A TUTORIAL ON DIGITAL VIDEO COMPRESSION TECHNIQUES

Joseph B. Waltrich
Jerrold Communications
Applied Media Lab

ABSTRACT

Digital video transmission offers significant advantages to the cable operator because of its reduced susceptibility to channel impairments. However, a penalty is exacted in terms of the bandwidth required to transmit an uncompressed digital video signal. In order to achieve a reasonable bandwidth for digital video transmission, some degree of video bandwidth compression is necessary.

This paper discusses the relationship between compression requirements and transmission bandwidth and presents a review of some of the more commonly advocated video compression schemes. The applicability of video compression for delivery of both NTSC and HDTV is discussed. Implications for both cable and fiber are considered.

INTRODUCTION

Because of its reduced susceptibility to transmission channel impairments, digital video transmission offers significant advantages over analog transmission. However, a penalty is exacted in terms of bit rate and the associated bandwidth required for transmission. For example, if a composite NTSC signal is sampled at four times the color subcarrier with 8 bit quantization, then the active video portion of each line would contain 768 samples. Since there are 480 active video lines in an NTSC frame and the NTSC frame rate is 29.97 frames/second, the total video bit rate can be calculated as follows:

$$8 \times 768 \times 480 \times 29.97 = 88.4 \text{ Megabits/second} \qquad (1)$$

It should be noted that the bit rate calculated from Equation (1) is for transmission of active video only. In practice, additional data such as digital audio (500 Kbits/sec or more), synchronization data and error corection data must also be transmitted. For cable applications, additional data must be transmitted for encryption and addressabiility. If all these factors are considered, the total bit rate for digital video transmissionn could easily exceed 100 Mbits/sec.

The required bandwidth for digital transmission is a function of the total data rate and the modulation technique. The bandwidth may be calculated as follows:

$$W = R_d/E_s \qquad (2)$$

where:

W = bandwidth (Hz)
R_d = total data rate (b/s)
R_s = spectral efficiency (b/s/Hz)

Table 1 lists the spectral efficiencies for various modulation schemes. Additional information regarding the data in Table 1 may be found in Feher [1].

From Equation (2), it is seen that transmission of uncompressed video at bit rates on the order of 100 Mb/s would require a transmission bandwidth of 30 - 60 MHz. Transmission of HDTV in RGB form (approximately 1.5 Gb/s) would require a bandwidth of about .5 -1 GHz. This is obviously impractical for cable and, although it might be argued that the use of fiber optics would alleviate bandwidth constraints, the cost of electronics for fiber based systems increases with increasing bit rate. Therefore, some form of bandwidth compression must be employed in order to achieve both spectral and cost efficiencies for digital video transmission.

COMPRESSION REQUIREMENTS

Prior to discussing compression techniques, the degree of compression as a function of bandwidth should be examined. The degree of compression is best expressed in terms of the average information or entropy of a compressed video source, expressed in terms of bits/pixel. (A pixel or pel, as it is sometimes abbreviated is defined as a picture element or sample). As previously stated, the total data rate includes not only digitized video but also digital audio and a certain amount of overhead data. The total data rate, as a function of video entropy and other data, is given by:

$$R_d = (R_v H_v + R_a)(1 + OH/100) \quad (3)$$

where:

R_d = total data rate (b/s)
R_v = video data rate (pel/s)
H_v = video entropy (b/pel)
R_a = audio data rate (b/s)
OH = overhead (%)

In order to fit a signal into a given bandwidth, the amount of compression must be determined for a particular modulation technique. That is, the entropy must be calculated, given the other parameters. Using (2) and (3), the following expression may be obtained for the entropy:

$$H_v = (1/R_v)(WE_s/(1 + OH/100) - R_a) \quad (4)$$

For successive calculations, it will be assumed that digital audio is transmitted at 500 Kb/s and that a 25% overhead is required for ancillary data.

In practice, video compression is usually performed on the signal in component form. Although the signal may be in RGB format, it is more efficient to perform signal processing on the luminance and chrominance components (e.g. - Y, I, Q) since the chrominance components may be sampled at half the rate used for the luminance, thereby reducing the uncompressed data rate. (The missing samples are reconstructed by interpolation at the receiver). Three sampling schemes will be considered:

- NTSC sampled at 4 times subcarrier
- NTSC sampled at 3 times subcarrier
- HDTV in Common Image Format

For the first case, there are 768 luminance pels per line and 480 active lines per frame. The luminance data rate is therefore:

$$R_Y = 768 \times 480 \times 29.97 = 11.05 \text{ Mpel/s} \quad (5)$$

And, since each chrominance component is sampled at half the rate for luminance:

$$R_C = 384 \times 240 \times 29.97 = 2.76 \text{ Mpel/s} \quad (6)$$

Therefore:

$$R_V = R_Y + 2R_C = 16.57 \text{ Mpel/s} \quad (7)$$

For NTSC sampled at 3 times subcarrier, there are 576 luminance pels/line. Using the same sampling scheme, one obtains a value of 8.29 Mpel/s for R_Y and 2.07 Mpel/s for R_C. This yields a value of R_V = 12.43 Mpel/s.

The Common Image Format for HDTV is currently specified as 1920 pels/active line by 1080 active lines. Applying the same sampling scheme to this format yields an R_V = 93.31 Mpel/s.

Fig. 1 presents a plot of H_V as a function of bandwidth for the three above-mentioned cases using a modulation technique with a spectral efficiency of 1.66 b/s/Hz. From this plot, it is seen that, in order to transmit digital NTSC in a 6 MHz channel using 2PAM or 4QAM, an entropy of .45 - .6 b/pel is required, depending on the sampling frequency. Assuming 8 bit source quantization, this amounts to a compression ratio of 13 -18. If 4PAM or 16QAM is used, the entropy is doubled and the compression ratio is halved. Transmission of digital HDTV in 6 MHz would require an entropy of .08 - .16 b/pel, corresponding to a compression ratio of 50 - 100.

COMPRESSION TECHNIQUES

During the past decade, considerable effort has been expended on the development of a variety of digital video compression techniques. Although much of this research was spurred by non-entertainment applications of television (e.g. - teleconferencing, military applications, etc.), some of these techniques have recently found their way into commercial television. Increased interest in HDTV has also generated a corresponding interest in compression techniques.

Currently, a number of video compression techniques are being used in a variety of applications. These include the following:

- o Predictive Coding (e.g. - DPCM)
- o Transform Coding
- o Vector Quantization
- o Subband Coding

This paper will concentrate on discussion of those techniques which are currently thought to be best suited to broadcast and cable applications.

Differential Pulse Code Modulation

Differential Pulse Code Modulation (DPCM) is a technique in which the value of a given pixel is estimated, based on the values of preceding pixels. This estimate or predictor is a linear function of preceding pixel values. For a pixel having a value X_N, the general form of the predictor is:

$$\hat{X}_N = \sum_{i=0}^{i=N-1} a_i X_i \qquad (8)$$

The predicted value is then subtracted from the pixel value to generate an error signal:

$$e_N = X_N - \hat{X}_N \qquad (9)$$

The error signal is encoded and transmitted. At the receiver, the pixel value is recovered by adding the error signal to the predictor which the receiver has determined from previously recovered pixels. The equation for the recovered signal X'_N is:

$$X'_N = e_N + \hat{X}'_N \qquad (10)$$

where

$$\hat{X'}_N = \sum_{i=0}^{i=N-1} a_i X'_i \qquad (11)$$

The bit rate reduction for DPCM is due to the fact that the variance of the error signal e_N is significantly less than that of the original image and therefore the error signal lends itself well to bit rate reduction via variable run length coding.

Fig. 2 presents an example of DPCM using a simple 2-dimensional predictor (i.e. - pixels from both the present and previous lines are used to estimate the value of X_N).

A block diagram of a DPCM system is shown in Fig. 3. System complexity depends on the nature of the prediction algorithm. Predictors may be either one, two or three dimensional, requiring from one line to one frame of memory. In practice, it is advantageous to use combinations of multi-dimensional predictors which are adaptively switched so as to minimize the value of the error signal. A more detailed discussion of adaptive prediction schemes may be found in publications by Ng and Hingorani [2] and Knee [3].

A comparison of the distribution of quantization levels for an uncompressed image and its associated error signal is shown in the histograms of Fig. 4. The variance of the error signal is approximately 1/50th that of the original image. Because of this, the error signal can be encoded using a variable run length code such as Huffman coding, thereby achieving a substantial reduction in entropy. This technique yields entropies on the order of 2 -5 b/pel.

Since the value of a reconstructed pixel depends on the value of previous pixels, transmission errors in DPCM can cause streaking effects over one or more lines of the picture, depending on the nature of the prediction algorithm.

Huffman Coding

Huffman coding is a variable length coding technique which reduces the average bit rate required to represent a set of quantization levels. This is accomplished by assigning short codewords to those quantization levels which have the greatest probability of occurrence and longer codewords to less frequently occurring quantization levels. Obviously, this scheme works best on signals having relatively little spread in the distribution of quantization levels (e.g. - the DPCM error signal). An illustration of the coding technique is shown in Fig. 5. The coding procedure is accomplished by starting with the two least frequently occurring quantization levels. These are combined into one node (a7 of Fig. 5). logic 0 and 1 levels are arbitrarily assigned to the two branches of this node. This process is repeated until we are left with a single node having a probability of 1. The codeword associated with each probability is determined by tracing back through the branches, starting with the last node.

The coding scheme of Fig. 5 is, of course, an oversimplification, since, in practice, the number of quantization levels is larger than six. However, a group of codes with very low probabilities of occurrence can be assigned a single codeword. The actual value of the quantization level is then transmitted following the codeword.

Obviously, a variable length code such as Huffman coding lends itself to other compression schemes as well as DPCM.

Transform Coding

In transform coding, the image is transformed from the time domain to a different domain (e.g. - the frequency domain). The transform coefficients are then encoded and transmitted. An inverse transformation is performed at the receiver to recover the original image.

A number of transforms have been used for various video compression applications. Among these are the Karhunen-Loeve, Walsh-Hadamard, Haar, Discrete Fourier and Discrete Cosine transforms. Detailed information on all of these transforms is given by Stafford [4]. Currently, the Discrete Cosine Transform (DCT) appears to be the most widely used of the above-mentioned transforms. Therefore, this paper will limit discussion of transform coding to the DCT.

The DCT is typically performed on blocks of pixels ranging in size from 4 x 4 to 16 x 16. The transform is similar to the real part of a Fourier transform. The contents of each pixel block are converted to a series of coefficients which define the spectral content of the block. The general form of a two-dimensional DCT performed on an N x N pixel block is given by the equation:

$$Y_{mn} = (1/2N)E_mE_n\sum_{k=0}^{N-1}\cos((2k+1)n\pi/2N)$$

$$\times \sum_{j=0}^{N-1}X_{jk}\cos((2j+1)m\pi/2N) \qquad (12)$$

where:

Y_{mn} = DCT coefficients at coordinates m, n
X_{jk} = pixel amplitude at coordinates j, k
$E_m, E_n = 1/\sqrt{2}$ $(m,n = 0)$
$E_m, E_n = 1$ $(m = n \neq 0)$

The inverse transform (IDCT) is given by the equation:

$$X_{jk} = (2N/E_mE_n)\sum_{m=0}^{N-1}\cos((2j+1)m\pi/2N)$$

$$\times \sum_{n=0}^{N-1}Y_{mn}\cos((2k+1)n\pi/2N) \qquad (13)$$

Fig. 6 presents some examples of DCT's of various pixel patterns. For pixel block patterns which are typical of live video, the DCT reduces the pixel block data to only a few non-zero coefficients. Therefore, it would be expected that a bit rate reduction would result solely from the transform process. In practice, however, the presence of noise in the signal can generate spurious transform coefficients. Filtering the input signal prior to generating the transform is, therefore, necessary in order to reduce the occurrence of these coefficients.

In order to maintain a desired entropy, it is often necessary to discard some of the DCT coefficients during the coding process. This is usually not a problem since, in most cases, only a few coefficients make a major contribution to the signal's spectral content. The coefficient selection process is typically determined by one of two types of coding: zonal coding and threshold coding. Zonal coding discards all of the coefficients except those within a selected zone (which always includes the low frequency components). Threshold coding discards those coefficients whose values are below a given set of threshold levels. (In some forms of threshold coding, these coefficients are coarsely quantized rather than simply discarded).

Since, in zonal coding, the high frequency coefficients of the DCT are discarded, edge blurring in the reconstructed image sometimes occurs. This effect is less noticeable in threshold coding since coefficient selection depends on the magnitude of a particular spectral component as well as its position. However, threshold coding carries a bit rate penalty since coefficient locations are not predetermined and, therefore, coefficient address information must also be transmitted.

Unlike a DPCM encoded picture, transmission errors when using DCT coding are confined to individual pixel blocks. This would cause a "spotting effect" in the decompressed picture, the nature of which depends on which DCT coefficients were corrupted.

Commercially available DCT codecs are currently being used for transmission of broadcast quality NTSC video at 45 Mb/s [5]. Several IC manufacturers will soon be offering DCT processors in chip form. Among these are the INMOS IMSA121, SGS-Thomson STV3200 and the LSI Logic L64730. These chips are capable of operating at clock rates in the 13.5 - 40 MHz range and computing both forward and inverse transforms on 8 x 8 pixel blocks.

The DCT is capable of generating reasonably good picture quality at entropies on the order of .5 - 2 b/pel. It is possible to combine the DCT with other compression schemes such as DPCM to achieve greater bit rate reductions.

Vector Quantization

In vector quantization, an image is divided into a number of non-overlapping blocks. Each block is regarded as an N-dimensional image vector **X** where N is equal to the number of pixels in the block. Each vector is compared with a set of N_C stored reference patterns or codevectors $\mathbf{Y_1} \ldots \mathbf{Y_{Nc}}$ in order to find the codevector $\mathbf{Y_k}$ which most closely matches the image vector **X**. Once found, the index k of the codevector is transmitted to the receiver which then uses a lookup table to reproduce the codevector $\mathbf{Y_k}$.

The entropy for vector quantization is dependent on the size of the vector and the number of vectors stored in the codebook. If the codebook size N_C is a binary number which can be represented by b bits, then a vector of N pixels is represented by one of 2^b codevectors and the entropy is simply:

$$H_V = b/N = \log_2(N_C)/N \qquad (14)$$

Conversely, the codebook length N_C may be determined from the desired entropy:

$$N_C = 2^{(NH_V)} \qquad (15)$$

The codevector $\mathbf{Y_k}$ is chosen to minimize the error (commonly referred to as the distortion) between $\mathbf{Y_k}$ and the image vector **X**.

The key elements for effective vector quantization are codebook generation and codebook search. Codebook generation may be based on statistical distribution of image vectors or on the use of training images. The LBG algorithm [6] is a popular method for optimizing codebook design.

The method of searching the codebook affects the number of computations required to determine the best choice of codevector. An exhaustive search involves computation of the distortion between the image vector and every vector stored in the codebook. By using a tree-structured codebook and binary search, the number of computations is decreased significantly. However, more memory is required for storage of a tree structured codebook. Details of codebook design and search techniques may be found in Lim [7].

In vector quantization, transmission errors can have a significant effect on the reproduction of individual pixel blocks since an incorrect codevector index results in selection of the wrong codevector by the decoder. As in the DCT, however, errors are confined to individual pixel blocks. The effect of occasional errors would be a random pattern of light and dark spots in the picture.

Subband Coding

In subband coding the image is divided into frequency subbands using various combinations of horizontal and vertical filtering and subsampling techniques. One form of subband coder, in which the image is divided into octave bands, is shown in Fig. 7. Prior to subsampling, the image is divided into frequency bands by combinations of lowpass and highpass filters. Following frequency division, the image is subsampled or decimated by an integral factor M as shown in the blocks represented by the down arrows. (For division into octave bands as shown in Fig. 7, M = 2).

At the receiver, the subbands are interpolated by an integral factor L by inserting L -1 zeroes between received samples of the subband data. (For octave bands, L =2). The interpolation process is indicated by an up arrow. The interpolated subbands are then filtered and recombined to reconstruct the received image.

Image compression is obtained by coding each subband using compression schemes such as DPCM and/or by discarding some of the high frequency information or transmitting it at a lower rate. This technique has been proposed by both Schreiber [8] and Zenith [9] to compress HDTV into a 6 MHz channel.

The effects of subband coding in the frequency domain are shown in Fig. 8. If ideal filtering is assumed, the low frequency information is not affected by the decimation process. However, decimation of the high frequency information results in aliasing of the data with associated spectrum folding. The interpolation process creates spectral images which allow the original information to be recovered by filtering the imaged spectra within the desired frequency range. Detailed discussion of the effects of decimation and interpolation may be found in Crochiere and Rabiner [10].

Since it is not possible to construct ideal filters, the decimation process will produce some degree of aliasing in all bands. The filters used in subband coding must be chosen such that the synthesis filters (i.e. - the filters following the interpolators) cancel aliasing generated during the decimation process. This may be accomplished by the use of Quadrature Mirror Filters.

Quadrature Mirror Filter design considerations are described by Vaidyanathan [11].

CONCLUSIONS

If a suitable compression technique can be developed, digital video could be transmitted over existing cable systems. However, new headend and subscriber equipment would be required and different modulation schemes may require re-structuring of carriers. Fiber offers a more likely medium for digital transmission since a cost-effective compression technique would probably prove more economical than adding fibers and associated electronics to achieve increased system capacity. In addition, fiber would be free of most of the problems which could cause errors in digital transmission over cable.

Since video compression is generally performed on component signals, digital transmission offers the opportunity to eliminate NTSC artifacts via comb filter encoding and decoding.

Most compression schemes are more effective on clean signals. It is likely that, for signals such as satellite feeds, noise reduction techniques will have to be developed. If digital transmission becomes popular, compression and noise reduction techniques may spur the development of low cost frame stores for both headend and subscriber equipment.

REFERENCES

[1] K. Feher, "Telecommunications Measurements, Analysis and Instrumentation", Prentice-Hall, 1987.

[2] S. Ng and R. Hingorani, "Adaptive Spatio-Temporal DPCM for Videoconferencing", GLOBECOM 1986, Houston, pp 8.7.1-5.

[3] M. Knee, "Adaptive Prediction for High Quality Television Transmission Coding Based on the LMS Algorithm", SMPTE J., 98, 580-589, Aug. 1989

[4] R. Stafford, "Digital Television", Wiley, 1980

[5] S. Cucchi and F. Molo, "DCT Based Television Codec for DS3 Digital Transmission", SMPTE J., 99, 640-646, Sept. 1989

[6] Y. Linde, A. Buzo and R.M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Trans. on Comm., Vol. COM-28, No. 1, 84-95, Jan. 1980.

[7] J. Lim, "Two Dimensional Signal and Image Processing", Prentice Hall, 1990.

[8] W. Schreiber et al, "Channel Compatible 6 MHz HDTV Distribution Systems", SMPTE J, 98, 5-13, Jan. 1989.

[9] W. Bretl, R. Citta, R. Lee, P. Fockens, "Spectrum Compatible High Definition Television Systems", SMPTE J., 98, 748-753, Oct. 1989.

[10] R. Crochiere and L. Rabiner, "Multirate Digital Signal Processing", Prentice Hall, 1983.

[11] P. Vaidyanathan, "Quadrature Mirror Filter Banks, M-Band Extensions and Perfect Reconstruction Techniques", IEEE ASSP Magazine, 4-20,

TABLE I

MODULATION TECNIQUES AND SPECTRAL EFFICIENCIES

Modulation Type	Theoretical RF Spectral Efficiency (b/s/Hz) [1]	Practical Efficiency (b/s/Hz) [2]
2PAM	2	1.66
3PRS	2	1.66
4QAM	2	1.66
4PSK	2	1.66
4PAM	4	3.33
16QAM	4	3.33
8PAM	6	5
64QAM	6	5
16PAM	8	6.66
256QAM	8	6.66
32PAM	10	6.66
1024QAM	10	6.66

[1] Assumes an ideal brick-wall filter

[2] Assumes a Nyquist filter with 20% rolloff

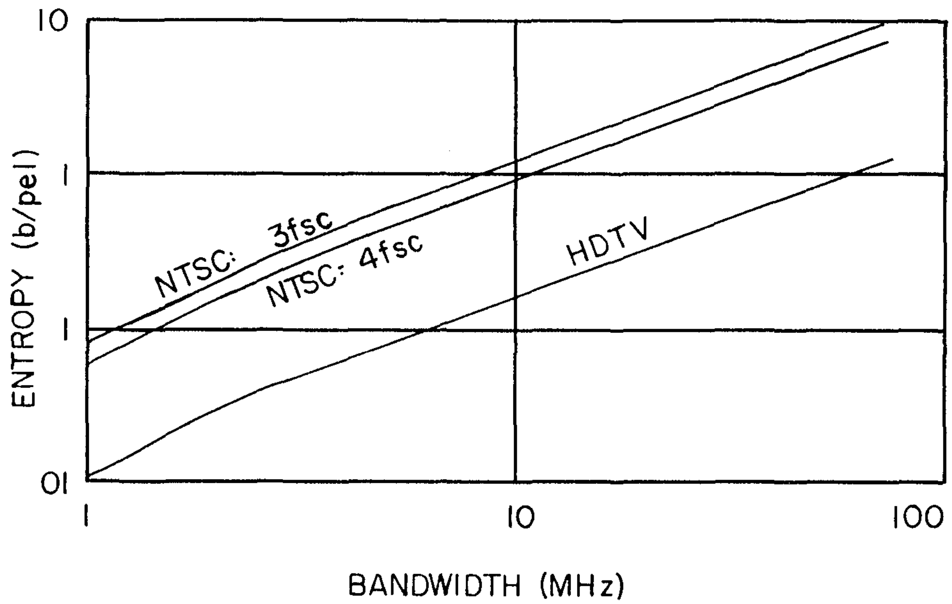

FIG. I REQUIRED ENTROPY AS A FUNCTION OF BANDWIDTH. SPECTRAL EFFICIENCY = 1.66 b/s Hz.

Predictor: X = .8X(x-1,y) - .6X(x-1,y-1) +.8X(x,y-1)

Transmitter:

line	Pixel Values		
N - 1	104	130	156
N	125	156	187
X	125	156	187
$\hat{X}$		142	172
e		14	15

Receiver:

line	Pixel Values		
N - 1	104	130	156
N	125	156	187
$\hat{X}'$		142	172
e		14	15
X'	125	156	187

Fig. 2 DPCM Example

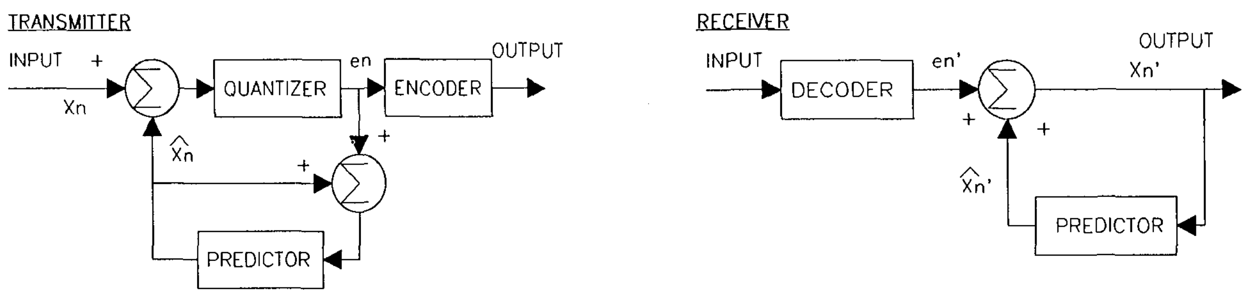

FIG. 3 DPCM SYSTEM BLOCK DIAGRAM

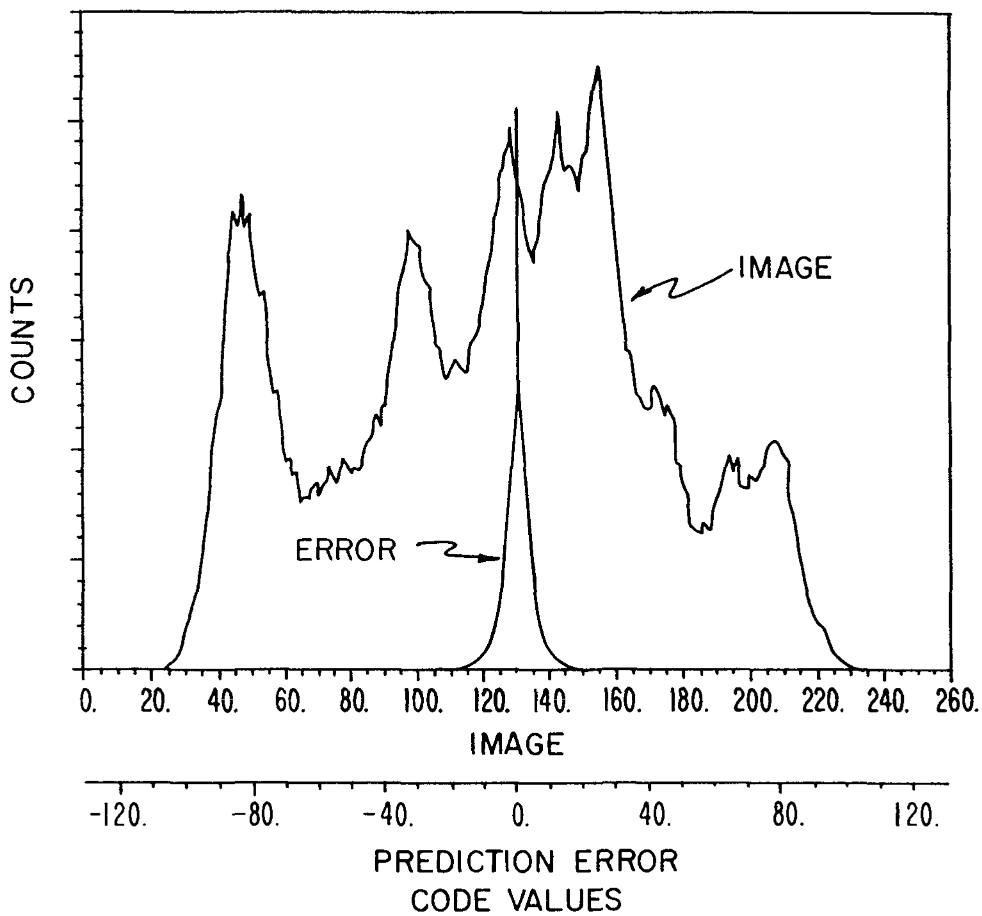

FIG. 4. COMPARISON OF IMAGE AND PREDICTION ERROR DISTRIBUTIONS

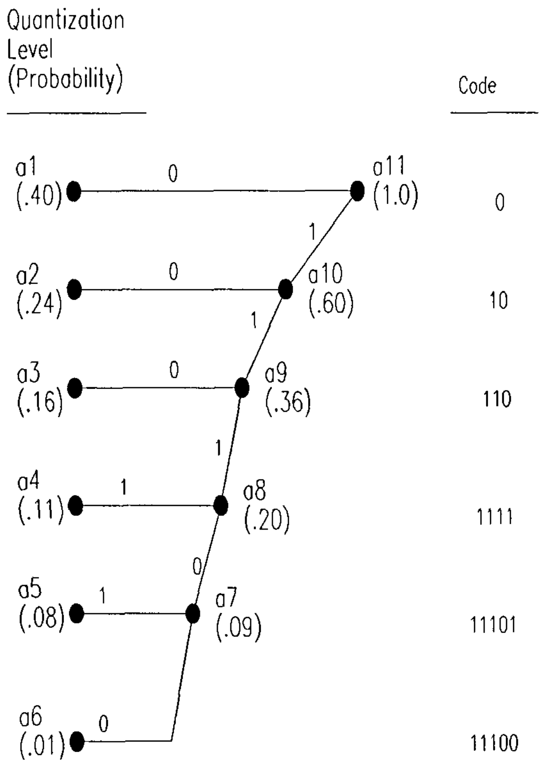

FIG. 5 Example of Huffman Coding Probabilities of Each Level are Shown in Parentheses

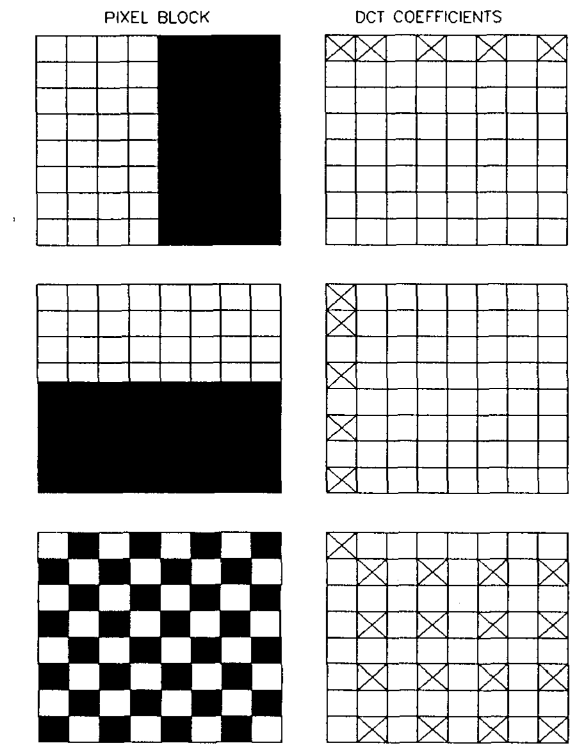

FIG 6. DCT EXAMPLES

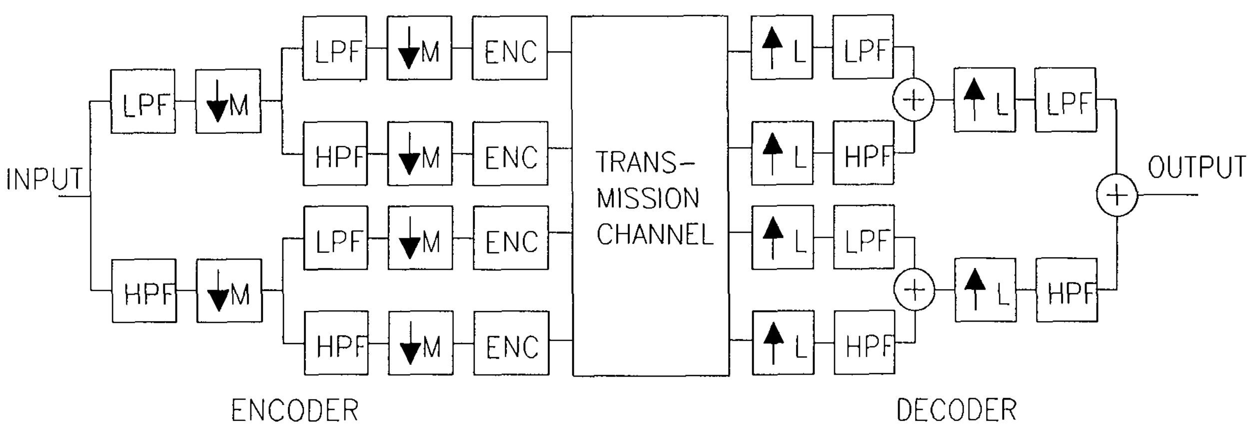

FIG. 7 SUBBAND CODING BLOCK DIAGRAM (M,L =2)

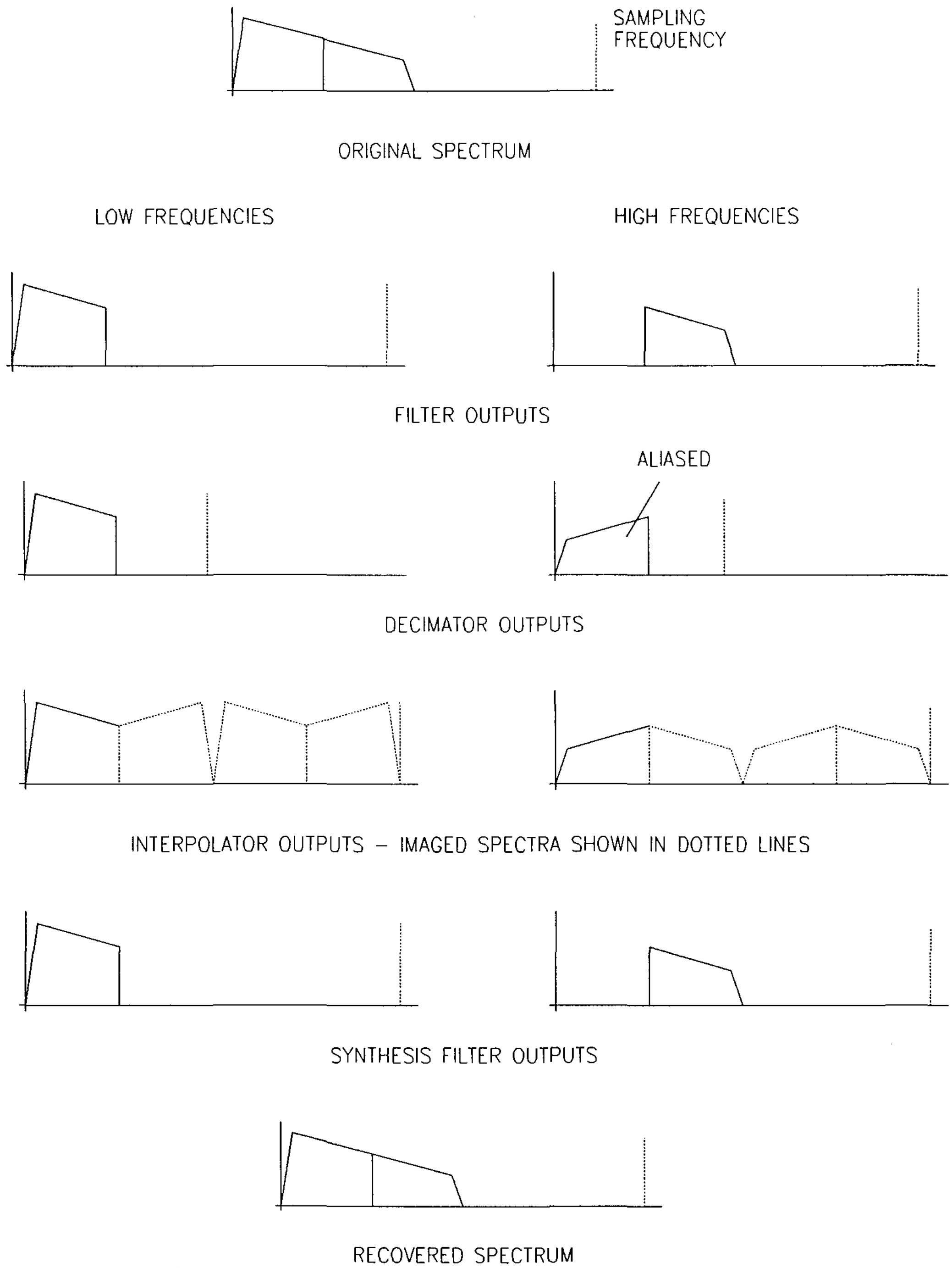

FIG. 8 SUBBAND CODING SPECTRA