An MPEG Standard Based Video Compression System and Applications

Mahesh Balakrishnan+, Weidong Mao+, Lionel Tranchard* and Etienne Fert*
Philips Laboratories, Briarcliff Manor, NY+.
Laboratoires d'Electronique Philips, Limeil-Brevannes, France*.

ABSTRACT

An MPEG based video compression system for delivering multiple channels of digitally compressed video over a single satellite transponder or a single cable channel is described. The video compression scheme is MPEG compliant and is based on the Discrete Cosine Transform and motion estimation followed by Huffman coding, and is described in detail. The system is capable of handling both interlaced video as well as progressive sources such as movie material.

INTRODUCTION

Recent years have seen significant developments in the area of digital compression as applied to video. These compression techniques have been applied to a variety of media including storage devices, satellite and terrestrial broadcast channels, and in cable and telecommunication networks. There has also been significant activity to establish image and video compression standards. The JPEG (Joint Photographic Experts Group) standard for compression of still images is currently in the final stages of standarization[1]. MPEG (Moving Picture Experts Group) is another body whose activities occur within the framework of the ISO (International Organization for Standardization)[2,3]. MPEG examines the issues involved not just in video compression, but also that of compressing associated audio and the system issues of audio-video synchronization. Differing industries such as the cable and telecommunications industry, the consumer electronics industry and the computer industry are looking at using the same technology in their applications. The use of a common standard will go a long way towards resolving the problem of interoperability of these systems. Furthermore, large volume production of standards based IC's can lead to substantial cost reductions, thus enabling even wider applications.

In this paper, we will give an overview of a video compression system based on the evolving MPEG (Moving Pictures Expert Group) standard, and describe its application to a cable network to provide multiple channels of compressed video. In recent months the cable industry has been examining applying digital compression techniques to satellite and cable, to increase the number of programs that can be carried over a single cable channel or a satellite transponder. The block diagram of such a system is shown in Figure 1. A program encoder compresses the video and audio associated with a particular program and multiplexes the output with the auxiliary data after separately encrypting them. Outputs of multiple program encoders are then multiplexed and transmitted over a single channel, after appropriate forward error protection and digital modulation. In an SCPC mode, the output of a single program encoder would be modulated on to one carrier. A cable head-end receives the compressed video, either over a satellite link as shown in Figure 1, or over a fiber-optic link. In either case, the received signal is demodulated, and then re-modulated for transmission over cable to the consumer's home.

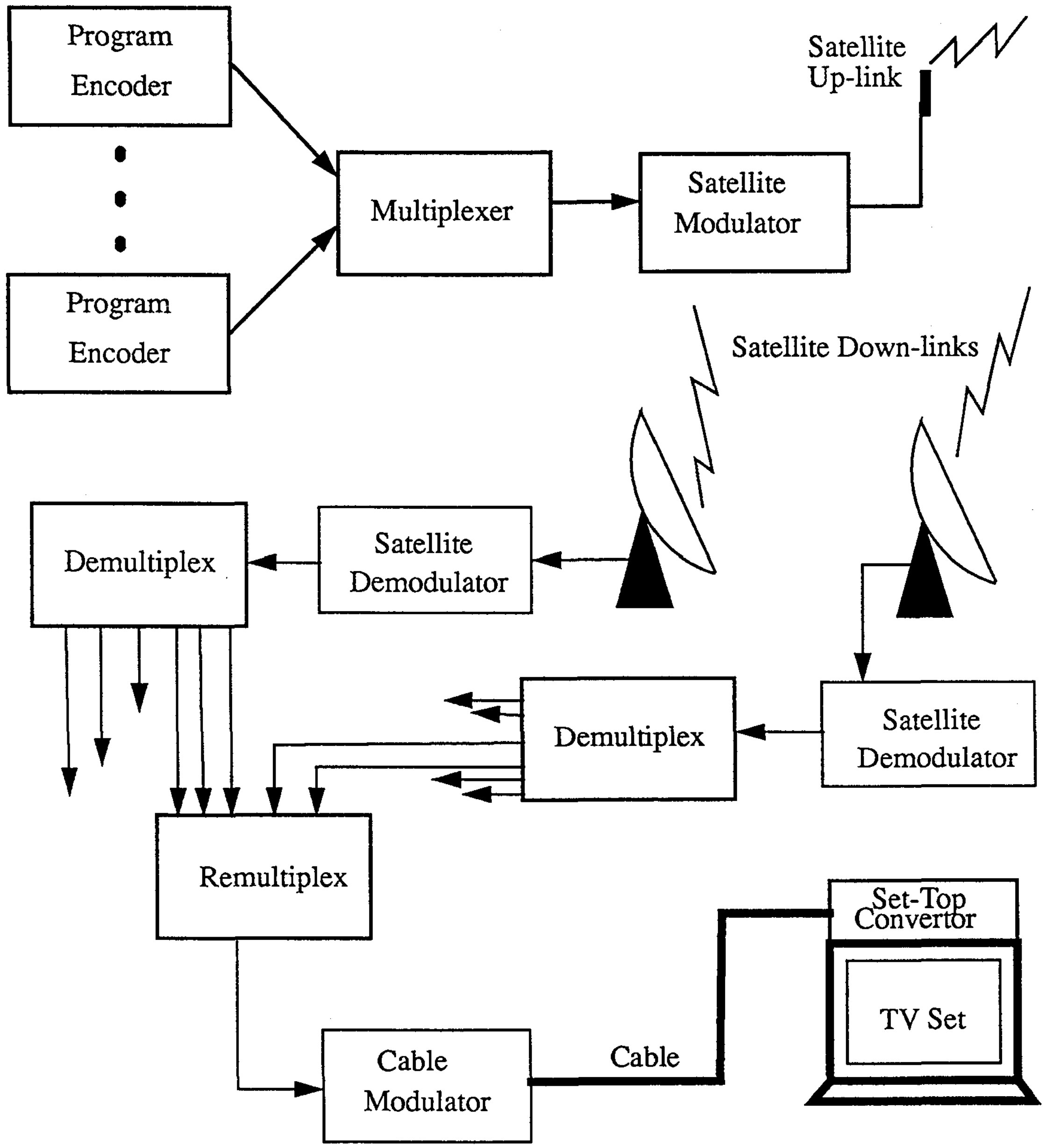

Figure1: Transmission link from satellite uplink through the head-end to the consumer's home.

The cable operator can also demultiplex the incoming signals and re-mix them in any appropriate combination, if it is so desired. The digital transmission techniques used for the cable and the satellite channels are very different and are so indicated in Figure 1. In the early stages of deployment of such a system, the cable operator can also decompress the compressed video and audio, at the headend, and transmit uncompressed analog video over cable, as is done today.

SYSTEM OVERVIEW

Figure 2 shows the function of each program encoder in greater detail. An individual program encoder compresses the video and audio associated with one program, separately encrypts them and the auxiliary data, and outputs a multiplexed stream. In the system we describe in this paper, the video and audio compression algorithms used are MPEG (Moving Pictures Expert Group) compliant. The MPEG video compression algorithm is currently undergoing standardization by the ISO (International Standards Organization), and is based on the Discrete Cosine Transform and motion compensation. The audio compression is provided by MPEG audio.

TABLE 1. System Parameters

Video	
Raster Format	525 lines 2:1 interlaced
Frame Rate	29.97 Hz
Aspect Ratio	4:3 or 16:9
Pixels per line	720
Lines per frame	480
Audio	
Number of channels	4 or more
Bandwidth	20 kHz
Sampling frequency	32, 44.1 or 48 kHz

The different encrypted bit streams are then multiplexed, as shown in Figure 2, to form the output of a single program encoder. The outputs of different program encoders can possibly have different bit rates, due to the variability in the program information, and are statistically multiplexed together to form a single bit stream. The multiplexer shown in Figure 1 therefore, also includes a transport layer protocol to allow for effective synchronization, and identification of the bit streams emanating from the different program encoders. At the cable headend, the multiplexed data is demultiplexed into different streams corresponding to the data output of each program encoder. These can now be remultiplexed in any combination desired. The different data streams multiplexed together can also originate from different satellite transponders as shown in Figure 1. This provides enormous flexibility to the headend programmer to selectively show certain programs in a particular viewing area.

Table 1 provides a listing of the video and audio related system parameters for the system.The video can be at either 4:3 or 16:9 aspect ratio. Four mono audio channels, each at 128 kbps are available for every video channel. These can be used to provide four separate mono audio tracks or two CD quality stereo pairs. The system transmits 30 Mbps of total data in the 6 Mhz cable channel, as well as the 24, 27 or 36 Mhz satellite channels.

TABLE 2. Total bit rates required for different levels of video quality.

Quality	60 Hz Video Source	24 Hz Movie Source
VHS	2.0 - 2.5 Mbps	1.5 - 2.0 Mbps
NTSC	2.5 - 6.5 Mbps	2.5 - 5.0 Mbps
CCIR 601	3.5 - 9.5 Mbps	3.0 - 7.5 Mbps

Table 2 lists the total bit rate that is required for 24 Hz movie and 60 Hz interlaced video, for three different levels of quality. VHS quality is equivalent to that which can be obtained on video played off a VHS tape, while NTSC quality is similar to that seen on today's TV sets. CCIR 601 quality refers to video at virtu-

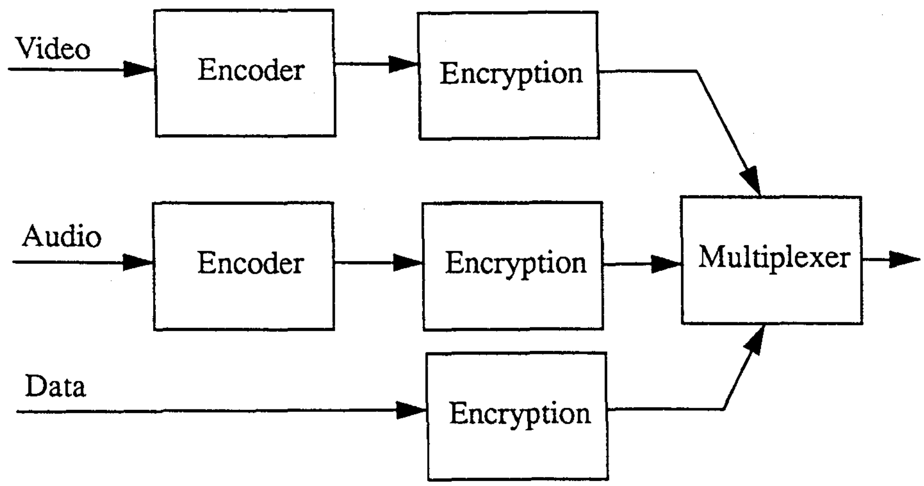

Figure 2: Block Diagram of the Program Encoder.

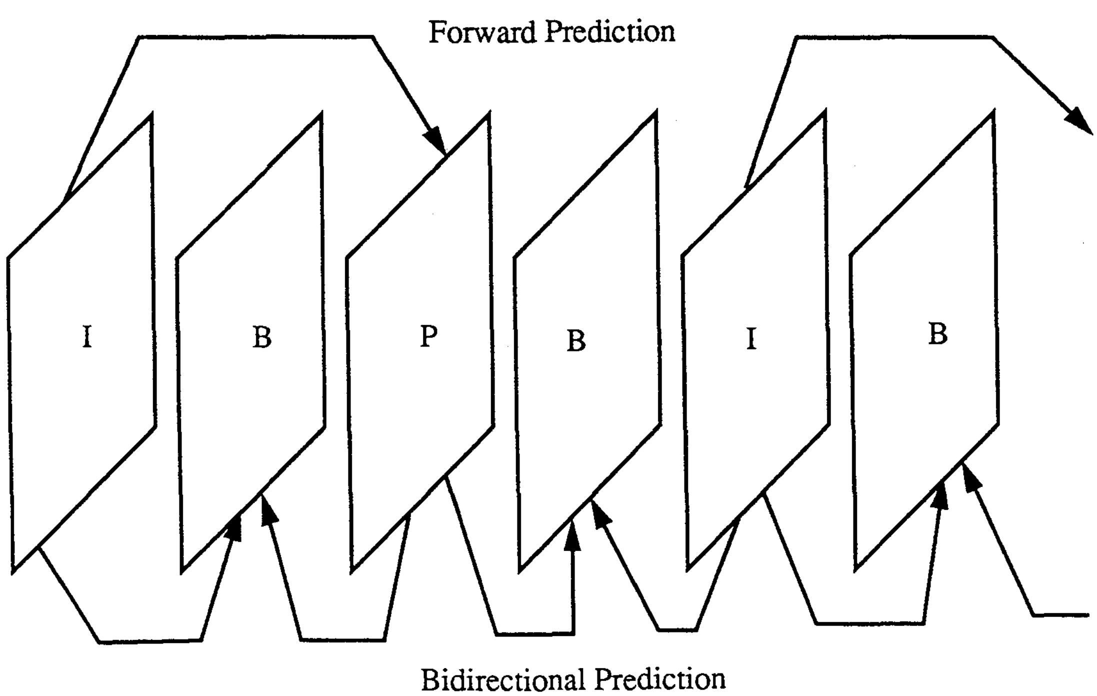

Figure 3: Frame structure in MPEG

ally studio quality. These rates include over 700 kbps of audio and data per channel.

VIDEO COMPRESSION AND PROCESSING

The input video to the video compression module of the system is of CCIR 601 resolution, since CCIR 601 is the standard for studio quality digital television. It also provides for easier exchange of material between the NTSC, PAL and SECAM television formats. For 525-line signals such as NTSC, each video frame has 720 samples per line, and 480 lines per frame. The input is also 2:1 interlaced. The video is converted to the YUV format, and the chrominance components subsampled to 240 lines per frame, and 360 samples per line. The YUV components are then compressed using an MPEG compliant algorithm.

The MPEG standard has been designed after extensive computer simulations, and provides a high degree of compression while maintaining good video quality. The proposed MPEG standard is based primarily on two compression techniques: the Discrete Cosine Transform (DCT), and motion compensation. Motion compensation attempts to exploit the redundancy between frames, by estimating the direction of motion of blocks from one frame to another, while the DCT reduces the spatial redundancy that exists between samples.

MPEG differs from other video compression schemes in the way it implements motion compensation. Unlike many other video compression schemes that use frames in the past to predict frames that occur later, MPEG uses frames from the past as well as from the future (these frames are collectively termed reference frames) to perform motion compensation. The use of non-causal and bidirectional motion compensation is one of the strengths of the proposed MPEG standard, and sets it apart from other video compression techniques.

The use of motion compensation implies that the decoder requires the reference frames in order to completely reconstruct the motion compensated frame. To allow a decoder to start decoding from the middle of a sequence (as in the case of a channel change), it is necessary to have frames that use no motion compensation. To support this and to restrict the propagation of any bit errors that occur during transmission, MPEG compresses frames in one of three different ways. Intra-coded or I-frames use no motion compensation at all and use only the DCT to perform data compression. As such, they can be decoded independent of other frames. Predicted or P-frames use motion compensation, but use only those frames that occur in the past as reference frames. Typically, this is the closest I or P frame. Finally, the interpolated or B-frames (B for bidirectional prediction), use frames from both the past as well as the future as reference frames in its motion compensation. This enables the encoder to predict uncovered areas, as well as objects that have recently moved into the frame. The reference frames used by the B-frame are the nearest I and P frames.Since B frames require far fewer bits than I or P frames, it is possible to get increased data compression while maintaining excellent video quality. Figure 3 illustrates one possible sequence of I, B and P frames in MPEG.

Figure 4 shows a block diagram of a typical MPEG encoder. The input video is converted to YUV component video. The chrominance components are then decimated to be a quarter of the resolution of the luminance. All three components are then grouped into blocks of 8x8. The four adjacent 8x8 luminance blocks comprising a larger 16x16 block, and the two associated chrominance blocks comprise what is termed a macro-block. Motion estimation and compensation is done for each macroblock. As mentioned earlier, I, B and P frames are coded differently. In the case of I frames, the DCT is applied directly on a block by

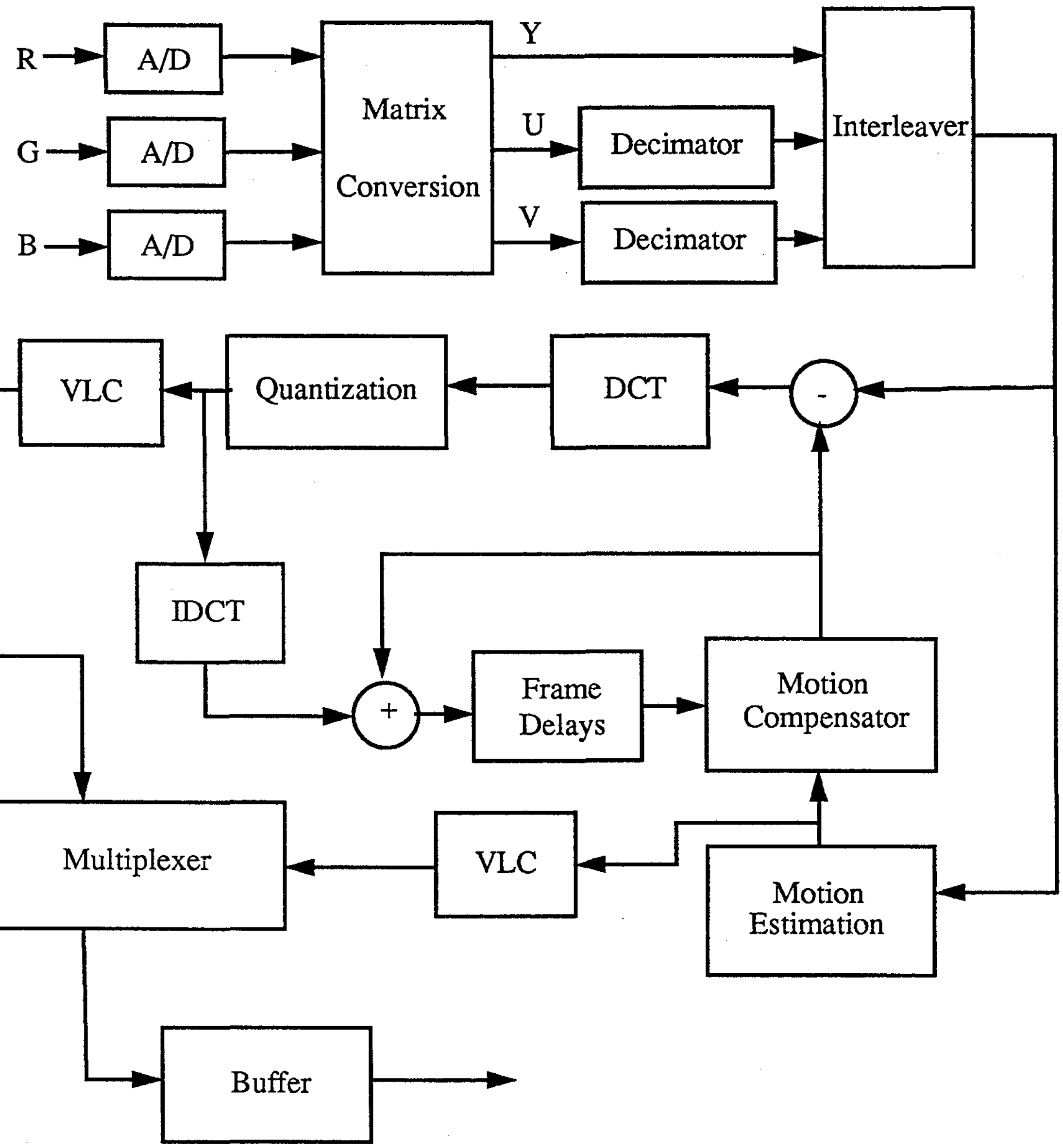

Figure 4: Digital Video Encoder Block Diagram

block basis. The DCT coefficients are then quantized, and coded using a variable length coder.

For B and P frames the DCT is applied on blocks obtained after motion compensation. Thus, the DCT is in this case applied on a difference signal. Motion compensation is done with respect to the reference frames (two reference frames for B-frames, and one reference frame for P-frames), and a "mode-selection" technique used to determine the best way to handle each macroblock. For instance, if the motion estimation is not successful, the algorithm intra-codes the macroblock using just the DCT. Furthermore, for B frames, motion-compensation could be done in a multitude of ways. The two reference frames can be used together to predict the macroblock. Or, the prediction could be done by using just one of the two frames. This is typically done when a scene change occurs between the frame being coded, and one of the reference frames. The quantized DCT coefficients and the motion vectors are then coded using a variable length coder, and transmitted.

Motion Estimation

Motion estimation is typically done on a macroblock basis, and involves determining for each macroblock in the current frame, the corresponding macroblocks in the reference frames. P-frames use one reference frame (the previous I or P frame) while B-frames have two reference frames (previous and future I or P frames). Motion estimation is one of the most computationally intensive operations that is required by MPEG. For each macroblock, it involves locating within a search window in the reference frames, the macroblock with the best match.An MPEG encoder can do this in many ways - full search, telescopic search and hierarchical being some of these.

Discrete Cosine Transform

On the B and P frames, the DCT is used to further reduce the spatial redundancy, subsequent to motion compensation. For I-frames, the DCT is applied directly on the frame. The DCT is an orthogonal transform, is filter-bank oriented, and thus has a frequency domain interpretation. This allows the system to exploit the properties of the Human Visual System (HVS). The HVS is less sensitive to high diagonal frequencies than it is to high horizontal and vertical frequencies. Thus, the DCT coefficients corresponding to diagonal frequencies are quantized more coarsely than are those corresponding to the horizontal and vertical frequencies. Furthermore, most naturally occurring images have higher energy distribution in the low frequency coefficients and less in the higher frequencies. To utilize this, the DCT coefficients are ordered according to the zig-zag scan shown in Figure 5. This one-dimensional ordering of the two-dimensional DCT coefficients increases the possibility of a large number of the quantized coefficients being zero. This is exploited by run-length coding the DCT coefficients as a "run of zeroes - amplitude" pair.

Movie Material

Movie material originates at 24 frames per second, and is displayed on television sets at 60 fields/second using a process termed three-two pulldown (Figure 6). The video encoder recognizes video material that has gone through a three-two pulldown process, and converts it back to 24 Hz. The material is then compressed and transmitted and the three-two pulldown done at the decoder, prior to display. This enables the encoder to take advantage of the lower frame rate as well as avoid an intra-frame scene change.

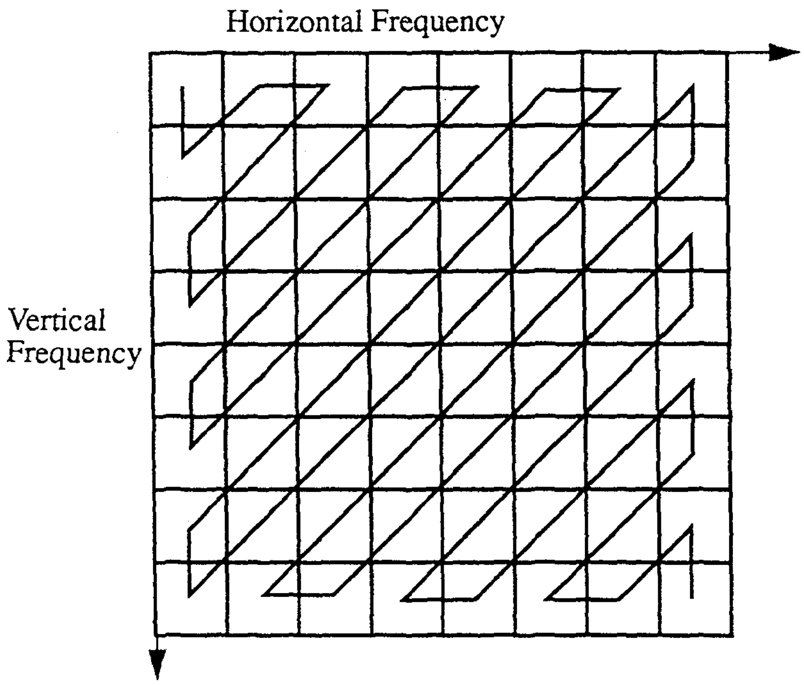

Figure 5: Zig-zag scan for the Discrete Cosine Transform

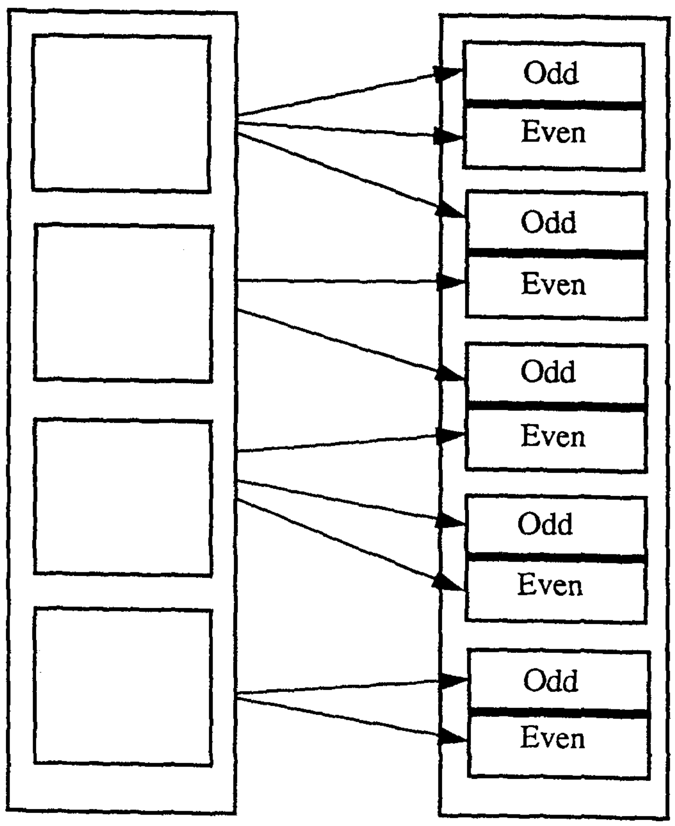

Figure 6: Three-two pulldown for movie material.

Error Concealment

The error rate for the system is restricted to less than one error event per day. However, even when the error does occur, it is concealed by the use of appropriate error concealment techniques. Since, the video compression algorithm uses frames from the past and the future to perform data compression, these reference frames are buffered at the decoder, in order to assist in the decoding process. Information from these frames is used to predict the pixel values in the current frame that have been lost due to transmission errors. Thus, the displayed video suffers minimal visible degradation.

REFERENCES

1. "JPEG digital compression and coding of continuous-tone still images", Draft ISO 10918, 1991.

2. Didier Le Gall, "MPEG: A video compression standard for multimedia applications", *Communications of the ACM,* vol. 34, no. 4, pp. 46 - 58, April 1991.

3. "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s", MPEG Committee Draft CD 11172, November 1991.