An Overview of the JPEG and MPEG Video Compression Specifications

William Woodward
Staff Engineer
Scientific-Atlanta Inc.

ABSTRACT

Digital Video Compression has become one of the fastest growing technologies in the last several years. As part of the development of this technology, several digital video compression standards are emerging. One of these standards JPEG, is intended for specifying the compression and decompression of single frame images. Another standard, MPEG, is intended for specifying the compression and decompression of motion video. These standards may have a significant effect on the CATV industry in the future. This paper describes these standards and gives an explanation of how they work.

INTRODUCTION

Over the past year, two digital video compression specifications have neared completion. The JPEG (Joint Photographic Experts Group) specification is being developed by a joint ISO/CITT committee. Also the MPEG (Motion Picture Experts Group) specification is being developed by a joint ISO/IEC committee. These specifications are important steps in the evolution of digital video systems.

The need for image data compression is apparent when we examine the data required to represent a single image. A 720 by 480 pixel image stored using 24 bits per pixel (eight bits per component for a red, green and blue color space) will require 1.036 Mbytes of memory. If this size image is used in a full motion video system with a frame rate of 30 frames per second, the channel data rate will be 248.8 Mbits/sec. The bandwidth required to transmit this signal using a modulation scheme which has a bit packing rate of 3 bits per hertz would be 82.9 MHz. This is too much bandwidth to be used in most practical systems. Some means of reducing the amount of data required to represent the signal must be found to make a digital video system viable.

DATA COMPRESSION

There are two general classes of data compression, lossless and lossy. Lossless data compression schemes rely on reducing the redundant information in the data while representing the data with as few logical indicators (bits) as possible. These schemes can be used for any kind of data since no information is lost in the data compression and expansion process. Lossy data compression schemes throw out information and rely on human psycho-visual properties in order to keep the distortions produced by data compression from being perceived. Both JPEG and MPEG are lossy compression schemes, however they both have lossy and lossless elements. Before describing the details of the JPEG and MPEG specifications, a discussion of some of these techniques is in order.

Three common lossless data compression techniques which are used in both JPEG and MPEG are Run Length Coding, Variable Word Length Coding and Predictive Coding.

Run Length Coding

Run Length Coding is used when the data tends to contain long strings of identical characters. Instead of transmitting the characters, the number of characters in the string is transmitted. For instance JPEG and MPEG use zero run length coding. When repeated zeros occur in the data, the number of zeros is transmitted instead of the actual zeros.

Variable Word Length Coding

Variable Word Length Coding, which is also called Huffman Coding, uses different length codes to represent characters. The length of codes used to represent characters is inversely proportional to the probability of occurrence of the character. In addition Huffman Codes have a prefix property which means that no short code group is duplicated as the beginning of a larger group(1). If the estimate of the probability of characters occurring is incorrect, Huffman Coding can increase the number of bits required to represent the data.

Predictive Coding

Predictive coding first estimates the present data value through some prediction algorithm. An error signal which is the difference between the actual data and the predicted data is then transmitted. The decoder uses the same prediction method as the encoder. By adding the error signal to the prediction, the original signal is exactly reconstructed in the decoder. If the prediction is a good estimate of the original data, the error signal will be small and can be represented by fewer bits than the original data.

Transform Coding

A common lossy coding technique which is used in MPEG and JPEG is transform coding. Transform coding takes image data in the spatial domain which tends to have high pixel to pixel correlation and transforms it in a manner which tends to group the energy into relatively few transform coefficients. The coefficients which are small or zero can be eliminated without any significant effect on the image(2).

JPEG SPECIFICATION

The JPEG specification describes two types of image compression algorithms for compressing and decompressing single gray-scale and color images. The first algorithm is the "Baseline/Extended" algorithm which is based on lossy transform coding and provides the greatest compression of the two algorithms. The baseline system is the minimum configuration for the transform coding based algorithm. The extended system includes the baseline system and some additional features. The second compression algorithm is the "Independent Function" system which is based on two dimensional differential pulse code modulation. The Independent Function system can be a lossless or lossy system and is intended for very high quality images. It does not provide large amounts of data compression. Because the JPEG system is intended for many applications, it supports numerous image color space representations and a wide range of image spatial resolutions. It is a symmetric algorithm in that the decoder performs the inverse process of the encoder and is therefore of the same complexity.

DCT

The block diagram of a baseline JPEG encoder is shown in figure 1. Each component of the image color space (i.e. R, G, or B for RGB, Y, U, or V for YUV color space) is partitioned into 8 x 8 pixel blocks. Each of these blocks is then coded. The compression algorithm is based on the two dimensional Discrete Cosine Transform or DCT given below(3).

$$F(u,v)=$$

$$\left(\frac{1}{4}\right) c(u)\, c(v) \sum_{i=0}^{7} \sum_{j=0}^{7} f(i,j) \cos((2i+1)u\frac{\pi}{16}) \cos((2j+1)v\frac{\pi}{16})$$

$$c(u)=\frac{1}{\sqrt{2}} \quad \textit{for} \quad u=0$$

$$c(u)=1 \quad \textit{for} \quad u\neq 0$$

$$c(v)=\frac{1}{\sqrt{2}} \quad \textit{for} \quad v=0$$

$$c(v)=1 \quad \textit{for} \quad v\neq 0$$

This transform takes a 8 X 8 pixel block of data which represents the horizontal and vertical spatial components of the image and transforms the data into a 8 X 8 block of coefficients which represent the horizontal and vertical frequency components of the

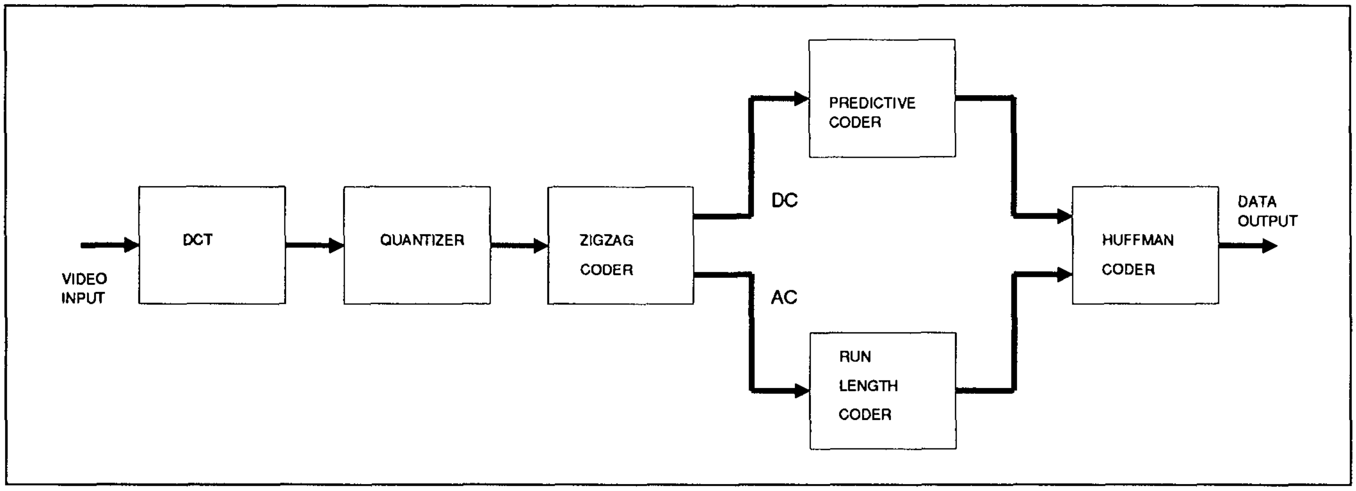
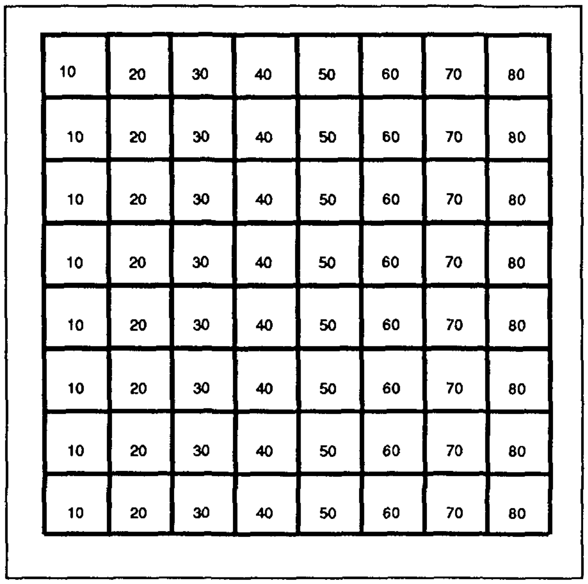
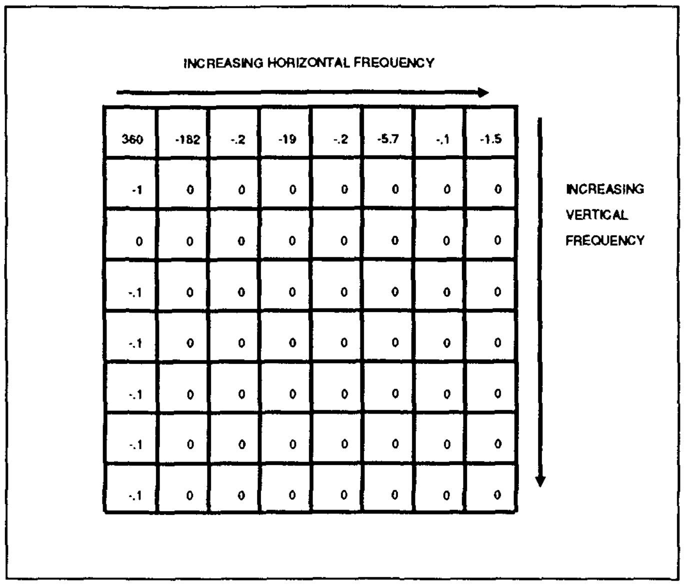


Figure 1 JPEG Encoder



10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80
10	20	30	40	50	60	70	80

Figure 2A 8x8Pixel Image Data


INCREASING HORIZONTAL FREQUENCY
INCREASING VERTICAL FREQUENCY

360	-182	-.2	-19	-.2	-5.7	-.1	-1.5
-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-.1	0	0	0	0	0	0	0
-.1	0	0	0	0	0	0	0
-.1	0	0	0	0	0	0	0
-.1	0	0	0	0	0	0	0
-.1	0	0	0	0	0	0	0

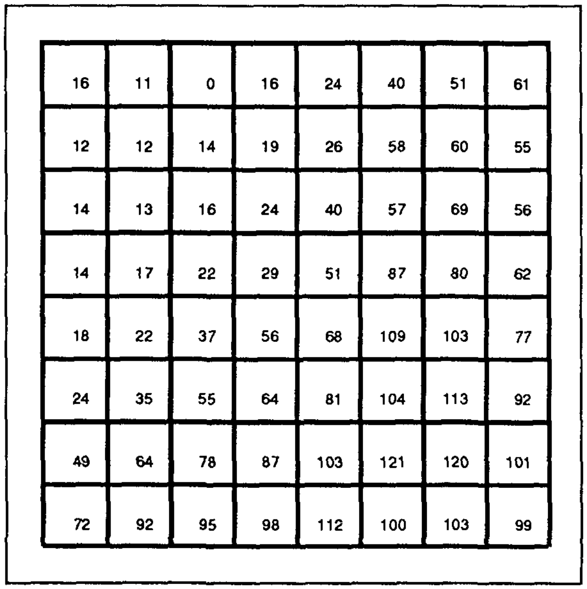
Figure 2B Transformed Image Data

data block. This has the property of concentrating the energy of the image into fewer terms.

Figure 2A represents data from an 8 X 8 pixel portion of the luminance component of an image. This block represents a portion of an image with increasing brightness from left to right. Figure 2B represents the DCT of this data block. Most of the higher frequency terms are very small or zero.

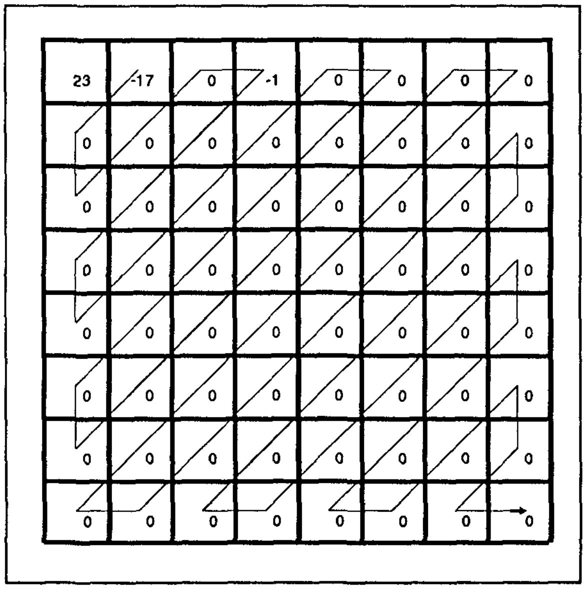
Quantization

The next step in compressing the image data is to quantize the DCT coefficients. Quantization is a process which breaks the possible DCT coefficient range into windows, then generates a code to represent each window. Each coefficient is linearly quantized (the quantization window size is constant) independently of the other coefficients. The quantization window or step size determines how



16	11	0	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figure 3 Quantizer Step Size



23	-17	0	-1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Fiqure 4 Quantized DCT Coef.

many bits will be used to represent the data. The larger the quantization step, the greater the distortion produced by the quantization process. However, fewer bits will be required to represent the data. Since the JPEG specification allows each DCT coefficient to be assigned its own quantization step size, more important frequency terms can be represented more accurately than less important terms. Many of the higher frequency terms will be set to zero by quantization. Figure 3 shows a quantization matrix containing typical quantizer step sizes. The data in figure 2B is quantized using the quantizer step sizes in figure 3 to produce the data in figure 4.

ZIG-ZAG Scanning

The frequency terms are scanned or reordered according to increasing spatial frequency. Since higher spatial frequency terms are often zero or quantized to zero, there will tend to be many zero terms in a row. This prepares the data to be Run Length Coded. The scanning process is performed by zigzagging diagonally across the block of DCT coefficients, hence it is known as zigzag scanning. Figure 4 illustrates the quantized and ZIG-ZAG scanned data.

ENTROPY CODING

Entropy coding is a general term for lossless coding techniques which are used to code the quantized DCT coefficients.

Predictive Coding

Unlike the higher frequency terms that can be coarsely quantized without significant image degradation, the DC term (first term in the transformed data block) must be represented accurately. The DC term is predictive coded using the previous 8 X 8 block's DC term as the predictor. The difference between the predictor and the present block's DC coefficient is Huffman Coded to form the first portion of the block's output data.

Run Length Coding

The AC DCT coefficients (remaining 63 terms) are Zero run length coded and then Huffman coded. The number of zeros preceding a non-zero AC term is combined with the length of the nonzero term in bits. This number is then Huffman coded and the actual bits which represent the non-zero term are appended to the end of the Huffman Code forming output data. This process is repeated until the end of the data block is reached. If all of the terms to the end of the block are zero, then an end of block signal is coded. Each 8 X 8 pixel block in the image is encoded in this manner until the entire image is coded.

The data compression obtained using JPEG is varied by changing the quantization levels of the quantizer. There is no way to know what the compression ratio will be (and therefore know how much output data there will be) until the image is compressed. For this reason JPEG is a variable output data rate system. No quantization levels are called out in the specification, however communications syntax is specified which allows this information to be sent from the encoder to the decoder. Also the Huffman codes are not specified but can be sent from the encoder to the decoder. This allows the compression/decompression process to be optimized for a particular image or group of images.

The Extended system has several features which are not included in the baseline system. One of these is a progressive scan mode. The baseline system operates in a sequential mode which sends all of the image information in raster scan order (left to right, top to bottom). The progressive mode sends some of the information for the entire image so that a low quality image can be produced. More information is progressively sent until all of the image information has been sent. The Extended system also allows for arithmetic coding instead of Huffman Coding. Arithmetic coding uses the statistics of the images being coded which increases the amount of compression obtained.

JPEG produces images which are recognizable at entropy levels of .15 bits/pixel. Excellent quality images are obtained at entropy levels of .75 bits/pixel and images which are essentially indistinguishable from the original are obtained at entropy levels of 1.5 bits/pixel(4). This implies compression ratios of 160:1, 32:1 and 16:1 respectively (assuming 24 bits / pixel source images).

While JPEG could be used to encode video by compressing each frame of the video sequence independently, this would not achieve as high of a compression level as is possible since no temporal or frame to frame redundancy is removed. For instance, if there is no motion in a video sequence, then all but the first frame of video is redundant. Video compression which only operates within a single frame of video is termed intra-frame compression while algorithms which operate using information from more than one frame are termed inter-frame compression.

MPEG SPECIFICATION

The MPEG standard is a lossy inter-frame compression scheme based on the DCT. The specification itself is a specification for a decoder only, and leaves the implementation of the encoder to the system designer. The specification assures that if an encoder uses the proper syntax to encode image data, then a MPEG decoder will be able to decode it. Because the signal quality is a strong function of the encoder design, all MPEG systems will not have the same image quality. MPEG relies on Signal Bandwidth Reduction, Lossy Compression and Lossless Compression to obtain image data compression.

Bandwidth Reduction

The input video format for MPEG is 352 X 288 pixel non-interlaced video frames in a Y,U,V color space. Converting the video to this format from a CCIR 601 YUV 4:2:2 format requires reducing the amount of data (bandwidth) by almost 1/6 before the compression process starts.

MPEG COMPRESSION

The block diagram in figure 5 depicts a MPEG decoder and a typical MPEG encoder. The MPEG specification defines three different types of video frame coding: Intra-coded(I), Predictive-coded(P), and Bidirectionally Predictive-coded(B). Each of these types of frame coding requires a different amount of data to encode. An Intra-coded frame requires the most data to code with a Predictive-coded frame requiring the second largest amount of data. The sequence of these frames is determined by the encoder. Figure 6 shows how a typical MPEG video sequence might be constructed from these frames.

The Intra-coded frames are very similar to JPEG coding except in the details of the Huffman coding. With Predictive-coded frames, the previous frame is used as a prediction for the present frame. The encoder then uses this predicted frame to generate an error signal which is encoded like an Intra-frame. The same prediction frame is generated in the decoder and the decoded error signal is added to it to generate the final output. The same process takes place with Bidirectionally Predictive-coded frames except that either the next frame, the previous frame or both are used to make a prediction of the present frame.

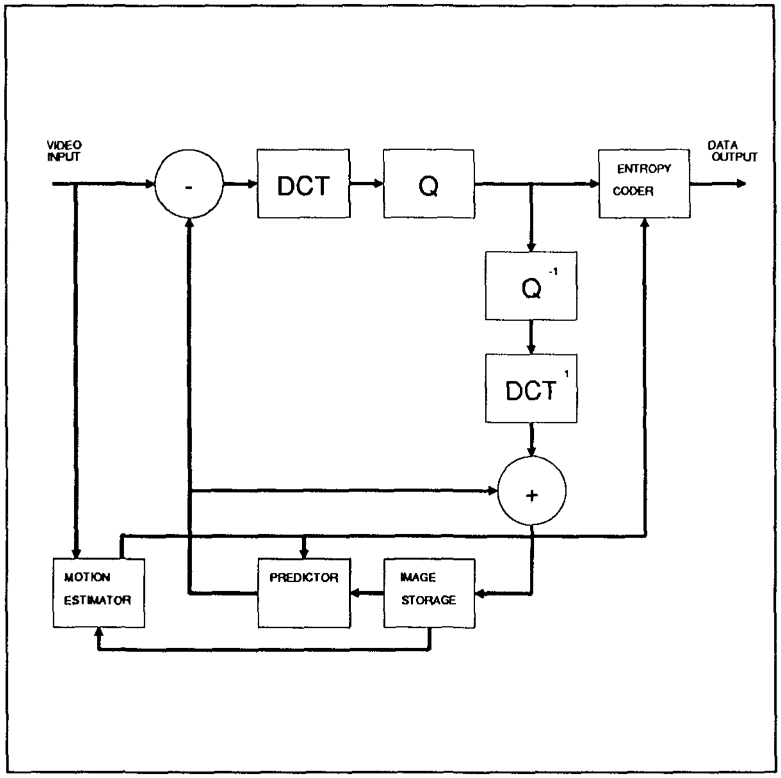
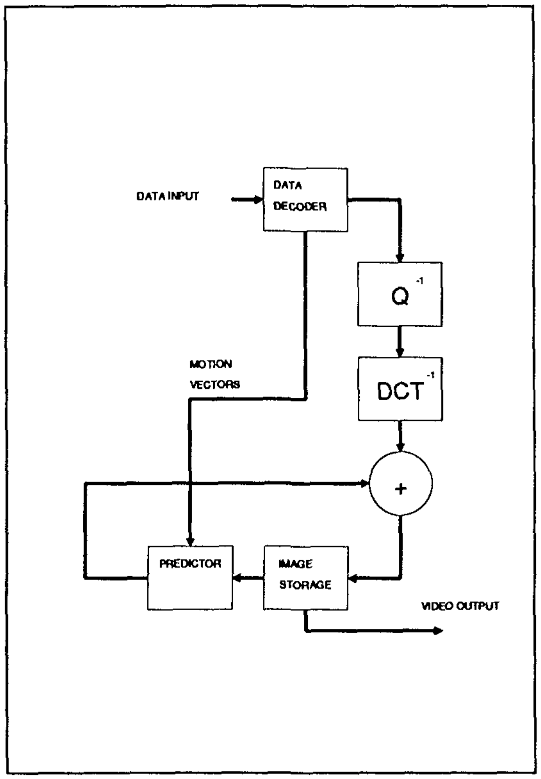


Figure 5 MPEG Encoder



MPEG Decoder

Motion Compensation

MPEG uses motion compensation to improve the accuracy of predicted frames. As with any predictive coded system, the closer the predicted frame is to the present frame, the less energy in the error signal and the fewer bits which are required to encode it. If there is no motion between the frames there is no difference to encode.

Motion compensation is performed by dividing the image to be coded into 16 x 16 pixel blocks (macroblocks). Each block in the image is associated with an optimally matched block in the prediction frame. A motion vector describes the location of this optimum block in the prediction frame. This vector indicates how far in .5 pixel resolution the optimum block is from the macroblocks original position. Thus by transmitting the relatively few bits required for motion vectors, the prediction can be greatly improved.

IMAGE QUALITY

There are several factors which determine the quality of an MPEG video sequence which are not specified in the specification. These are:

1. Motion Vector Generation
2. Image Pre and Post Processing
3. Image type sequence control
4. Rate control

All of the above items are determined by the encoder and are interrelated.

Motion Vector Generation

MPEG syntax describes how motion vectors are to be encoded and how the decoder processes them, however the manner in which they are generated is left to the system designer. The more accurate the predicted frame generated by the motion vectors, the less information that has to be encoded in the error signal. This allows more accurate coding of the error signal. Motion Vector generation is one of the most computationally intensive aspects of MPEG encoding. An inexpensive encoder might not generate any motion vectors and it would still be MPEG, however the image quality would be poor. Likewise for point-multipoint systems where few encoders will drive many decoders, highly elaborate motion vector generation might take place to obtain the highest quality images possible.

Image Pre and Post Processing

The manner in which the source video is converter into a format suitable for MPEG and the manner in which the MPEG decoder output is converted to a format suitable for the display device is not specified in the MPEG specification. It will however have a great bearing on the image quality.

Image Type Sequence Control

An MPEG decoder should be able to decode I, P and B frames. The specification does not indicate an algorithm for choosing which types of frames should be coded. The sophistication of the algorithm which makes this determination in the encoder will have an effect on the video quality. While it is possible to encode a fixed pattern of I, P and B frames, an algorithm which adjusts this pattern according to the image source material has the possibility of giving better results.

Rate Control

Like JPEG, MPEG has a coder data rate which is dependent on desired compression levels and the subject material of the image. In order to make the system operate through a fixed data rate communications channel, the output data from the encoder is buffered and the input data to the decoder is buffered. The compression level is dynamically adjusted by changing the quantizer scale factor or

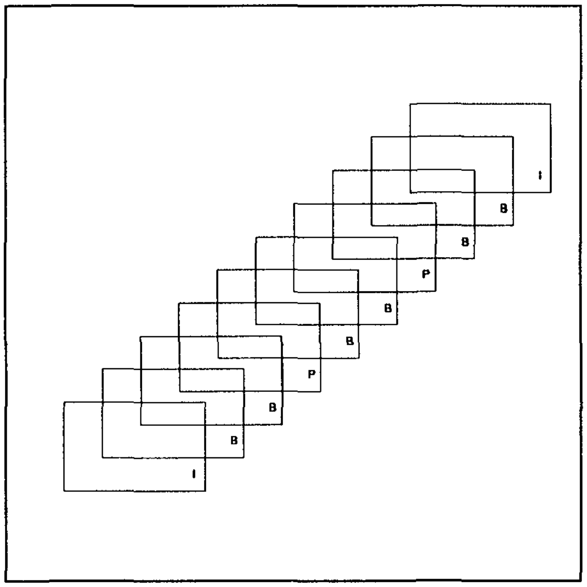


Figure 6 Typical MPEG Frame Sequence

changing the type of frame which is coded so that the average data rate into the buffer is equal to the channel data rate. This allows the data to be read out of the buffer at a constant rate without over or under running the buffer. The output data rate for MPEG is constrained to 1.5 Mbits/sec. While the syntax will support higher data rates, a system operating at these rates would be considered some form of MPEG extension.

The MPEG Encoder is typically significantly more complex than a MPEG decoder. This is because most encoders will generate motion vectors which is a computationally intensive process.

CONCLUSION

This has described the JPEG and MPEG specification as they stand at this point. It is unlikely that any significant technical changes will be made before the specifications are approved.

A MPEG specification is presently being written for audio data compression which will provide for variable quality audio of up to Compact Disk quality stereo. In addition, there is an effort under way to write an MPEG II specification which will provide near broadcast quality video at a higher data rate and handle interlaced pictures. All of these specifications will provide a basis for many video systems of the future.

(1) D.A. Huffman "A Method for the Construction of Minimum Redundancy Codes," Proc IRE 40, 1089, 1952.

(2) W. K. Pratt, "Digital Image Processing," New York: Wiley, 1978.

(3) K. R. Rao, "Discrete Cosine Transform," New York: Academic Press, 1990.

(4) ISO/CITT, "JPEG Draft Specification," May 1990.