A COMPARISON OF LEADING EDGE COMPRESSION TECHNOLOGIES

Arun N. Netravali
AT&T Bell Laboratories

Abstract

Digital Television is fast becoming a reality after decades of research. Broadcast, Cable and Satellite distribution of entertainment and sports television presents a unique set of requirements. In this paper, these requirements are reviewed and are followed by a brief description of the compression techniques. A compression system is then presented that combines several leading techniques to satisfy these special requirements. It is clear from this that the compression technology is real, cost-effective and can offer consumers and service providers significant benefits.

INTRODUCTION

Digital television is fast becoming a reality after decades of research.[1] It is rapidly proliferating variety of applications such as video-conferencing, multimedia computing systems and entertainment television over broadcast, cable, satellite and tape media. Many of the technical impediments which previously prevented the dream of digital television from becoming a reality are going away. This is due to significant progress in compression technology, integrated circuits for signal processing and memory, and reduced cost of transmission and storage. Moreover, digital compression provides the vehicle for a step function increase in the capacity of video delivery systems which can enable new services that were not previously possible. Thus we have a simultaneous "push" from the new services side and "pull" from the technologies of algorithms and hardware. The challenge now is to "engineer" the systems in a cost-effective way to provide an array of services to the customers.

Television signal when digitized produces an enormous bit rate, too high for economical transmission or storage. For example, the CCIR-601 television signal, when sampled and digitized requires over 200 Mbits/sec. Digital compression is therefore, essential in reducing this to a bit-rate dictated by the application. Obviously, performance of the compression hardware has to be judged by the quality of the picture at the affordable bit rate, but in addition, compressed digital television has to allow all the other functionality that analog television currently has. This functionality will vary from application to application. For example, the quality of compressed digital television should not be lower than analog television for the type of entertainment and sports material that we are used to watching over the cable, but, a lower quality may be sufficient for videoconferencing. Therefore, the choice of compression algorithm will depend on the application. Already different algorithms are in the process of getting standardized for different applications such as video teleconferencing ($p \times 64$ kbit/sec-algorithm[2]), multimedia (MPEG-1 algorithm[3]), and digital cable TV and HDTV, being led by the Cable Labs and the FCC, respectively.

In this paper, we start with a description of the desirable characteristics of the video compression algorithm for the cable, terrestrial broadcast and satellite applications. We then discuss some of the basic techniques used by most of the compression algorithms. We compare these basic techniques on the basis of compression performance as well as complexity of implementation. Following this, we describe an algorithm that appears ideally suited for cable and broadcast applications. It is clear from this, that the compression technology is ripe and will revolutionize the nature of television in the coming years.

COMPRESSION TECHNOLOGY REQUIREMENTS

The basic digital video compressor and decompressor are shown in Figure 1. As mentioned earlier, the function of the compressor is to reduce the bit rate as much as possible without sacrificing the picture quality required for the service. In addition, a number of other requirements have to be satisfied, particularly in the cable and broadcast environment. In this section we outline these requirements so that they will serve as a guide in evaluating the different compression algorithms described in the next section.

The first requirement, of course, is to achieve transparent coding for the class of pictures that are typically used in the particular service. In entertainment and sports applications, the picture material may contain high detail, complex (not necessarily predictable) motion and a large number of frequent scene cuts or changes. The picture material may be created by a television camera or synthesized by a computer or an arbitrary mixture of the two. In a number of situations, the source material may contain noise (e.g., old film or electronic news gathering cameras). The ability to compress such diverse material is indeed very challenging and very different from either video teleconferencing or the multimedia applications. Moreover, the picture quality standard accepted and practiced in the cable and the broadcast industry is significantly more demanding than what is practical in videoconferencing or multimedia applications.

The second requirement is that the compressed video bit stream must not be very fragile. Of course, compressed bit streams are always more fragile than raw uncompressed digital video. However, their robustness to transmission impairments can be significantly improved by error correction and ghost cancellation techniques. In addition, the compression algorithm must be such that a large fraction of the errors that escape the error correction mechanism can be concealed and their effects localized both in terms of space (i.e., horizontal and vertical dimensions of a picture) and time (i.e., the number of frames).

In a cable or broadcast environment, a viewer may change from channel to channel with no opportunity for the transmitter to adapt itself to such channel changes. It is important that the buildup of resolution following the channel change takes place quite rapidly so that the viewer can make a decision to either stay on the channel or change further depending upon the content that he wishes to watch.

A cable, satellite or broadcast environment has only a few transmitters which do compression but a large number of receivers which have to do decompression. Therefore, the economics is dictated to a large extent by the cost of decompression. The choice of the compression algorithm ought to make the decompression extremely simple by transferring much of the cost to the transmitter, thereby creating an asymmetrical algorithm. The existing video standards such as px64 kbit/sec for videoconference and MPEG for multimedia do not

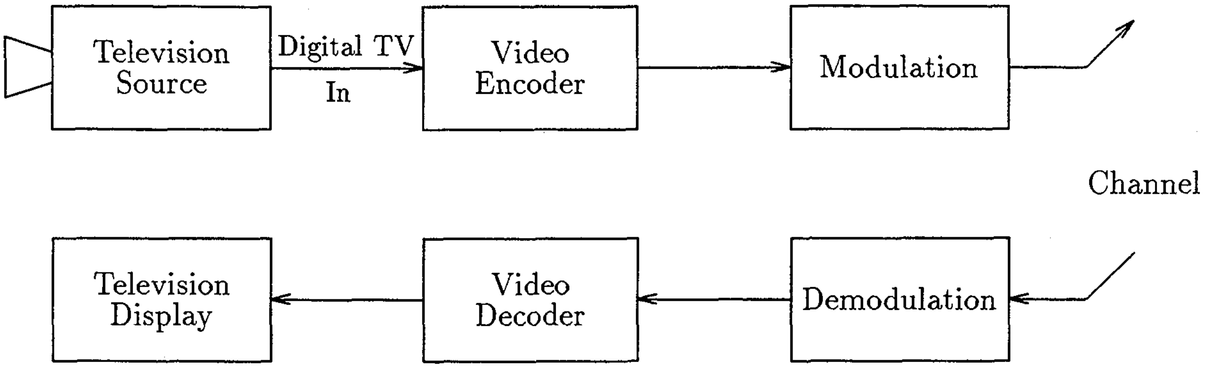


Figure 1: Video Compression and Decompression

explicitly incorporate this requirement. In a number of situations, cost of the encoder is also important (e.g., camcorder). Therefore, a modularly designed encoder which is able to trade off performance with complexity but which creates the data decodable by a simple decompressor may be the appropriate solution. Obviously, this tradeoff will change as a function of the integrated circuits technology.

In a number of instances, the original source material may have to be compressed and decompressed several times. In studios it may be necessary to store the compressed data and then decompress it for editing. Such multiple encodings of the signal is bound to increase the visibility of the coding artifacts; however, a choice of the coding algorithm should minimize the loss of quality associated with multiple encodings.

It is commonly believed that much of the material for the services based on digital compressed television will be from films. The conversion from film to video using 3:2 pull down creates a unique type of correlation in the signal. It is desirable for the compression algorithm to automatically detect such correlation and adapt itself to achieve a high degree of efficiency without increasing the complexity of the receiver.

The technology for storage of digital signals on a variety of media has made significant strides over the last few years. If the compressed digital signal is stored on a digital tape recorder, then some of the functions that we have become accustomed to should be easy to provide from the compressed digital signal. These include fast forward and backward searches, still frames, etc. This was an important consideration in the development of the MPEG-1 algorithm.

As we move toward digital television, we see a large number of possible picture resolutions (e.g., NTSC, CCIR-601, HDTV). It is desirable for the compression scheme to be compatible over these different resolutions. This will allow, for example, an HDTV decompressor to decode the compressed NTSC as well as CCIR-601 signals without much duplication of the hardware. Also, a compatibility between the transmission formats chosen for the NTSC, CCIR-601 and HDTV signals would be desirable. Moreover, such a common transmission format should allow easy interconnection or transmission over different media and telecommunications networks.

In a number of situations, particularly in the cable head end, one may wish to add an insert into a compressed digital bit stream.

It is desirable to add the insert without having to fully decode the signal. If only a small part of the coded bit stream can be affected and these effects can be localized on the picture signal, then the adding of inserts can become less damaging to the original signal.

It is clear by looking at these requirements that the broadcasting, cable, and satellite applications have requirements that are different from teleconferencing and multimedia applications. Therefore, it is not surprising that a different class of algorithms would suit these applications. In the next section we describe the basic compression techniques, and compare them in preparation for the description of an algorithm that handles the above requirements rather well.

BASIC COMPRESSION TECHNIQUES

A number of compression techniques have been developed for coding of video signals.[4] A compression system typically consists of a combination of these techniques to satisfy the type of requirements that we listed in the previous section. The first step in compression usually consists of **Decorrelation** i.e. reducing the spatial or temporal redundancy in the signal. The candidates for doing this are:

1. Making a prediction[1] of the next sample of the picture signal using some of the past and subtracting it from that sample. This converts the original signal into its unpredictable part (usually called prediction error).
2. Taking a transform[2] of a block of samples of the picture signal so that the energy would be compacted in only a few transform coefficients.

The second step is **Selection** and **Quantization** to reduce the number of possible signal values. Here, for DPCM, the prediction error may be quantized sample at a time or a vector of prediction error of many samples may be quantized all at once. Alternatively, for transform coding, only important coefficients may be selected and quantized. The final step is **Entropy Coding** which recognizes that different values of the quantized signal occur with different frequencies and, therefore, representing them with unequal length binary codes reduces the average bit rate. We give below more details of the following techniques since they have formed the basis of most of the compression systems:

a) Predictive Coding (DPCM)

b) Transform Coding

c) Motion Compensation

d) Vector Quantization

e) Entropy Coding

f) Incorporation of Perceptual Factors

Predictive Coding (DPCM)

In predictive coding, the strong correlation between adjacent pels (spatially as well as temporally) is exploited. As shown in Figure 2, an approximate prediction of the sample to be encoded is made from previously coded information that has already been transmitted. The error (or differential signal) resulting from the subtraction of the prediction from the actual value of the pel is quantized into a set of discrete amplitude levels. These levels are then represented as binary words of fixed or variable lengths and

---

[1] This is also the first step in a family of compression algorithms known as Differential Pulse Code Modulation (DPCM).

[2] This is called Transform Coding. Transform is simply a linear combination of all the pels in the block.

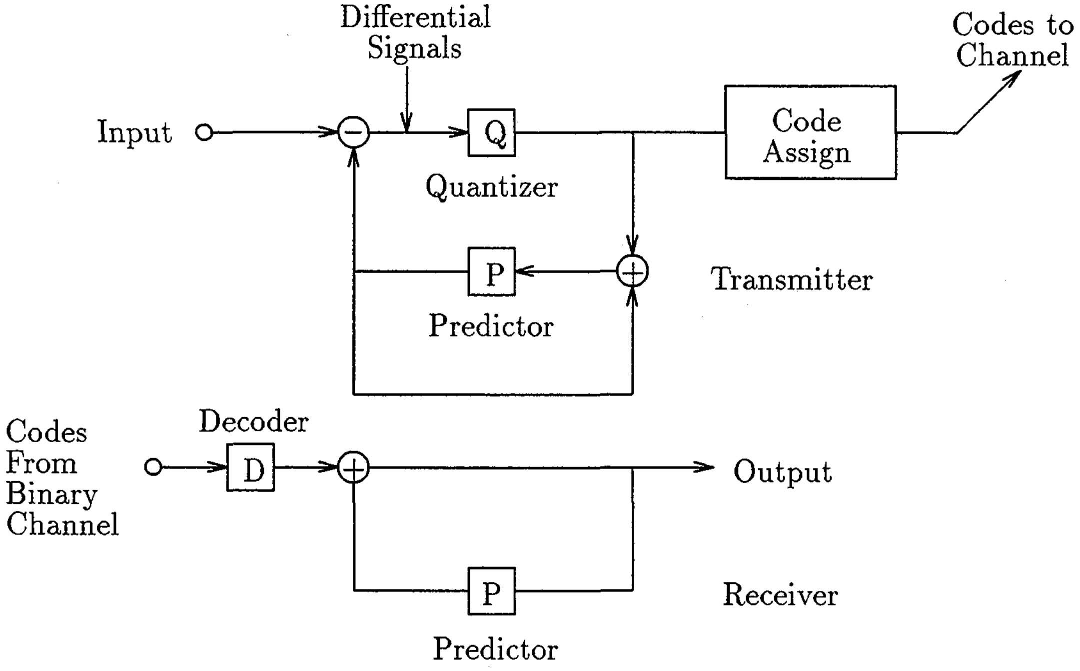


Figure 2: Block Diagram of a Predictive Encoder and Decoder

sent to the channel for transmission. The predictions may make use of the correlation in the same scanning line or adjacent scanning lines or previous fields. A particularly important prediction is the **motion compensated** prediction. If a television scene contains moving objects and an estimate of frame-to-frame translation of each moving object is made, then more efficient prediction can be performed using elements in the previous frame that are appropriately spatially displaced. Such prediction is called motion compensated prediction. The translation is usually estimated by matching a block of pels in the current frame to a block of pels in the previous frames at various displaced locations. This is shown in Figure 3. Various criteria for matching and algorithms to search for the best match have been developed.[4] Typically, such motion estimation is done only at the transmitter and the resulting motion vectors are used in the encoding process and also separately transmitted for use in the decompression process.

Transform Coding

In transform coding (Figure 4) a block of pels are transformed by transform **T** into another domain called the transform domain, and some of the resulting coefficients are quantized and coded for transmission. The blocks may contain pels from one, two or three dimensions. The most common technique is to use a block of two dimensions. Using one dimension does not exploit vertical correlation and using three dimensions requires several frame stores. It has been generally agreed that **Discrete Cosine** transform is best matched to the statistics of the picture signal and moreover, since it has a fast implementation, it has become the transform of choice.[3] The advantage

---

[3] Also, recent advances have dramatically decreased the cost of implementing transforms using LSI[5].

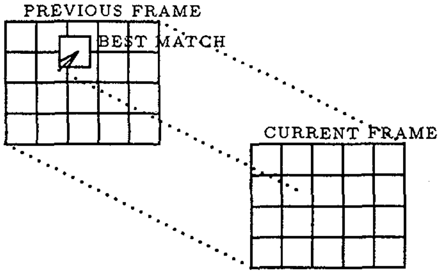
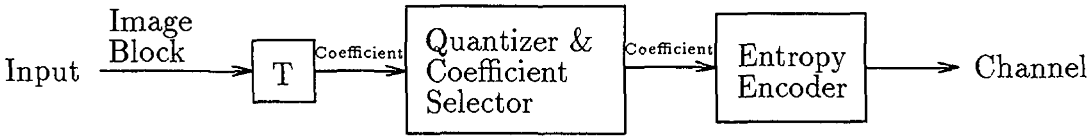
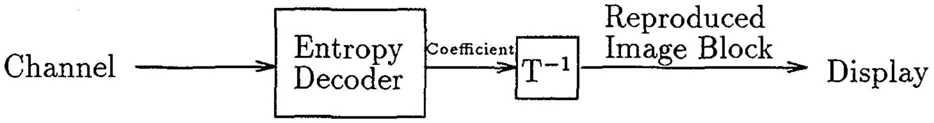


Figure 3: Block Matching for Motion Estimation



Encoder



Decoder

Figure 4: Transform Coding

of transform coding comes about mainly from two mechanisms. First, not all of the transform coefficients need to be transmitted in order to maintain good image quality, and second, the coefficients that are selected need not be represented with full accuracy. Loosely speaking, transform coding is preferable to predictive coding for lower compression rates and where cost and complexity are not extremely serious issues. Most modern compression systems have used a combination of predictive and transform coding. In fact, motion compensated prediction is performed first to remove the temporal redundancy, and then the resulting prediction error is compressed by two-dimensional transform coding using Discrete Cosine transform as the dominant choice.

Vector Quantization

In predictive coding, described in the previous section, each pixel was quantized separately using a scalar quantizer. The concept of scalar quantization can be generalized to vector quantization in which a group of pixels are quantized at the same time by representing them as a code vector. Such a vector quantization can be applied to a vector of prediction errors, original pels, or transform coefficients. As in Figure 5, a group of 9 pixels from a 3×3 block is represented to be one

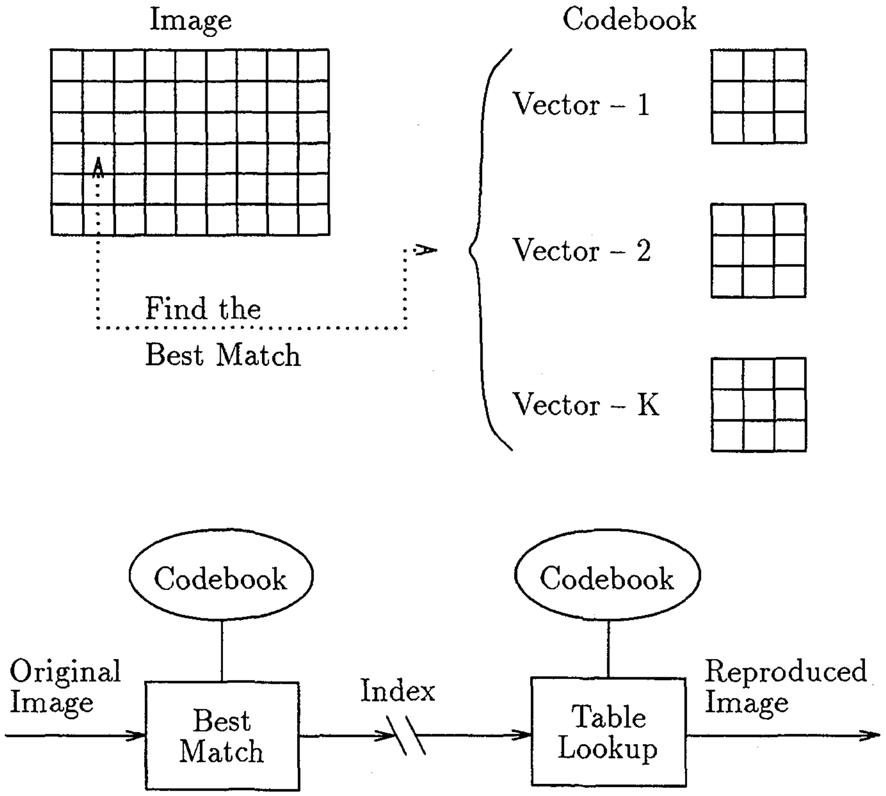


Figure 5: Vector Quantization

of the k vectors from a codebook of vectors. The problem of vector quantization is then to design the codebook and an algorithm to determine the vector from the codebook that offers the best match to the input data. The design of codebook usually requires a set of training pictures and can grow to a large size for a large block of pixels. Thus, for an 8×8 block compressed to two bits per pel, one would need 2^{128} size codebook. Matching the original image with each vector of such a large size codebook requires a lot of ingenuity. However, such matching is only done at the transmitter, and the receiver is considerably simple since it does a simple table lookup.

Entropy Coding

If the quantized output values of either a predictive or a transform coder are not all equally likely, then the average bit rate can be reduced by giving each one of the values a different word length. In particular, those values that occur more frequently are represented by a smaller length code word. If a code with variable length is used, and the resulting code words are concatenated to form a stream of bits, then correct decoding by a receiver requires that every combination of concatenated code words be uniquely decipherable. A variable word length code that achieves this and at the same time gives the minimum average bit rate is called **Huffman Code**. Variable word length codes are more sensitive to the effect of transmission errors since synchronization would be lost in the event of an error. This can result in several code words getting decoded incorrectly. A strategy is required to limit the propagation of errors in the presence of Huffman Codes.

Incorporation of Perceptual Factors

The perception based coding attempts to match the coding algorithm to the characteristics of human vision. We know, for example, that the accuracy with which the human eye can see the coding artifacts depends upon a variety of factors such as the spatial and temporal frequency, masking due to the presence of spatial or temporal detail, etc. A measure of the ability to perceive the coding artifact can be calculated based on the picture signal.[6] This is used, for example, in transform coding to determine the precision needed for quantization of each coefficient. Perceptual factors control the information that is discarded on the basis of its visibility to the human eye. It can, therefore, be incorporated in any of the above basic compression schemes.

Comparison of Techniques

Figure 6 represents an approximate comparison of different techniques using compression efficiency vs. complexity as a criterion under the condition that the picture quality is held constant at a "8-bit PCM level". The compression efficiency is in terms of compressed bits per Nyquist sample, and therefore, different resolution and bandwidth pictures can be simply scaled by proper multiplication to get the relevant bit rates. The complexity allocated to each codec should not be taken too literally; rather, it is an approximate estimate relative to the cost of a PCM codec which is given a value of five. Furthermore, it is the complexity of only the decoder portion of the codec, since that is the most important part of the digital television. The relation of cost to complexity is controlled by an evolving technology, and codecs with high complexity are fast becoming inexpensive through the use of application specific video DSP's and submicron device technology. Also, most of the proposed systems are a combination of several different techniques of Figure 6, making such comparisons difficult. As we remarked before, the real challenge is to combine the different techniques to engineer a cost-effective solution for a given service. The next section describes one example of such a codec.

A COMPRESSION SCHEME

In this section we describe a compression scheme that combines the above basic techniques to satisfy the requirements of Section 2. We have used this compression scheme successfully for both the noninterlaced HDTV signals and interlaced NTSC and CCIR-601 signals.[7,8]

Three basic types of redundancy are exploited in the video compression process. Motion compensation removes temporal redundancy, two-dimensional DCT removes spatial redundancy, and perceptual weighting removes amplitude redundancy by putting quantization noise in less visible areas.

Temporal processing occurs in two stages. The motion of objects from frame to frame is estimated using hierarchical block matching. Using the motion vectors, a displaced frame difference (DFD) is computed which generally contains a small fraction of the information in the original frame. The DFD is transformed using DCT to remove the spatial redundancy. Each new frame of DFD is analyzed prior to coding to determine its rate versus perceptual distortion character-

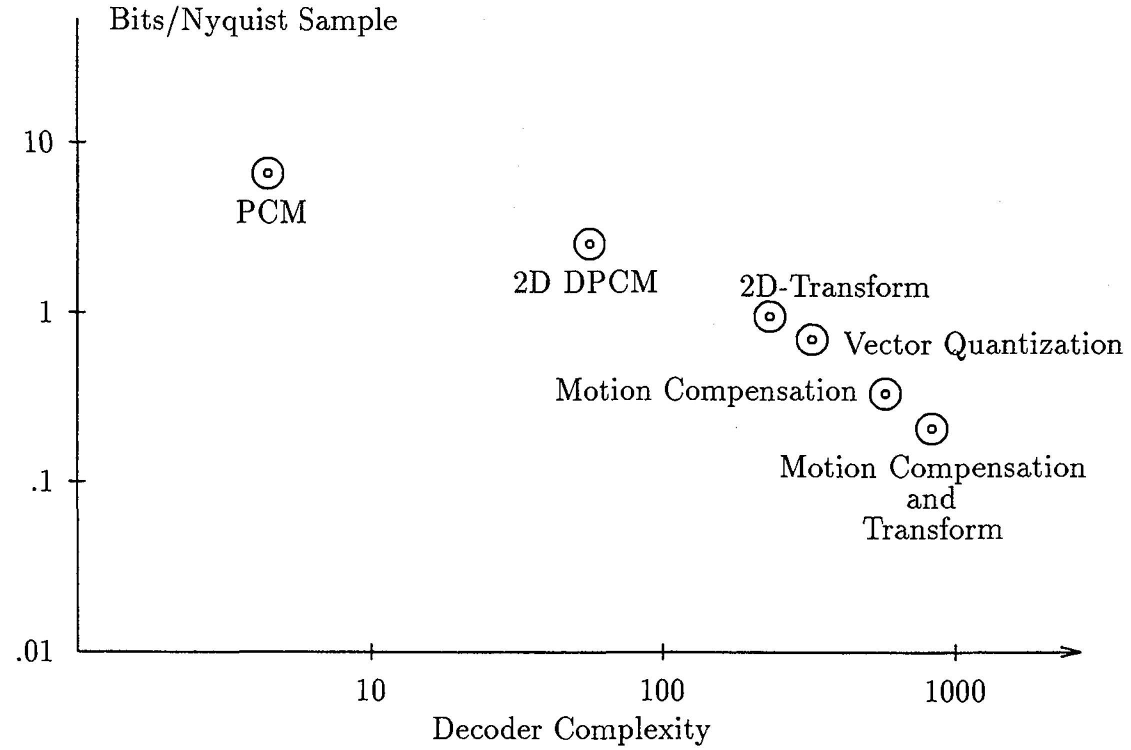


Figure 6: Compression vs. Complexity

istics and the dynamic range of each coefficient (forward analysis).[4] Quantization of the transform coefficients is performed based on the perceptual importance of each coefficient, the precomputed dynamic range of the coefficients, and the rate versus distortion characteristics. The perceptual criterion uses a model of the human visual system to determine a human observer's sensitivity to color, brightness, spatial frequency and spatial-temporal masking. This information is used to minimize the perception of coding artifacts throughout the picture. Parameters of the coder are optimized to handle the scene changes that occur frequently in entertainment/sports events, and channel changes made by the viewer. The motion vectors, compressed transform coefficients and other coding overhead bits are packed into a format which is highly immune to transmission errors.

The encoder is shown in Figure 7. Each frame is analyzed before being processed in the encoder loop. The motion vectors and control parameters resulting from the forward analysis are input to the encoder loop which outputs the compressed prediction error to the channel buffer. The encoder loop control parameters are weighed by the buffer state which is fed back from the channel buffer.

In the predictive encoding loop, the generally sparse differences between the new image data and the motion-compensated predicted image data are encoded using adaptive DCT coding. The parameters of the encoding are controlled in part by forward analysis. The data output from the encoder

---

[4] This also helps us to automatically detect the 3:2 pull down signals and adapt to them.

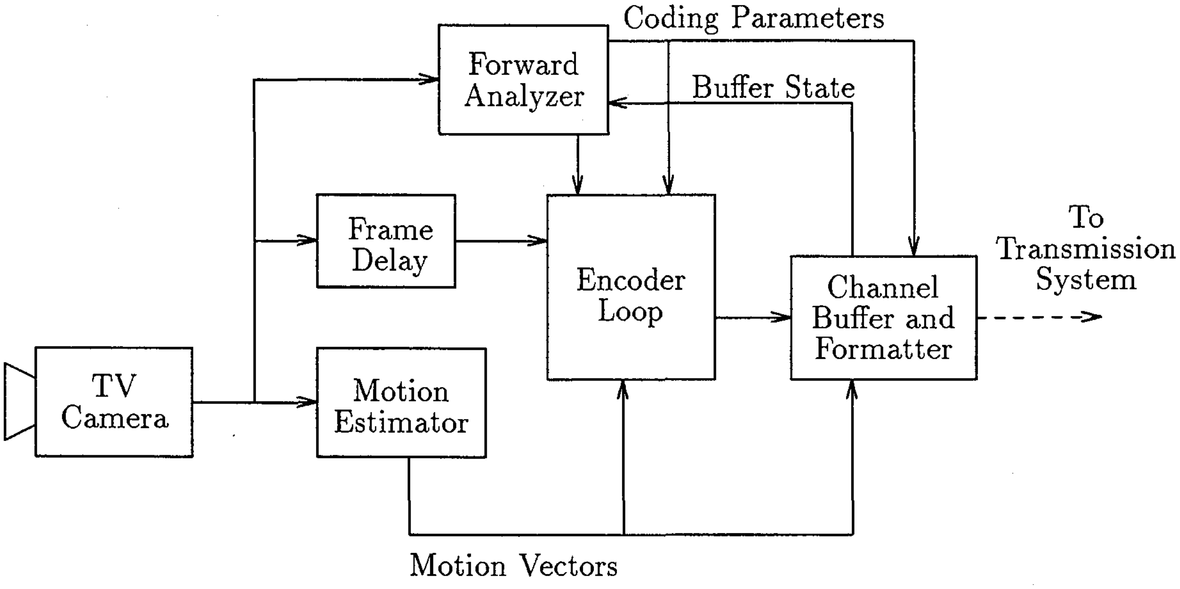


Figure 7: Encoder

consists of some global parameters of the video frame computed by the forward analyzer and transform coefficients that have been selected and quantized according to a perceptual criterion.

Each frame is composed of a luminance frame and two chrominance difference frames which are half the resolution of the luminance frame horizontally. The compression algorithm produces a chrominance bit-rate which is generally a small fraction of the total bit-rate, without perceptible chrominance distortion.

The output buffer has an output rate of between 2 to 7 Mbits/sec and has a varying input rate that depends on the image content. The buffer history is used to control the parameters of the coding algorithm so that the average input rate equals the average output rate. The feedback mechanism involves adjustment of the allowable distortion level, since increasing the distortion level (for a given image or image sequence) causes the encoder to produce a lower output bit rate.

The encoded video is packed into a special format before transmission which maximizes immunity to transmission errors by masking the loss of data in the Decoder. The duration and extent of picture degradation due to any one error or group of errors is limited. The Decoder is shown in Figure 8. The compressed video data enters the buffer which is complementary to the compressed video buffer at the encoder. The decoding loop uses the motion vectors, transform coefficient data, and other side information to reconstruct the NTSC images. Channel changes and severe transmission errors are detected in the decoder causing a fast picture recovery process to be initiated. Less severe transmission errors are handled gracefully by several algorithms depending on the type of error.

Processing and memory in the decoder are minimized. Processing consists of one inverse spatial transform and a variable length decoder which are realizable in a few VLSI

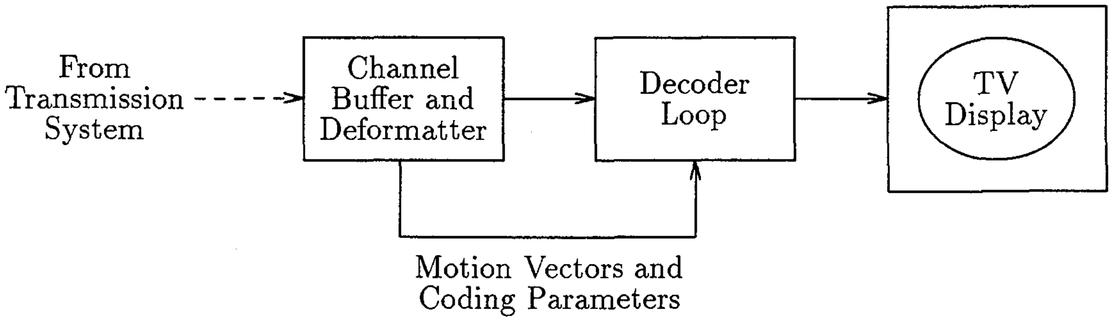


Figure 8: Decoder

chips. Memory in the decoder consists of one full frame and a few compressed frames.

CONCLUSIONS

We have presented in this paper a set of requirements specific to digital compression of television signals for broadcast, cable and satellite distribution. These requirements are considerably different than those for videoconferencing or multimedia computing. We have outlined a collection of state of the art basic compression techniques and synthesized a compression scheme that meets the special requirements of the broadcast, cable and satellite industries. Rapid advances in video compression, and semiconductor technology (processing as well as memory) have made digital television realizable.

REFERENCES

[1] W. M. Goodall, "Television by Pulse Code Modulation," Bell System Technical Journal, Vol. 30, Jan. 1951, pp. 33–49.

[2] M. Liou, "Overview of the px64 kbit/s Video Coding Standard," Communication of the ACM, April 1991, pp. 59–63.

[3] D. LeGall, "MPEG: A Video Compression Standard for Multimedia Applications," Communications of the ACM, April 1991, pp. 46–58.

[4] A. N. Netravali and G. G. Haskell, "Digital Pictures: Representation and Compression, Plenum 1988.

[5] S. C. Knauer, "Efficient Implementation of DCT and Other Techniques," Internal Document, AT&T Bell Laboratories 1990.

[6] R. J. Safranek and J. D. Johnston, "A Perceptually Tuned Sub-band Image Coder with Image Dependent Quantization and Post-Quantization Data Compression," Proc. IEEE ASSP 89, May 1989, pp. 1945–1948.

[7] A. N. Netravali, et al., "A High Performance Digital HDTV Coder," Proceedings of the NAB, 1991.

[8] A. N. Netravali, et al., "A High Performance Digital NTSC/CCIR-601 Coder," SCTE Fiber Optics Conference, 1992.