Video Compression

By David Strachan
Illustrated by Margarida DeBruin and Robert Marhong

Digital television is now commonplace in many television studios. Transmission of the digital signal is not practical, however, because the pictures contains too much data. The solution is to be found in video compression.

Why compression? Since the early days of the motion-picture industry, what viewers have perceived to be moving pictures have actually been sequences of still photographs. The trick, of course, is that the pictures change sufficiently often to convince the eye that what we are watching is real movement. In television, we don't even get to see one complete picture at a time. Designers rely on the persistence of vision of the eye to create the illusion of a complete moving picture, when all that is moving are the three electron beams in the cathode ray tube. The results are nevertheless quite acceptable.

With the advent of digital television, engineers have been tempted to stretch their powers of deception even further. The digital video signal is free of noise and is very robust, but alas, it occupies too much bandwidth for economical tape recording, let alone disk recording or transmission over the air. Fortunately, in most television pictures there is a lot of repetitive detail in plain backgrounds, blue sky, and common successive frames, which we can simply discard without the eye noticing that it has again been cheated. This process is called digital video compression.

Benefits

For those still questioning the need to convert their facility to digital video, an examination of what becomes possible with compressed video may help to make the path ahead look much clearer. If digital video signals could be processed in such a way as to enable them to be economically recorded on computer hard disks without any apparent loss of quality, then the possibilities for editing, painting, and animating would seem endless. Also, if digital video could be squeezed into the same bandwidth as that occupied by conventional analog signals, viewers could receive guaranteed studio quality pictures at home. All of this and more is made possible using compression techniques. But first you have to have a digital signal.

JPEG, MPEG, and MPEG-2

The computer industry has been using lossless compression techniques for many years to cram data onto hard disks and for transmission over modems. It was the computer industry that came up with the Joint Photographic Experts Group (JPEG) standard for compressing high-resolution digital still pictures, and it wasn't long before somebody thought it would be "cool" to show video on his computer too, so along came motion JPEG. The Moving Picture Experts Group (MPEG) was formed in 1988 to determine international standards for the digital compression of moving pictures, particularly to satisfy the growing interest in CD-ROMs.

Motion JPEG and MPEG techniques are now used extensively for computer imaging and can be very cost-effective for disk recording, CD-ROMs, etc., but neither offer optimum results for broadcasting. What has emerged, however, is MPEG-2 (an ISO/IEC ratified standard); the industry has adopted this at an amazing speed, driven almost entirely by the strong desire to provide viewers with a huge choice of programs delivered direct-to-home (DTH) via satellite or cable TV, using set-top decoders. MPEG-2 encoding is based on an integrated circuit chip set, pioneered by C-Cube Microsystems. Compression techniques are still evolving, but regardless of whether or not the system offers the ultimate method of compression for digital video, MPEG-2 has become so widely accepted that television engineers cannot afford to ignore this data-reduction system.

The World of MPEG-2

Here are some examples of companies who have adopted the MPEG-2 digital video compression standard. Nethold's Multichoice company launched its 20-channel digital TV for Belgium, The Netherlands, and Luxembourg (Benelux); Scandinavia, the Middle East, and Africa, on PanAmSat 4 in October 1995. The system will use over 1 million MPEG-2 set-top decoders supplied by Philips, Panasonic, and PACE. Echostar's DISH Television Network plans to offer 150 television channels. ExpressVu, the Canadian DTH operator, plans to launch a 100-channel MPEG-2 service. Galaxy, the Australian Pay TV operator, is installing a 10-channel digital satellite DTH system using two transponders on the Otex 3B satellite. The system uses MPEG-2 encoders with Irdeto conditional access (scrambling). Shinawatra Satellite, Thailand, has commenced tests of its MPEG-2 DTH broadcast system using Thailand's Ku-band Taicom 1 satellite.

At the time of writing, at least ten manufacturers are building MPEG-2 decoders, and many more are expected to follow. Both the European DVB and U.S. advanced television (ATV) systems will use MPEG-2 compression to deliver terrestrially transmitted digital television.

The Technical Challenge

As the purist will testify, you cannot compress video to any extent without throwing something away and thereby reducing the picture quality. Fortunately, however, the human visual system is incapable of absorbing all

This is one of a 12-part series of tutorials by this author and his colleagues. David Strachan is with Leitch Technology, Inc., North York, Ontario, Canada M2H 1E3.

of the material presented in a complex moving image, so by skillfully choosing compression techniques that selectively discard information the eye is unlikely to notice, impressive results can be achieved.

Each television picture comprises a finite number of tiny pixels. In the conventional 4:2:2 representation of NTSC and PAL television, there are 720 pixels along the active part of each horizontal line. In NTSC there are 486 active lines/frame (576 active lines in PAL) and 30 frames/sec (25 in PAL). Each pixel is made up of 8 bits for luminance and 4 bits each for the two color-difference signals (R-Y and B-Y, also known as C_r and C_b), a total of 16 bits. Therefore the bit rate for the active part of the video only is shown in Table 1.

The sole purpose of MPEG-2 is to reduce these bit rates to something more manageable, and its success relies on data reduction primarily in two areas of the motion picture. The first area is the information contained in each frame (spatial: relating to space, e.g., surplus blue sky, etc.); the second is detail, which does not change from frame to frame (temporal, relating to time).

Table 1 — Bit Rates for Active Part of Video

NTSC	720 x 486 x 29.97 x 16	168 Mbits/sec
PAL	720 x 576 x 25 x 16	166 Mbits/sec

Levels and Profiles

Much credit must go to the MPEG team for the international standardization of MPEG-2. The published ISO/IEC documents 13818 (-1 to -4) cover video and audio compression and the multiplexing structure needed for combining video, audio, and tim-

LEVELS	Spatial resolution layer	PROFILES: Simple	Main	SNR	Spatial	High
High	Enhancement		1920 x 1152 60			1920 x 1152 60
	Lower		–			960 x 576 30
High-1440	Enhancement		1440 x 1152 60		1440 x 1152 60	1440 x 1152 60
	Lower		–		720 x 576 30	720 x 576 30
Main	Enhancement	720 x 576 30	720 x 576 30	720 x 576 30		720 x 576 30
	Lower	–	–	–		352 x 288 30
Low	Enhancement		352 x 288 30	352 x 288 30		
	Lower		–	–		

Notes: 1920 x 1152 represents samples per line x lines per frame
60/30 represents frames per second.
The enclosed box represents conventional television (MP@ML)

Level	Low	Main	High-1440	High
Mbits/s	4	15	60	80

Levels and bit rates - main profile

Figure 1. Profiles and levels.

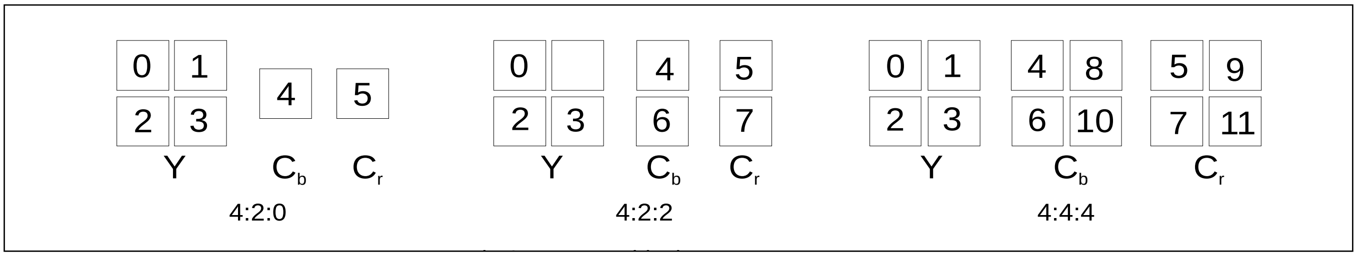

Figure 2. Macroblock structures.

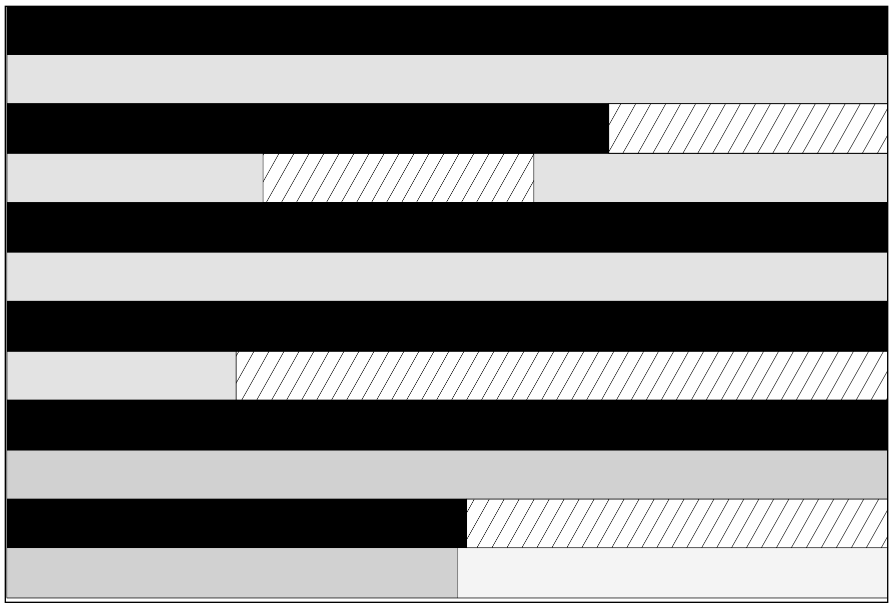

Figure 3. Slices of macroblocks.

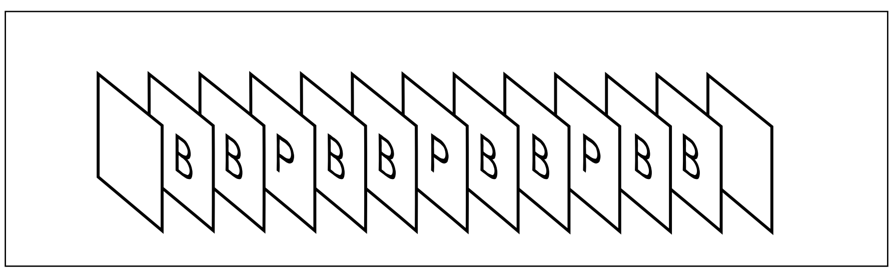

Figure 4. Typical group of pictures (GOP).

ing data for successful reproduction of video with synchronized audio. Not only is MPEG-2 truly a world standard, but the system encompasses everything from computer compressed data rates of less than 4 Mbits/sec, through conventional TV at 10 to 15 Mbits/sec and high-definition television operating at up to 80 Mbits/sec. These are known as different levels, and the MPEG-2 architecture supports all the levels shown in Fig. 1.

MPEG-2 also provides for flexibility in the type of compression used for each level. Compression types are known as profiles and may vary from use of the full 4:2:2 signal at the high end to the elimination of complete frames at the simple end. Encoders can vary considerably depending upon the application, so details of the encoding scheme must be transmitted along with the data, to enable the decoder to reconstruct the signal. In this way encoders can be designed to handle the various levels using different profiles at the same time as keeping the cost of the decoders to a minimum for the desired application. Most 525 and 625-line broadcasting uses main profile at main level (MP@ML).

Layers and Scalability

One of the most ingenious features of MPEG-2 is its ability to transmit video signals of widely ranging quality. A relatively inexpensive MPEG-2 decoder can reconstruct a useful picture by using only part of the encoded video bitstream, the rest of the data being reserved for quality enhancements. Coded video data consists of a series of video bitstreams called layers. The first layer is known as the base layer, and this can always be decoded independently. The other layers are called enhancement layers.

These layers may be used for spatial, temporal, and other scalable extensions. (More information on this in next month's article on HDTV.) If there is only one layer, the coded video data is said to have a nonscalable video bitstream. If there are two or more layers, the data is said to have a scalable hierarchy. Scalability has a further benefit, in that it helps to make the video resilient to transmission path errors. Transmission paths with the best error performance can be reserved for critical base layer information, while the enhancement layer data can be sent over a channel with inferior error performance.

Video Bitstream

The video bitstream is made up of blocks of pixels, macroblocks (MB), pictures, groups of pictures (GOP), and video sequences, as follows:

- Block
- Macroblock
- Slice
- Picture
- Group of pictures
- Video sequence

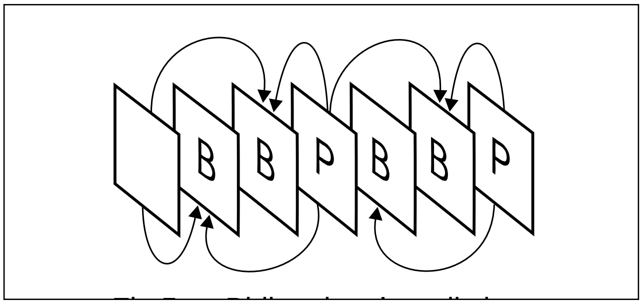

Figure 5. Bidirectional prediction.

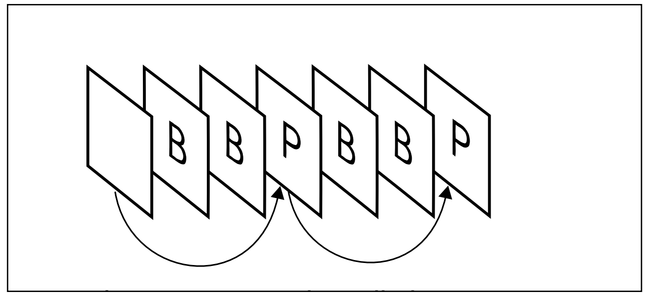

Figure 6. Forward prediction.

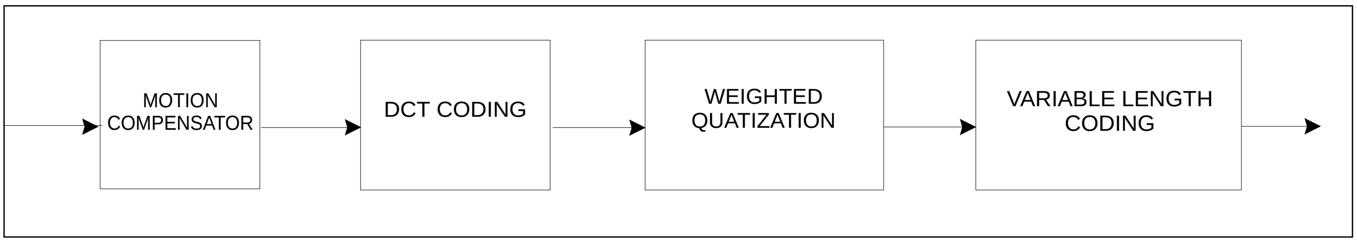

Figure 7. Video encoder.

The smallest element, a block, consists of 8 lines x 8 pixels per line. Blocks are grouped into macroblocks (MB), according to one of the MPEG-2 predefined profiles. The 4:2:0 macroblock format has 4 blocks for luminance, 1 block for C_r and 1 block for C_b. The 4:2:2 MB format has 4 luminance blocks, 2 C_r blocks, and 2 C_b blocks. The 4:4:4 again has 4 luminance blocks, but this version contains 4 C_r blocks and 4 C_b blocks (Fig. 2). As can be seen, a 4:2:2 MB will contain 8 blocks and therefore 8 x 8 x 8 (512) pixels.

Slices are strings of macroblocks arranged horizontally along the raster (Fig. 3). Slices can vary in length from a minimum of one macroblock to a maximum of one line. Pictures and groups of pictures will be examined during our discussions about temporal compression.

Temporal Compression

Temporal compression is designed to minimize the duplication of data contained in successive pictures. This is achieved by transmitting motion vector data and some difference information, rather the whole picture over again. To facilitate motion predicting, MPEG-2 separates the video into three types of pictures (Fig. 4):

- I (Intracoded) pictures
- P (Predictive coded) pictures
- B (Bidirectionally interpolated) pictures

I-pictures are the key reference for the other two picture types. They are derived by compressing the information in a single chosen field or frame (spatial compression). Still pictures are best preserved by using complete frames, but as the field rate is 2x the frame rate, movement is better served by using field-based pictures. Some MPEG-2 encoders are capable of analyzing the incoming video to determine the changes between successive fields. If there are no changes between odd and even fields, the encoder presumes that the two fields are part of the same frame and encodes them as such.

Changes between fields are noted and converted into motion vectors, which are encoded into data for later interception by the decoder. In this way, substantial bit rate reduction is achieved. The changes are transmitted in the form of P-pictures and B-pictures. P-pictures are predicted directly from the previous I-picture (Fig. 5). B-pictures are derived using either I-picture or P-picture information and these reference sources may be either ahead of or behind the B-picture being created (Fig. 6). Hence the term bidirectional interpolation. Both P and B-type pictures are also compressed spatially prior to transmission. The technique of motion compensation using the above method is known as temporal compression.

The three types of pictures are transmitted sequentially in a group of pictures, as shown in Fig. 4, with the first picture always being an I-picture. There are typically 12 pictures in a GOP, but as stated earlier, some encoders can detect changes between successive fields and, if the change is substantial, the encoder assumes that there has been a scene change, so it forces a new I-picture. This causes the sequence to start over again. The GOPs are sent in a video sequence, which contains data defining picture size, rates, and quantization matrices. The video sequence and all elements down to the slice size provide unique start codes to facilitate detection by the decoder.

The only drawback of generating these virtual pictures is that engineers have yet to find an easy way to edit on B or P pictures. Consequently television stations are likely to continue using motion JPEG techniques for in-house television contribution (as you can edit on any field), until a solution is found for this problem. Nevertheless, compression ratios of up to 10:1 can still be achieved using JPEG. Compression ratios in the order of 25:1 are achievable with MPEG-2. MPEG-2 is considered to be a distribution compression format.

Spatial Compression

The word spatial refers to the space in a single picture and the goal of spa-

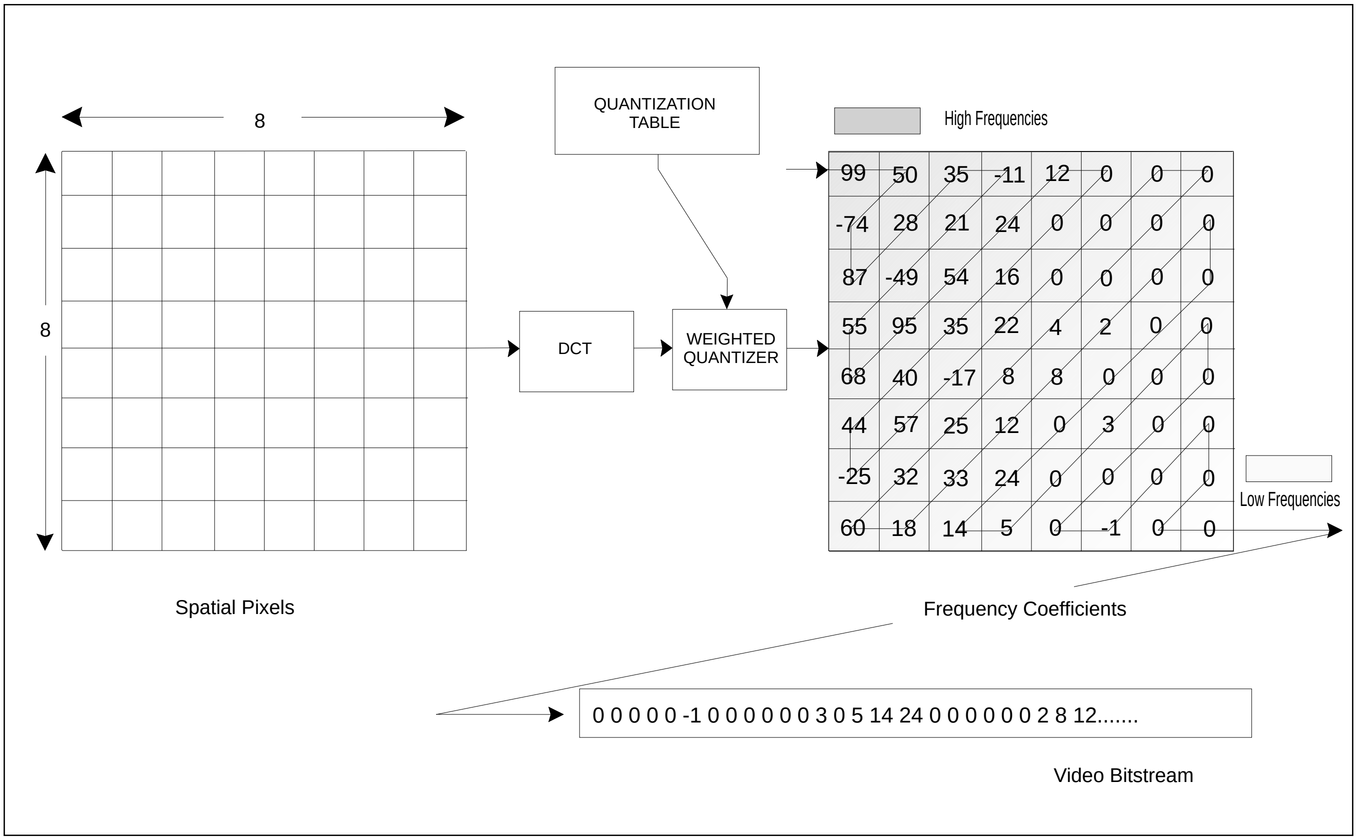

Figure 8. DCT coding and quantization (numbers are hypothetical).

tial compression is to minimize the duplication of data in each picture. Bit rate reduction in spatial compression is achieved by first transforming the video data from the space and time domain into the frequency domain using the discrete cosine transform (DCT) method and then applying quantization and variable length coding techniques to reduce the bit rate (Fig. 7).

Video is normally displayed on a time-based device, such as a waveform monitor, rather than on a frequency-based spectrum analyzer, but to accomplish data reduction, we must first transform the video data into the frequency domain. This is where DCT (a trigonometrical formula derived from Fourier analysis theory) is used to transform the data in each block of 8x8 pixels into blocks of 8x8 frequency coefficients. In the frequency domain, most of the high energy (and therefore most noticeable) picture elements are represented by low frequencies at the top left corner of the block, and the less important details are revealed as higher frequencies towards the bottom right (Fig. 8). Note that at this stage we have not yet discarded any bits.

After DCT encoding, the data is subjected to a quantization process, weighted to reduce data in the high-frequency areas, where the eye is less sensitive. We use more bits per pixel to quantize the important low-frequency coefficients and fewer bits per pixel for the high-frequency coefficients. The DC components are normally quantized at 10 bits, because if we employ coarser quantization of very low frequencies, the blocks themselves can start to become visible in the pictures. We have now achieved the first step in spatial bit rate reduction.

To create the compressed video bitstream, the 64 frequency coefficients are scanned in a zig-zag fashion from top left to bottom right and, as can be seen from Fig. 9, the high-frequency areas are represented by strings of zeros. Further data reduction can now be achieved by transmitting only the number of zeros instead of the usual values of the coefficients.

The last stage in the spatial compression process employs variable length coding (VLC). VLC assigns shorter code words for frequently occuring events and longer code words for less frequent events; it is also reversible. JPEG and MPEG systems use these methods of spatial compression for bit-rate reduction.

Program Streams and Transport Streams

So we have compressed the video. Now what? Before we can store or transmit the data, we have to multiplex the audio, video, and system information together. Figure 9 shows how this process works.

There are normally two audio/video multiplexers. One takes the video and audio packetized elementary streams and produces the program stream, and the other uses the same data to generate the transport stream. Program streams are normally reserved for robust transmission paths where errors are unlikely to occur. The program stream data packets may be of different lengths and can contain a relatively

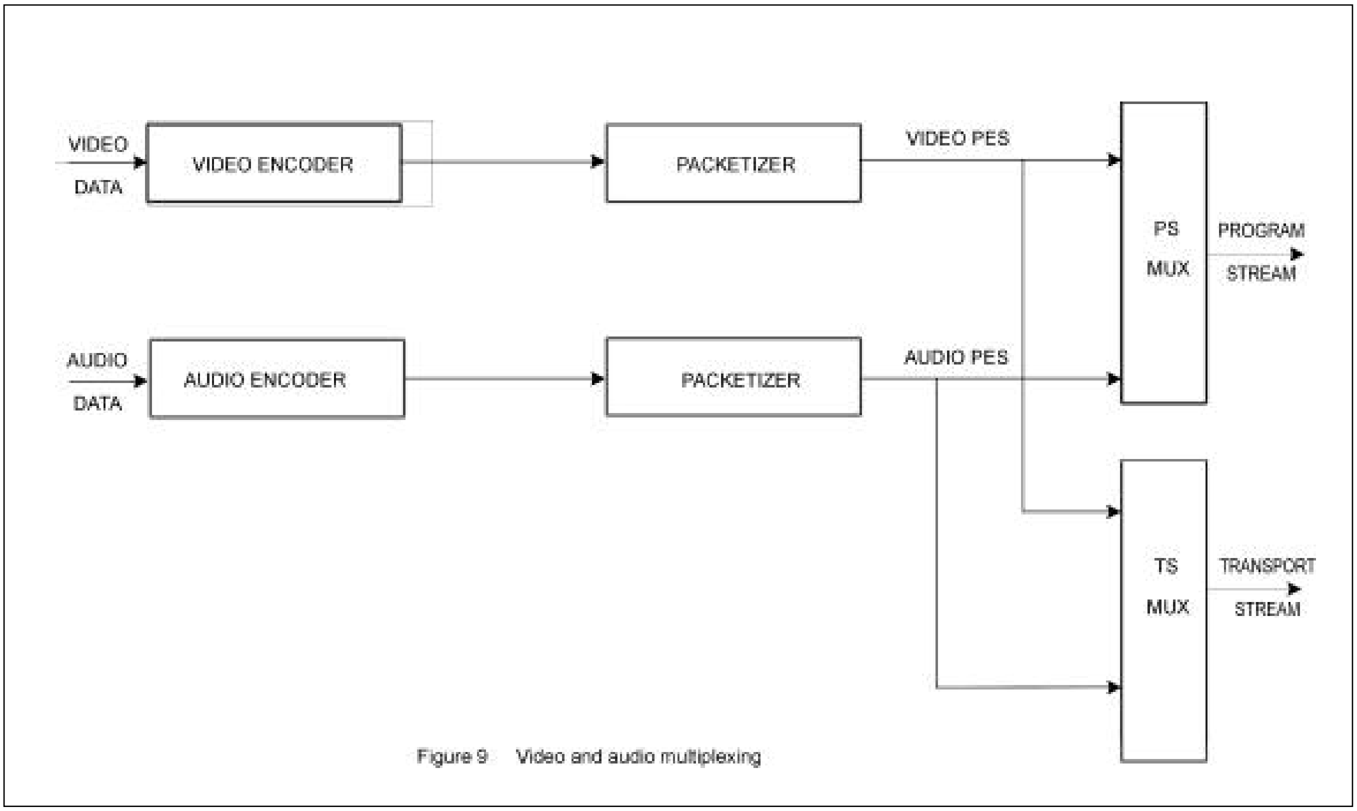

Figure 9. Video and audio multiplexing.

large number of bytes. A transport stream, on the other hand, can contain one or many programs with one or many independent time bases. Multiple TV channels can therefore be multiplexed together in this way. Transport stream packets are always 188 bytes in length. The transport stream is designed for use in environments where errors are likely to occur.

Conclusion

JPEG and MPEG compression techniques will continue to be used for low-cost computer compression requirements and wherever editing is needed in professional television.

MPEG-2 has become the international standard for video compression for any signals that are to be simply stored, distributed, and viewed. CD-ROMs are being developed employing MPEG-2 compression methods. MPEG-2 has been adopted worldwide as the compression standard for satellite delivered DTH television and for future cable and digital terrestrial television (DTTV), including high-definition television (HDTV).

Acknowledgments

Special thanks to colleagues Alan Plaunt and Kazem Najafi, at Leitch Technology, for their help and guidance.

Bibliography

Baron, S. N., and Wilson, W. R., "MPEG Overview," *SMPTE J., 103:*391-394, June 1994.

Holmes, S. J., "The Discrete Cosine: Transform Without Tears," *International Cable,* Feb. 1995.

ISO/IEC International Standard 13818-1, System Encoding section.

ISO/IEC International Standard 13818-2, Video Encoding section.

Stojancic, M. M., and Ngai, C., "Architectual and VLSI Implementation of the MPEG-2:MP@ML Video Decoding Process," *SMPTE J., 104:*62-72, Feb. 1995.

Video Compression — A Primer of Bit-Rate Reduction Technology for Video Signals, Sony publication.

THE AUTHOR

David Strachan is director of marketing for Leitch Technology, based in Toronto, Canada. Born in the U.K., Strachan obtained Grad. IERE membership in London in 1968. Since that time he has worked in the television broadcasting industry, first with EMI and Rank in the U.K. and for several transmitter and satellite equipment suppliers in Canada, prior to joining Leitch as international sales manager in 1985.