that has an amplitude larger than the dead-zone threshold or the quantizer stepsize. The first term in the right-hand side of (14) is therefore the overall error energy due to the truncation of perceptually unimportant (PU) signals in the block. The second term in the right-hand side of (14) is the overall quantization error energy of PI signals in the block. Depending upon the number of PI signals in a block, all blocks can be categorized into either perceptually important or unimportant. Only perceptually important blocks in which the number of PI signals is larger than a prespecified threshold (e.g., 12, used in the simulation) are encoded. The threshold can be adjusted depending on the block size and the allocated channel capacity.

TABLE I

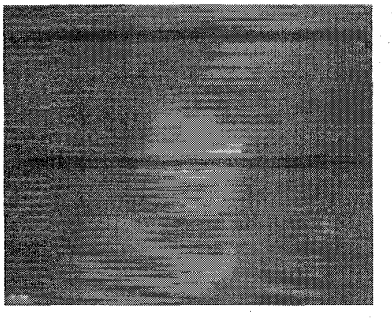
LISTING OF THE BCH CODES OF LENGTH 255; t IS THE NUMBER OF BIT ERRORS THAT CAN BE CORRECTED AND $p_\varepsilon(t)$ IS THE BIT ERROR PROBABILITY

t	BCH code (n,k)	$\varepsilon=10^{-2}$ $p_\varepsilon(t)$	$\varepsilon=10^{-3}$ $p_\varepsilon(t)$	percentage check bits
0	(255,255)	1.00×10^{-2}	1.00×10^{-3}	0.00%
1	(255,247)	9.11×10^{-3}	3.32×10^{-4}	3.14%
2	(255,239)	9.43×10^{-3}	4.50×10^{-5}	6.27%
3	(255,231)	7.64×10^{-3}	3.90×10^{-6}	9.41%
4	(255,223)	4.30×10^{-3}	2.49×10^{-7}	12.55%
5	(255,215)	2.01×10^{-3}	1.26×10^{-8}	15.69%
6	(255,207)	7.95×10^{-4}	5.27×10^{-10}	18.82%
7	(255,199)	2.73×10^{-4}	1.88×10^{-11}	21.96%

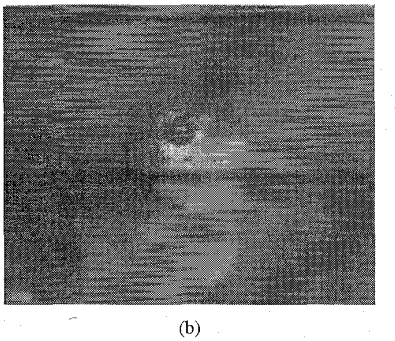
TABLE II

THE AVERAGE SOURCE BIT RATE REQUIRED FOR EACH SUBBAND WHEN JND PROFILES ARE USED IN SOURCE CODING

Subband Number	1	2	3	4	5	6	8	Total Bits
Source Code Bits (/Second)	34812	4236	4475	1743	3156	4378	2903	55702



(a)



(b)

Fig. 12. The reconstructed (a) frame #7 and (b) frame #10 which are encoded by the proposed coder with JND profiles and without error concealment. The source bitstream is equal-error protected and transmitted through the channel with BER $= 10^{-2}$.

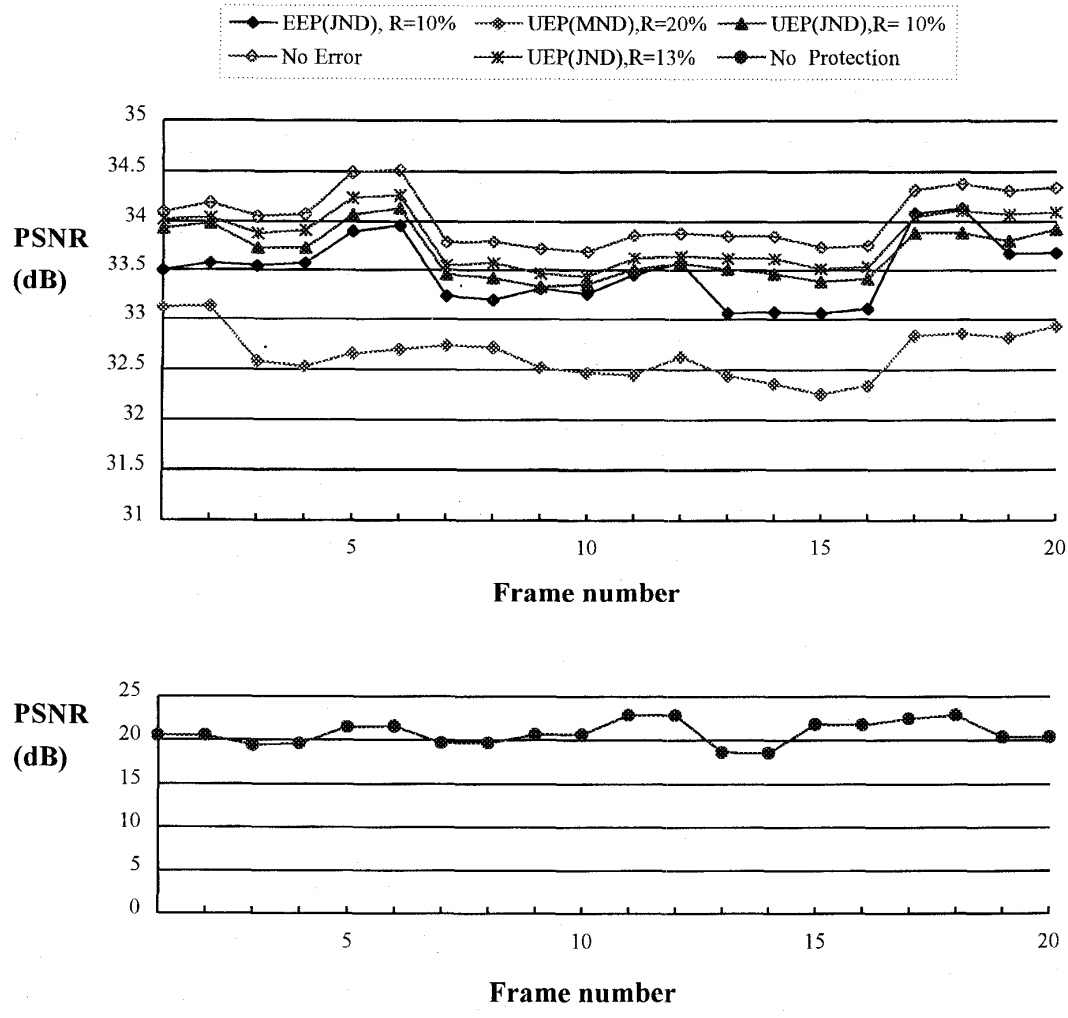


Fig. 13. Plots of PSNR for different coding and protection schemes. (EEP: equal error protection; UEP: unequal error protection; R: percentage of redundant bits for error protection; MND: frames encoded particularly with MND profiles; No Error: frames transmitted without error corruption; No Protection: frames encoded with JND profiles and without error concealment, and transmitted without error protection).

In most cases, the distortion caused by leaving out few isolated PI signals does not affect the perceptual quality [19]. The PI signals in each PI block are then located by raster scanning and quantized by the associated quantizer. The runs between consecutive PI signals and quantization indices are variable-length encoded. As will be described in the next section, the bitstream of coded data will be divided into segments of different lengths for error protection of different degrees. Either complete or partial bitstream of a block is contained in a segment, the information of the block's location and quantizer stepsize should be included in the segment and transmitted as part of the bitstream.

IV. Unequal Error Protection

For an effective unequal error protection scheme, total check bits should be reasonably distributed over the source code bits in a manner such that the perceptible distortion due to channel errors can be minimized. This is a bit allocation problem similar to that in source coding. In this section, a perceptually tuned UEP scheme is presented, where the allocation of check bits is divided into two stages. In the first stage, total check bits are allocated to each frequency subband according to its allowable distortion energy and the number of source code bits. In the second stage, different BCH codes of the same length are applied to protect segments of source code bits of different perceptual importance. Suppose the total number of check bits is R_t bits, and the number of the qth subband's source code bits is C_q bits for $1 \le q \le 11$. Then the number of check bits allocated to the qth subband can be derived as

$$R_q = R_t \cdot \frac{\left(\frac{JE_q}{C_q}\right)^{-1}}{\sum_{i \in \Theta} \left(\frac{JE_i}{C_i}\right)^{-1}}, \quad \text{for} \quad q \in \Theta \tag{18}$$

where $\Theta = \{\gamma: C_\gamma \neq 0\}$ and JE_q denotes the JND energy of the qth subband. Since each group of the lowest subband has approximately the same perceptual significance and produces

TABLE III
THE AVERAGE CHECK BIT RATE ALLOCATED TO EACH SUBBAND WHEN 10% OF THE CHANNEL RATE (61.89 kbps, FOR THE CODING WITH JND PROFILES) IS USED FOR UEP, AND THE AVERAGE BIT ERROR PROBABILITIES AFTER THE UEP

Subband Number	1	2	3	4	5	6	8
Allocated Check Bits (bits/sec)	4948	406	430	249	44	51	60
Bit Error Prob. for BER=10^{-2}	4.5×10^{-3}	8.0×10^{-3}	7.9×10^{-3}	4.7×10^{-3}	9.8×10^{-3}	9.8×10^{-3}	9.7×10^{-3}
Bit Error Prob. for BER=10^{-3}	4.4×10^{-7}	1.3×10^{-5}	1.3×10^{-5}	7.1×10^{-7}	7.4×10^{-4}	7.6×10^{-4}	6.1×10^{-4}

TABLE IV
THE AVERAGE CHECK BIT RATE ALLOCATED TO EACH SUBBAND WHEN 10% OF THE CHANNEL RATE (44.72 kbps, FOR THE CODING WITH MND PROFILES) IS USED FOR UEP, AND THE AVERAGE BIT ERROR PROBABILITIES AFTER THE UEP

Subband Number	1	2	3	4	5	6	8
Allocated Check Bits (bits/sec)	7144	328	360	180	14	60	44
Bit Error Prob. for BER=10^{-2}	1.5×10^{-3}	4.5×10^{-4}	4.2×10^{-4}	2.6×10^{-4}	9.4×10^{-3}	9.4×10^{-2}	9.2×10^{-3}
Bit Error Prob. for BER=10^{-3}	7.2×10^{-9}	2.1×10^{-17}	1.8×10^{-17}	1.6×10^{-17}	3.3×10^{-4}	3.2×10^{-4}	1.4×10^{-4}

almost the same number of source code bits, the number of check bits allocated to each group is $R_1/4$. For each group of the lowest subband and each high-frequency subband, the source code bitstream is partitioned into segments of different lengths and protected by the check bits allocated in the second stage. To assign a t-error correcting (n, k) BCH code to protect k information bits, at least $d \cdot t$ redundant bits are required to give a total code length of $2^d - 1$ bits for an integer $d \geq 3$ and $t < 2^{d-1}$ [26]. Specifically, as t error bits can be corrected by a BCH code of length n, the probability of bit error can be lowered down to $P_\varepsilon(t)$ approximately by

$$p_\varepsilon(t) \approx \frac{1}{n} \sum_{i=t+1}^{n} (i+t) \binom{n}{i} \varepsilon^i (1-\varepsilon)^{n-\varepsilon}. \tag{19}$$

Some BCH codes of length 255 are shown in Table I. Since BCH codes of fixed length are to be used, the perceptually important source code bits should be contained in short segments and protected with more check bits. For each source code bitstream, the number of the allocated check bits, R, is first inspected to determine if it is large enough to equal-error protect the bitstream of length C by a BCH code with a largest possible number of check bits. That is to test if the following requirement can be met with a positive integer ρ and a (n, B_i) BCH code from Table I

$$C \leq \rho \cdot B_i \quad \text{and} \quad R \geq \rho \cdot B_c \tag{20}$$

where B_i is the bit length of the code segment to be protected, B_c is the number of check bits, and $n = B_i + B_c$. If no integer is found to meet the requirement with the (255, 247) BCH code, then only $\lfloor R/8 \rfloor$ code segments with the lowest JND/MND energies are protected by the (255, 247) BCH code. As an (n, B_i) BCH code with a largest possible value of B_c meets the requirement with an integer value of ρ, a finer protection can be obtained by using an additional BCH code of higher error protection capability. In this case, the code segments with lower JND/MND energy can then be shortened to receive more check bits and thus giving a higher protection to visually sensitive part of the signal. Specifically, with the use of (n, B_i) and $(n, B_i - 8)$ BCH codes, two positive integers $\boldsymbol{a}$ and $\boldsymbol{b} \geq 0$ can be found to meet the following requirement

$$C \leq \boldsymbol{a} \cdot B_i + \boldsymbol{b} \cdot (B_i - d) \quad \text{and} \quad R \approx \boldsymbol{a} \cdot B_c + \boldsymbol{b} \cdot (B_c + d) \tag{21}$$

such that C source code bits can be protected by $\boldsymbol{a}$ (n, B_i) BCH codes and $\boldsymbol{b}$ $(n, B_i - 8)$ BCH codes with approximately R check bits. In so doing, the average bit error probability can be lower than that obtained by equal code segmentation and protection. Given a fixed channel rate, the average bit rate of source codes can be kept close to a specified percentage of the channel rate by adjusting the average level of the allowable distortion profile. The difference between the source bit rate and the channel rate is then used for error protection. For transmitting each pair of video frames, more check bits will be available for protecting the source bitstream if the number of source code bits decreases.

V. SIMULATION RESULTS

The simulation is performed to encode the QCIF luminance sequence of "Miss America" at a rate of 15 frames per second. The simulation results in terms of PSNR, peak signal-to-perceptible-noise ratio (PSPNR) and the reconstructed visual quality are shown, which are obtained by applying the proposed coding scheme with/without error concealment and with equal or unequal error protection. The PSPNR is defined to evaluate the visual quality in terms of the perceptible distortion energy. Only the part of distortion that exceeds the JND profile is taken into account. That is the equation shown below in (22), and

$$\delta(x, y, \kappa) = \begin{cases} 1, & \text{if } |p(x, y, \kappa) - \hat{p}(x, y, \kappa)| \geq \text{JND}_{S-T}(x, y, \kappa) \\ 0, & \text{if } |p(x, y, \kappa) - \hat{p}(x, y, \kappa)| < \text{JND}_{S-T}(x, y, \kappa), \end{cases} \quad \text{for } 0 \leq x < H, 0 \leq y < W \tag{23}$$

where $\hat{p}(x, y, \kappa)$ denotes the reconstructed pixel at position (x, y) of the κth frame. In the simulation, the MND profile is obtained by adding a constant value of two to each element of the JND profile. The average data rates in encoding the sequence with the estimated JND profiles and MND profiles are 55.7 kbps and 40.65 kbps, respectively. In both cases, perceptually important signals in high frequency subbands are only found in the subbands of #2, #3, #4, #5, #6, and #8. The

$$\text{PSPNR} = 20 \log_{10} \frac{255}{\sqrt{E([|p(x, y, \kappa) - \hat{p}(x, y, \kappa)| - \text{JND}_{S-T}(x, y, \kappa)]^2 \cdot \delta(x, y, \kappa))}} \tag{22}$$

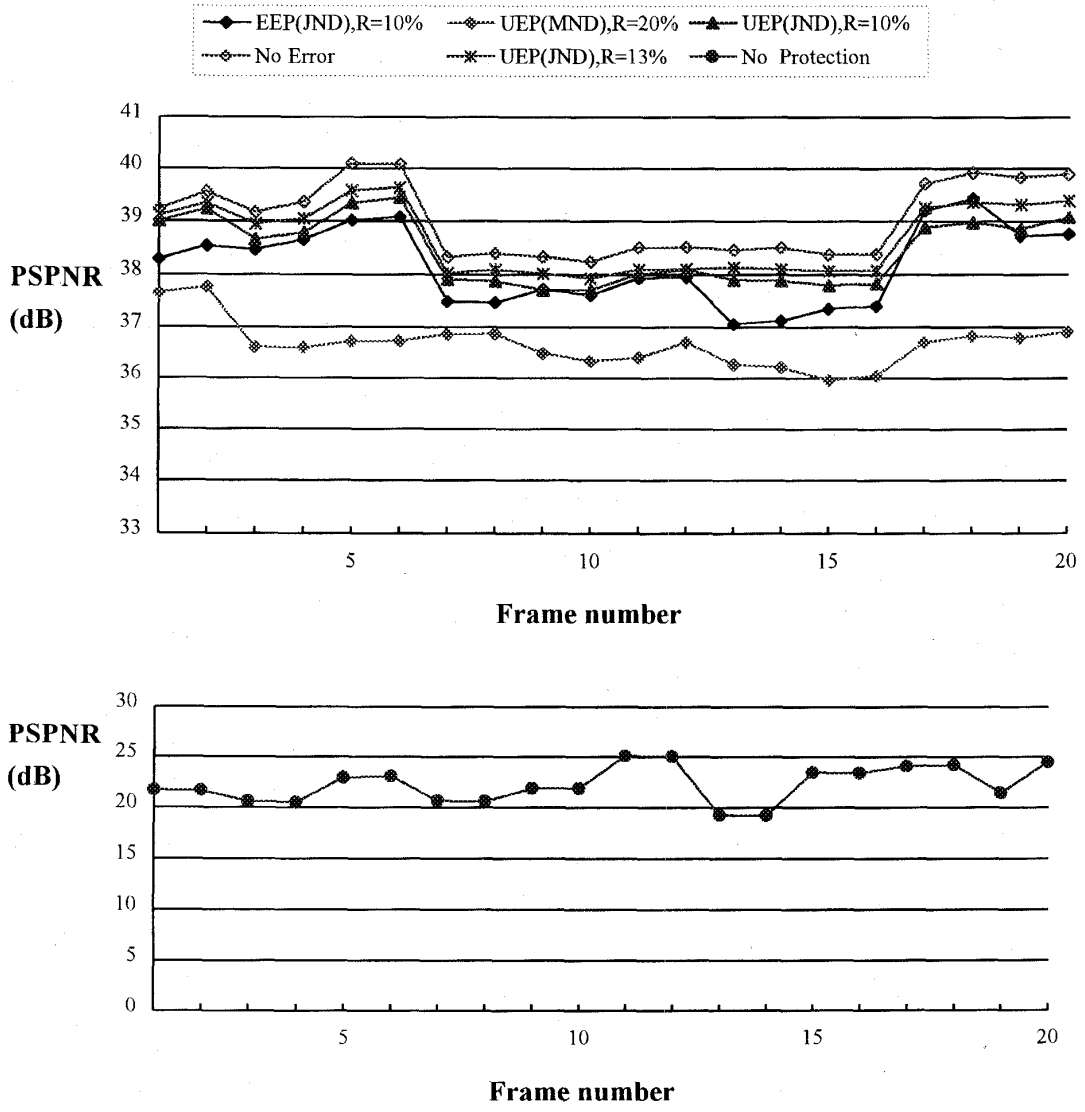
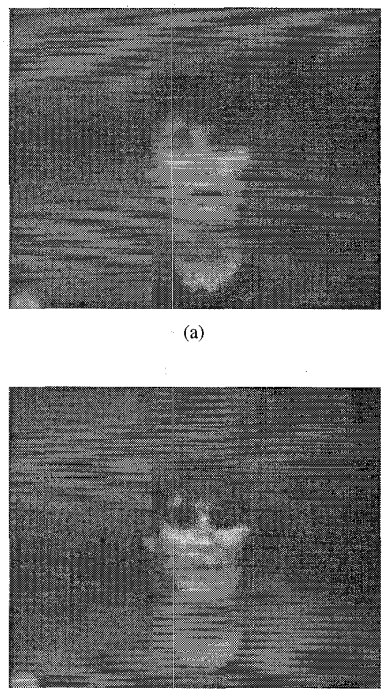


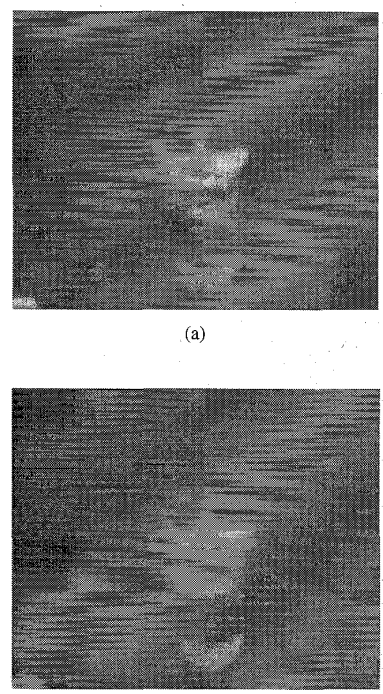
Fig. 14. Plots of PSPNR for different coding and protection schemes. (EEP: equal error protection; UEP: unequal error protection; R: percentage of redundant bits for error protection; MND: frames encoded particularly with MND profiles; No Error: frames transmitted without error corruption; No Protection: frames encoded with JND profiles and without error concealment, and transmitted without error protection).

average number of source code bits required for encoding each subband with JND profiles is shown in Table II. In Fig. 11(a) and (b), the original video frames #7 and #10 of the test sequence are shown. The reconstructed video frames which are encoded by the proposed video coder with JND profiles and transmitted without channel errors' corruption are illustrated in Fig. 11(c) and (d). In reconstructing a video sequence transmitted through a noisy channel, data are discarded if incorrectable errors are detected. The reconstructed video frames which are encoded by the proposed coder with JND profiles and without error concealment are shown in Fig. 12. The source bitstream of these encoded frames is equal-error protected and transmitted through the channel with BER $= 10^{-2}$, where the source code bits corresponding to some rows of the lowest subbands are corrupted due to quasiperiodical burst errors. Suppose a noisy channel of 64 kbps is used to transmit the sequence encoded with JND profiles, then only 13% of the channel rate can be used for error protection. If the sequence encoded with JND profiles is equal-error protected by 10% of a channel rate (61.89 kbps, including check bits), the (255 231) BCH code can be applied to protect the uniformly segmented source bitstream. The bit error probability can then be lowered to 7.64×10^{-3} for the channel BER $= 10^{-2}$ and 3.9×10^{-6} for the channel BER $= 10^{-3}$. The performance of the proposed codec with UEP for different BER's are shown in Tables III and IV. It can be noted that lower subbands receive more bits than the higher subbands. In particular, the lowest subband has the lowest error probability due to the allocation of approximately 80% of total check bits. The coding with MND profiles reduces the source code bits



(a) (b)

Fig. 15. The reconstructed (a) frame #7 and (b) frame #10 which are encoded with JND profiles and with error concealment. The source bitstream is equal-error protected and transmitted through the channel with BER = 10^{-2}.



(a) (b)

Fig. 16. The reconstructed (a) frame #7 and (b) frame #10, which are encoded with JND profiles and with error concealment. The source bitstream is unequal-error protected and transmitted through the channel with BER = 10^{-2}.

of high-frequency subbands, that substantially increases the weight for allocating more check bits to the lowest subband. The performance shown in Figs. 13 and 14 indicates that the UEP scheme outperforms equal error protection (EEP) scheme in both PSNR and PSPNR as the channel BER = 10^{-2}. The robustness of the proposed coder with UEP is improved if the number of the allocated check bits increases. The amount of perceptual redundancy removed by the proposed coder can be derived as the difference between PSPNR and PSNR. From Figs. 13 and 14, it can be found that the perceptual redundancy removed by the source coding with JND profiles is larger than that by the coding with MND profiles. It indicates that the higher the bit rate used for source coding, the more the amount of perceptual redundancy is removed. The reconstructed video frames which are encoded by the proposed coder with JND profiles and error concealment are shown in Figs. 15 and 16. The source bitstream that reconstructs the frames in Fig. 15 is equal-error protected and transmitted through the channel with BER = 10^{-2}, while the bitstream for reconstructing the frames in Fig. 16 is unequal-error protected. In Fig. 17, the reconstructed video frames are encoded with the above-mentioned MND profiles and with error concealment, the source bitstream of which is unequal-error protected and transmitted through the channel with BER = 10^{-2}. With the reconstructed video frames shown in Figs. 15, 16, and 17, the error concealment can effectively hide the distortion due to the data loss in the lowest subband. The channel errors which occurred in high frequency subbands only cause a minor difference between the UEP and EEP schemes. The reconstructed sequence encoded with MND profiles shows a uniformly degraded quality, which is primarily rebuilt by the decoded data of the lowest subband that are allocated more check bits than those encoded with JND profiles. From the simulation results, it can be found that source bit rate can be controlled by adjusting the average level of the allowable distortion profile. The higher the average level of the distortion profile is, the larger the distortion energy is allocated to each block in high frequency subbands, the larger the dead-zone or stepsize of the associated quantizer becomes, and the less the source code bits are generated. The buffer fullness control can be implemented by using the relationship between the source bit rate and the average level of the distortion profile. This distortion-rate relationship is currently under investigation.

VI. Conclusion

In this paper, a combined source-channel coding scheme is presented to maintain acceptable visual quality of the video signals which are transmitted over the channel of low bandwidth and high error rates. To encode video signals at low bit rates, the source coding efficiency is maximized by exploiting basic properties of the human visual system. The spatio-temporal JND profile is estimated to quantify the perceptual redundancy inherent in video signals. The proposed 3-D subband codec is designed to encode the perceptually important signals through the allocation of distortion energy given by the JND/MND profile. The JND/MND profile is also used to allocate check bits for unequal error protection. It

(a)

(b)

Fig. 17. The reconstructed (a) frame #7 and (b) frame #10, which are encoded with MND profiles and with error concealment. The source bitstream is unequal-error protected and transmitted through the channel with BER = 10^{-2}.

is found that the lowest subband absorbs high percentages of source code bits and check bits to maintain and protect the dominant part of visual information. Furthermore, error concealment is applied to the perceptually most sensitive signals of the lowest subband that has the lowest value in the average amplitude of the corresonding JND/MND profile. With the error concealment incorporated in encoding signals of the lowest subband, the distortion due to erroneous transmission of coded data can be effectively suppressed. It is also shown that the performance of the proposed codec with UEP is superior to that with EEP in both PSNR and PSPNR. The proposed perceptual model might not be the optimal one, but empirically found effective and reasonable in estimating the perceptual redundancy inherent in video signals. One way to further improve the coding efficiency of the proposed coder is to increase the resolution in temporal frequency subbands by including more frames in the spatio-temporal filtering. The application of spatio-temporal CSF to distortion allocation can be more effective if the number of temporal frequency subbands is increased. Nevertheless, the coding delay and the storage requirement can be the limitation. The proposed codec can be applied to the coding of color video sequences in case that the perceptual model can also measure the JND profile of chrominance signals. Actually, the intercomponent masking effect between luminance and chrominance signals has been studied and applied to the coding of color images. The perceptual model for measuring the spatio-temporal JND profiles of color video sequences is currently under intense investigation.

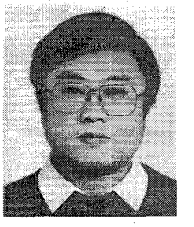
ACKNOWLEDGMENT

The authors are indebted to the anonymous reviewers whose comments helped to improve an earlier version of this paper.

REFERENCES

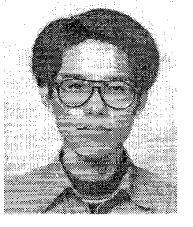
[1] J. E. Padgett, C. G. Gunther, "Overview of wireless personal communcation," *IEEE Commun. Mag.*, vol. 33, pp. 28–41, Jan. 1995.
[2] Y. Q. Zhang, "Very low bit rate video coding standards," in *Proc. of SPIE Conf. Visual Communications and Image Processing,* vol. 2501, 1995, pp. 1016–1023.
[3] H. Ibaraki, T. Fujimoto, and S. Nakano, "Mobile video communication techniques and services," in *Proc. of SPIE Conf. Visual Communications and Image Processing,* vol. 2501, 1995, pp. 1024–1033.
[4] K. Illgner and D. Lappe, "Mobile multimedia communications in a universal telecommunications network," in *Proc. of SPIE Conf. Visual Communications and Image Processing,* vol. 2501, 1995, pp. 1034–1043.
[5] T. C. Chen *et al.*, "A real-time software based end-to-end wireless visual communications simulation platform," in *Proc. of SPIE Conf. Visual Communications and Image Processing,* vol. 2501, 1995, pp. 1068–1074.
[6] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE,* vol. 81, pp. 1385–1422, Oct. 1993.
[7] B. Masnick, "On linear unequal error protection code," *IEEE Trans. Inform. Theory,* vol. IT-3, pp. 600–607, Oct. 1967.
[8] K. M. Rose, and A. Heiman, "Enhancement of one-dimensional variable-length DPCM images corrupted by transmission errors," *IEEE Trans. Commun.*, vol. 37, pp. 373–379, Apr. 1989.
[9] P. H. Westerink, J. H. Weber, D. E. Boekee, and J. W. Limpers, "Adaptive channel error protection of subband encoded images," *IEEE Trans. Commun.*, vol. 41, pp. 454–459, Mar. 1993.
[10] ISO-IEC/JTC1 SC 29/WG 11 (MPEG), "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s," *ISO Rec. 11172,* Part 2: Video.
[11] Y. Q. Chang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Trans. Circuits Syst. Video Technol.* vol. 2, pp. 285–296, Sept. 1992.
[12] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing,* vol. 3, pp. 559–571, Sept. 1994.
[13] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Processing,* vol. 3, pp. 572–588, Sept. 1994.
[14] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," presented at *Proc., ICASSP,* 1988.
[15] A. Jacquin and C. Podilchuk, "Very low bit rate 3D subband-based video coding with a dynamic bit allocation," in *Proc. of SPIE Int. Symp. Video Communications Fiber Optic Services,* vol. 1977, Apr. 1993, pp. 156–167.
[16] C. Podilchuk, N. S. Jayant, and N. Farvardin, "Three dimensional subband coding of video," *IEEE Trans. Image Processing,* vol. 4, pp. 125–139, Feb. 1995.
[17] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantization data compression," in *Proc. IEEE Int. Conf., Acoustics, Speech, Signal Processing,* vol. 3, 1989, pp. 1945–1948.
[18] C. H. Chou, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," presented at *Proc. IEEE Int. Symp., Information Theory,* 1994.
[19] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.
[20] C. H. Chou, "Adaptive transform coding based on removing just noticeable distortion," in *Proc. of SPIE Conf. Visual Communications and Image Processing,* vol. 2501, 1995, pp. 607–618.
[21] N. Jayant, "Signal compression: technology targets and research directions," *IEEE J. Select. Areas Commun.*, pp. 314–323, June 1992.
[22] X. Fan and N. Farvardin, "A perceptually motivated three-component image model—Part II: applications to image compression," *IEEE Trans. Image Processing,* vol. 4, pp. 430–447, Apr. 1995.
[23] S. Karunasekera and N. Kingsbury, "A distortion measure for blocking artifacts in image based on human visual sensitivity," *IEEE Trans. Image Processing,* vol. 4, pp. 713–724, June 1995.

[24] W. Xu *et al.*, "Subjective rating of picture coding algorithm," in *Proc. of SPIE Conf. Visual Communications and Image Processing,* vol. 2501, 1995, pp. 619–628.
[25] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.,* pp. 314–364, 1989.
[26] G. C. Clark, Jr. and J. B. Cain, *Error Correction Coding For Digital Communication.* New York: Plenum, 1981.
[27] H. Tominaga, H. Jozawa, M. Kawashima, and T. Hanamura, "A video coding method considering cell loss in ATM-based networks," *Signal Processing: Image Communication,* vol. 3, pp. 291–300, 1991.
[28] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression.* New York: Plenum, 1988.
[29] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE,* vol. 65, pp. 536–548, Apr. 1977.
[30] R. Forchheimer and T. Kromamder, "Image coding—from waveforms to animation," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 2008–2023, Dec. 1989.
[31] J. O. Limb, "On the design of quantizer for DPCM coder: a functional relationship between visibility, probability and masking," *IEEE Trans. Commun.,* vol. COM-26, pp. 573–578, 1978.
[32] D. H. Kelly, "Visual responses to time-dependent stimuli," *J. Opt. Soc. Amer.,* vol. 51, pp. 422–429, 1961.
[33] J. G. Robson, "Spatial and temporal contrast sensitivity functions of the visual system," *J. Opt. Soc. Amer.,* vol. 56, pp. 1141–1142, 1966.
[34] D. H. Kelly, "Motion and vision II. stabilized spatio-temporal surface," *J. Opt. Soc. Amer.,* vol. 69, pp. 1340–1349, Oct. 1979.
[35] R. Pease and J. O. Limb, "Exchange of spatial and temporal resolution in television coding," *Bell Syst. Tech. J.,* vol. 50, p. 191, 1971.
[36] H. G. Musmann, "Predictive image coding," in *Image Transmission Techniques,* W. K. Pratt, Ed. New York: Academic, pp. 81–97, May 1979.
[37] H. G. Musmann and W. D. Erdmann, German Patent Appl. no. P 2740945.6, 1977.
[38] J. L. Mannos and D. J. Sakrison, L. Mannos and D. J. Sakrison, "The effect of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory,* vol. IT-20, pp. 525–536, July 1974.
[39] C. F. Hall and E. L. Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-7, pp. 161–170, Mar. 1977.
[40] K. N. Ngan, K. S. Leong, and H. Singh, "Adaptive cosine transform coding of images in perceptual domain," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 1743–1749, Nov. 1989.
[41] D. L. McLaren and D. T. Nguyen, "Removal of subjective redundancy from DCT-coded images," *Proc. Inst. Elec. Eng.-I,* vol. 138, no. 5, pp. 345–350, Oct. 1991.
[42] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. 1980 ICASSP,* pp. 291–294, Apr. 1980.
[43] H. Gharavi, "Subband coding algorithms for video applications: videophone to HDTV-conferencing," *IEEE Trans. Circuits Syst. Video Technol.* vol. 1, pp. 174–183, June 1991.
[44] R. Stedman, H. Gharavi, L. Hanzo, and R. Steele, "Transmission of subband-coded images via mobile channels," *IEEE Trans. Circuits Syst. Video Technol.* vol. 3, pp. 15–26, Feb. 1993.



Chun-Hsien Chou graduated from National Taipei Institute of Technology, Taipei, Taiwan, in 1979, and received the M.S. and Ph.D. degrees in electrical engineering from National Tsing Hua University, Hsinchcu, Taiwan, in 1986 and 1990, respectively.

In 1990, he joined the Department of Electrical Engineering at Tatung Institute of Technology, Taipei, Taiwan, as an Associate Professor. During the academic year from 1991 to 1992, he was a postdoctoral research member at AT&T Bell Laboratories, Murray Hill, NJ. His current research areas include perceptual coding of image and video signals, very low bit rate video coding, and real-time VLSI architectures for multimedia communication.



Chi-Wei Chen received the B.S. and M.S. degrees in electrical engineering from Tatung Institute of Technology, Taipei, Taiwan, in 1993 and 1995, respectively.

He is currently on active service in the army. His main interests include image processing and video coding.