Digital Video Compression—An Overview

Joseph B. Waltrich, *Member, IEEE*

Abstract—**Although transmission of compressed digital video over fiber may seem paradoxical, it is, in fact, a reality. Fiber-optic links such as Vyvx are currently transmitting digitally compressed video in DS3 format. In addition, the desirability of pass-through transmission between fiber and other media would seem to indicate increased use of fiber for digital video transmission.**

This paper presents a review of some of the more commonly employed video compression schemes. The use of motion compensation to improve compression is also discussed.

I. Introduction

BECAUSE of its reduced susceptibility to transmission channel impairments, digital video transmission offers significant advantages over analog transmission in terms of picture quality. However, a penalty is exacted in terms of bit rate and the associated bandwidth required for transmission over bandlimited media (e.g., satellite, cable, etc.) which may interface to fiber-optic links. It has been shown [1] that, for transmission of uncompressed NTSC video in digital component form, sampled at CCIR-601 rates (13.5 MHz for luminance, 6.75 MHz for each chrominance component), the bit rate is in excess of 200 Mb/s.

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:

$$W = R_d/E_s \tag{1}$$

where:

W = bandwidth (Hz)

R_d = total data rate (b/s)

E_s = spectral efficiency (b/s/Hz).

Spectral efficiencies for modulation schemes currently proposed for digital video transmission range from 2–5 b/s/Hz. Additional information regarding modulation formats for digital transmission may be found in Feher [2].

From (1), it is seen that transmission of uncompressed video at bit rates on the order of 100 Mb/s would require a bandwidth of 20–50 MHz. Transmission of HDTV in component form (approximately 1.5 Gb/s) would require a bandwidth of 0.3–0.75 GHz. Although this may not be a problem for fiber, except perhaps for the cost of the electronics at each end of the link, it is obviously impractical for alternative media which may interface with the fiber link. Therefore, all digital video transmission schemes currently being proposed make use of some form of data compression.

Manuscript received December 16, 1991; revised June 30, 1992.

The author is with Jerrold Communications, Applied Media Laboratory, Hatboro, PA 19040

IEEE Log Number 9205522.

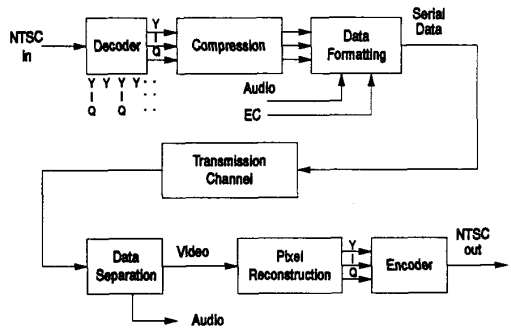


Fig. 1. Digital video transmission system block diagram.

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. (In most cases, video is digitized at 8 bits/pixel/component prior to compression). It should be borne in mind that the total data rate includes not only digitized video but also digital audio as well as error correction and other overhead data.

In practice, video compression is usually performed on the signal in component form. The most efficient approach is to perform signal processing on the luminance and chrominance components (e.g., Y, I, Q for NTSC) since, because of their lower bandwidth, the chrominance components may be sampled at half the rate of the luminance, thereby reducing the uncompressed data rate. A generic digital video transmission system is shown in block diagram form in Fig. 1. The input video signal is decoded into component form and each component is compressed individually using one or more of the techniques which will be discussed subsequently. The compressed video is then formatted for transmission, adding audio and ancillary data and the entire data stream is transmitted using one of the digital modulation techniques described in [2]. The receiver performs the reverse of this process (demodulation, demultiplexing and decompression) to recover the original signal.

Regardless of the particular technique used, compression engines accomplish their intended purpose in the following manner:

- Those portions of the video signal which are not perceptible to the human eye are not transmitted.
- Frame to frame redundancies in the video signal are not transmitted.
- The remaining information is coded in an efficient manner for transmission.

In general, in order to fit an NTSC signal into a 6-MHz bandwidth, a video entropy of 0.5–1.5 b/pel is required. Transmission of HDTV would require an entropy of about 0.08–0.16 b/pel, depending on the sampling frequency and the modulation technique.

II. 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. HDTV has also stimulated interest in compression techniques.

Currently, a number of video compression techniques are being used singly or in combination. These include the following:

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

This paper will concentrate on discussion of those techniques which are currently being proposed for the majority of compression systems.

A. 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 weighted preceding pixel values. For a pixel having a value X_N, the general form of the predictor is

$$\hat{X}_N = \sum_{i=0}^{i=N-1} a_i X_i \tag{2}$$

where: The predicted value is then subtracted from the encoded

$\hat{X}_N$ = predicted value of X_N

a_i = weighting coefficient of ith pixel in the prediction equation.

value of the Nth pixel, $\hat{X}_N$, to generate an error signal

$$e_N = \check{X}_N - \hat{X}_N. \tag{3}$$

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

$$X'_N = e_N + \hat{X}'N \tag{4}$$

where: X'_N = recovered value of Nth pixel

$\hat{X}'_N$ = predicted value of recovered pixel.

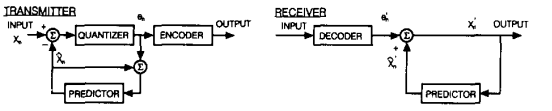


Fig. 2. DPCM system block diagram.

TRANSMITTER:			
Current Frame, Line N	125	156	187.....
Previous Frame, Line N (Predicted Pixels)	104	130	156.....
Error	 21	26	31.....
RECEIVER:			
Error, Line N	 21	26	31.....
Predicted Pixels (Previous Frame, Line N)	104	130	156.....
Current Frame (Reconstructed Pixels)	125	156	187.....

Fig. 3. DPCM example.

The value of $\hat{X}'_N$ is obtained from

$$\hat{X}'_N = \sum_{i=0}^{i=N-1} a_i X'_i \tag{5}$$

where the a_i are the same weighting coefficients used to generate the predicted pixel values in the encoder.

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 length coding such as Huffman coding.

A block diagram of a DPCM system is shown in Fig. 2. System complexity depends on the nature of the prediction algorithm. Predictors may be one, two or three dimensional, requiring from one line to one or more frames of memory. Lately, the most popular form of DPCM has utilized a temporal predictor, consisting of a frame delay, in conjunction with transform coding. Fig. 3 presents an example of a temporal DPCM system. The values shown in Fig. 3 are decimal values of pixel intensity assuming 8 bit quantization. The predictor in this example is a simple frame difference. That is:

$$\hat{X}_N = X_{N-1F} \tag{6}$$

where X_{N-1F} = the value of the Nth pixel in the previous frame.

Obviously, the receiver needs more information than just the error signal in order to reconstruct the original pixels. This information is provided by interspersing the transmission of difference values with periodic transmissions of data in PCM form in order to provide a starting point for the receiver to do its pixel reconstruction. This periodic refresh also counteracts the effect of errors which, because of the nature of the DPCM process, can propagate over several lines and/or frames.

A comparison of the distribution of pixel values for an uncompressed image and its associated error signal is shown

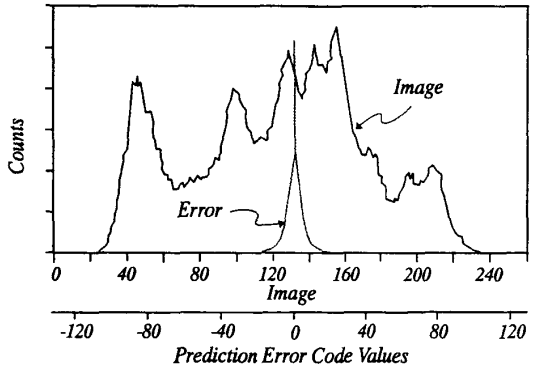


Fig. 4. Prediction error distributions.

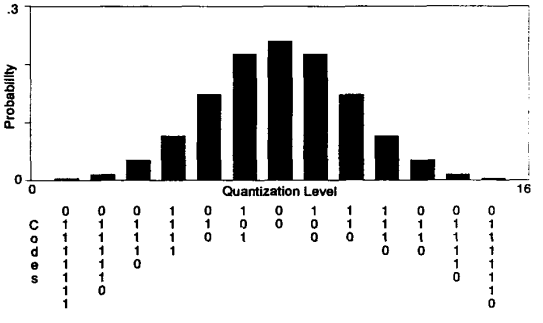


Fig. 5. Huffman coding example.

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 length code such as Huffman coding, thereby achieving a substantial reduction in entropy. This technique yields entropies on the order of 6–15 b/pel/component as compared to the 8 bits/pel/component entropy of the original image.

B. Huffman Coding

Huffman coding is a variable length coding technique which reduces the average bit rate required to represent a set of data values. This is accomplished by assigning short codewords to those data values which have the greatest probability of occurrence and longer codewords to less frequently occurring values. A crude example of Huffman coding may be found in Morse code, in which the letter E, the most frequently occurring letter in the English language, is represented by a single dot.

Obviously, variable length coding works best on signals having relatively little spread in the distribution of data values (e.g., the DPCM error signal). An example of Huffman coding is shown in Fig. 5 for a signal having 13 values. Although this is somewhat of an oversimplification, code tables for 20–30 values are not uncommon. In practice, a group of codes with very low probabilities is often assigned a single codeword and the actual pixel value is then transmitted following the codeword.

C. Transform Coding

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

A number of transforms have been used for various video compression applications. These transforms are discussed in detail by Stafford [3]. Currently, the discrete cosine transform (DCT) is the most widely used transform and is the only transform which will be discussed in this paper.

The DCT is typically performed on blocks of either 8×8 or 16×16 pixels. 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 composition of the block. The general form of a two-dimensional DCT performed on an $N \times N$ pixel block is given by the equation

$$\begin{aligned} Y_{mn} &= (4/N^2)E_m En \\ &\quad \cdot \sum_{k=0}^{N-1}\sum_{j=0}^{N-1} X_{jk} \cos\left((2j+1)m\pi/2N\right) \\ &\quad \cdot \cos\left((2k+1)n\pi/2N\right) \end{aligned} \tag{7}$$

where:

$$\begin{aligned} Y_{mn} &= \text{DCT coefficients at coordinates } m, n \\ X_{jk} &= \text{pixel amplitude at coordinates } j, k \\ E_m, E_n &= \begin{cases} 1/\sqrt{2} & (m,n=0) \\ 1 & (m,n=1,2\cdots N-1). \end{cases} \end{aligned}$$

The inverse transform (IDCT) is given by

$$\begin{aligned} X_{jk} &= E_m En \\ &\quad \cdot \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} Y_{mn} \cos\left((2j+1)m\pi/2N\right) \\ &\quad \cdot \cos\left((2k+1)n\pi/2N\right). \end{aligned} \tag{8}$$

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 nonzero coefficients. However, the presence of noise in the signal can generate spurious coefficients. Filtering the input signal is, therefore, required in order to reduce the occurrence of these coefficients.

In order to maintain a desired entropy, it is sometimes necessary to discard some of the DCT coefficients. 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 determined by one of two methods: zonal coding or threshold coding. Zonal coding discards all coefficients except those within a selected zone (which always includes the low frequency components). Threshold coding discards (or sometimes coarsely quantizes) those coefficients whose values are below a given set of threshold levels.

Since, in zonal coding, the high frequency coefficients are discarded, edge blurring sometimes occurs in the reconstructed

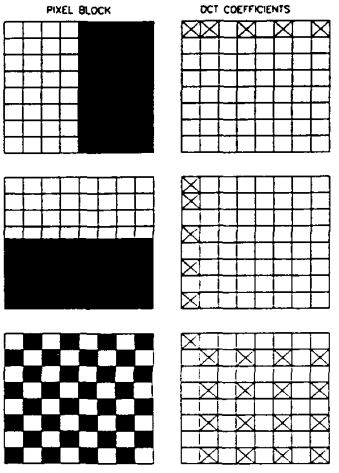


Fig. 6. DCT examples.

image. This effect is less noticeable for threshold coding but threshold coding carries a bit rate penalty since coefficient locations are not predetermined and, therefore, coefficient address information must be transmitted.

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

Several IC manufacturers now offer DCT processors. Among these are the INMOS A121, SGS-Thomson STV3208, LSI Logic L64730 and the C-Cube CL550. These chips are capable of operating at clock rates in the 13.5–40 MHz range and computing both forward and inverse transforms. Additional functions, such as Huffman coding, have also been incorporated into some of these chips.

The DCT is capable of generating reasonably good picture quality at entropies on the order of 0.5–2 b/pel (relative to an uncompressed source at 8 b/pel/component). The DCT is often combined with other compression techniques such as temporal DPCM, to achieve greater bit rate reductions.

D. Vector Quantization

In vector quantization, an image is divided into a number of nonoverlapping blocks. Each block is regarded as an N-dimensional image vector $\boldsymbol{X}$ 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 $\boldsymbol{Y1} \cdots \boldsymbol{Y_{Nc}}$ in order to find the codevector $\boldsymbol{Y_k}$ which most closely matches the image vector $\boldsymbol{X}$. Once found, the index k of the codevector is transmitted to the receiver which then uses a lookup table to reproduce the codevector $\boldsymbol{Y_k}$.

The entropy for vector quantization depends on the vector size and the number of vectors in the codebook. If the codebook size N_c is a binary number, representable by b bits, then a vector of N pixels is represented by one of 2^b codevectors and the entropy is

$$H_v = b/N = \log_2 (N_c)/N. \tag{9}$$

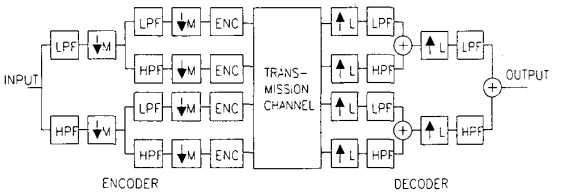


Fig. 7. Subband coding block diagram $(M, L, = 2)$.

Conversely, the codebook length N_c may be determined from the desired entropy

$$N_c = 2^{(NH_v)}. \tag{10}$$

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

The picture quality produced by vector quantization is dependent on a number of factors (e.g., vector size, codebook size, codebook versus image matching, etc.). For images which are similar to the training vectors, reasonably good quality pictures have been achieved at an entropy of 1–2 b/pel.

The key elements for effective vector quantization are codebook design and codebook search. Codebook generation may be based on statistical distribution of image vectors or on the use of training images. The LBG algorithm [4] 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. Depending on the search technique used, computational complexity is traded for encoder memory. Details of codebook design and search techniques may be found in Lim [5].

Transmission errors in vector quantization 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 and have much the same visual effect.

E. Subband Coding

In subband coding the image is divided into frequency subbands using various combinations of 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 low-pass and high-pass 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 octave band division, $M = 2$).

At the receiver, the subbands are interpolated by an integral factor L by inserting $L - 1$ zeroes between received samples of the subband data. 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 the subbands using compression schemes such as DPCM and/or by discarding some of the high frequency information or transmitting it at

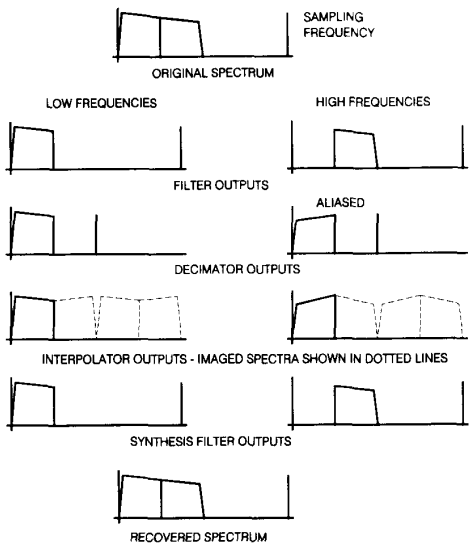


Fig. 8. Subband coding spectra.

a lower rate. Although originally proposed for video, [6], [7], this technique is now being used for audio compression [8].

The spectral effects of subband coding are shown in Fig. 8. If ideal filtering is assumed, the low frequency information is not affected by decimation but the high frequency information is aliased, with associated spectrum folding. The interpolation process creates spectral images which allow recovery of the original information by appropriate filtering of the imaged spectra. Detailed discussion of decimation and interpolation effects may be found in Crochiere and Rabiner [9].

Since it is impossible to construct ideal filters, the decimation process will produce some aliasing in all bands. The filters used must be chosen such that the synthesis filters (i.e., those following the interpolators) cancel aliasing generated during decimation. This may be accomplished by the use of Quadrature Mirror Filters Quadrature Mirror Filter design considerations are described by Vaidyanathan [10].

Subband coding is capable of producing good quality pictures at entropies of about 1 b/pel.

III. Motion Compensation

In order to achieve the maximum benefit from frame-to-frame redundancies in a television picture, it is necessary to take advantage of the fact that motion frequently produces a spatial displacement of identical pixel blocks from one frame to the next. If this difference in block position can be determined, then, instead of re-transmitting the contents of the entire block, it is only necessary to transmit a motion vector defining the block's displacement from its position in a previous frame. The current frame's block is compared with all the possible blocks within a small area (known as a search window) of the previous frame. If a match is found, a motion vector is computed and transmitted. This process is illustrated in Fig. 9.

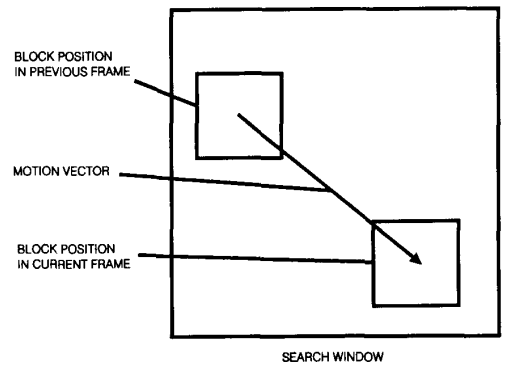


Fig. 9. Motion compensation.

The difference between an $N \times N$ luminance block and all the luminance blocks included in the search window is computed as follows:

$$D(i,j) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |SW(x+i, y+j) - Y(x,y)| \qquad (11)$$

where

D = difference

SW = search window in previous frame

Y = luminance block of size N x N in current frame

x, y = pixel coordinate within $Y(x,y)$

i,j = displacement of luminance block within search window.

The best match is defined by the minimum value of $D(i,j)$. If the minimum value is below a desired threshold, a motion vector is transmitted.

Commercially available motion estimation processors include the Thomson STV3220 and the LSI Logic 64720. Both chips are capable of working with blocks of either 8×8 or 16×16 pixels. The size of the search window (i.e., the range of values of i and j) is from $-8 \cdots +7$ for the STV3220 and $-N/2 \cdots (N/2-1)$ for the 64720. Search window size can be increased by time multiplexing computations (STV3220) or by cascading chips (64720).

Motion compensation is used in conjunction with other compression techniques to optimize bit rate reduction. For example, the proposed motion picture experts group (MPEG) compression standard [11] makes use of motion compensation in conjunction with compression using DCT and temporal DPCM. Motion compensation can achieve an additional reduction in entropy by a factor of 2–10, depending on the nature of the image.

IV. Conclusions

Digital compression has moved from the realm of the theoretical to the practical. Currently, the most popular form of compression for real time video application appears to be a combination of DCT and DPCM techniques in conjunction with motion compensation. A number of multi-function compression IC's are now available, with more devices to

come. Hardware costs are still relatively high but increased integration can be expected to result in the availability of reasonably priced consumer hardware for satellite and cable applications over the next 2–3 years.

REFERENCES

[1] A. Netravali and B. Haskell, *Digital Pictures: Representation and Compression.* New York: Plenum Press, 1989.

[2] K. Feher, *Telecommunications Measurement, Analysis and Instrumentation.* New York: Prentice-Hall, 1987.

[3] R. Stafford, *Digital Television.* New York: Wiley, 1990.

[4] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[5] J. Lim, *Two Dimensional Signal and Image Processing.* New York: Prentice-Hall, 1990.

[6] W. Schreiber *et al.,* "Channel compatible 6 MHz HDTV distribution systems," *SMPTE J.,* vol. 98, pp. 5–13, Jan. 1989.

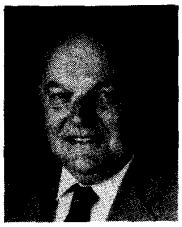
[7] W. Bretl, R. Citta, R. Lee, and P. Fockens, "Spectrum compatible high definition television system," *SMPTE J.,* vol. 98, pp. 748–753, Oct. 1989.

[8] G. Thiele, G. Stoll, and M. Link, "Low bit rate coding of high quality audio signals—An introduction to the MASCAM system," *EBU Tech. Rev.,* no. 230, pp. 71–94, Aug. 1988.

[9] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing.* New York: Prentice-Hall, 1983.

[10] P. Vaidyanathan, "Quadrature mirror filter banks, M- band extensions and perfect reconstruction techniques," *IEEE ASSP Mag.,* pp. 4–20, July 1987.

[11] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM,* Apr. 1991.



Joseph B. Waltrich (M'89) received the B.A. degree in physics from LaSalle University, Philadelphia, PA and the M.S. degree in physics from the University of Notre Dame, Notre Dame, IN.

He joined Jerrold Communications in 1985 where he has worked in research and development on advanced television, digital compression and digital transmission. He currently holds the position of Manager of Advanced Television Systems. His prior experience includes work in the development of products for broadcast television. He is a member of SMPTE and the Society of Cable Television Engineers.