Image Coding Overview

Introduction

Today we are offered the ability to transmit pictures (images) either down a telephone line or over private leased lines. In the future, with the new switching and transmissions standards called SDH (Synchronous Digital Hierarchy) in Europe and the Far East, and SONET (Synchronous Optical Network) in the USA, we shall be able to hold videoconferences with friends and colleagues anywhere in the world. Videophones will become a standard feature and cable TV will be piped into the home using the same cable.

Let us first look at how our standard TV pictures are produced. This example describes the UK 625-line PAL (Phase Alternate Line) system. Three signals are used, one called luminance and two chrominance signals. The luminance signal represents the light / dark nature of the picture and the chrominance signals represent the color. The required frequency to represent all the information in the picture is high, as can be shown by the following simple analysis.

The TV screen is made up of 625 lines each with 625 picture elements in it. If the picture is scanned once every 20 ms, or 50 times per second, we can just about avoid the eye perceiving a 'flickering' of the picture. The number of elements in the picture is then:

$$N = 625 \times 625 = 3.9 \times 10^5$$

In the worst case we should have alternate elements black then white. We also include the screen aspect ration of 4/3 which affects the spatial and temporal frequency that the eye requires. The number of cycles in the picture is then:

$$N' = \frac{4}{3} \times \frac{N}{2} = 2.6 \times 10^5$$

For a frame rate of 50 per second, our information rate would be:

$$f = 2.6 \times 10^5 \times 50 = 13\ MHz$$

In practice only 575 lines are displayed, the remainder being used for flyback and synchronization. In addition, we must allow time for horizontal flyback and synchronization, all of which has the effect of reducing the maximum frequency to 11 MHz.

From this we can see that the required bandwidth for a television picture is very high—it is reduced by interlacing alternate lines and transmitting them only once every other frame. Although this brings the bandwidth down to around 6 MHz, we can easily see that video signals use a great deal more bandwidth than speech signals.

In current analog TVs the video signal is transmitted using amplitude modulation and the audio signal is transmitted using frequency modulation. To reduce the required bandwidth, the video signal is actually coded using vestigial AM, which transmits all of one sideband and a small amount of the low frequencies of the other sideband. In this way the low frequencies, which contain the most significant information, are preserved. Even so, the total bandwidth used is approximately 8 MHz.

If we wish to transmit the monochrome video signal as a digital waveform it would require a sampling frequency of around 12 MHz (twice the information rate of 6 MHz). In order to avoid brightness contours around areas of constant luminance, we need to use seven bits per sample to represent the signal. If we then add on extra bits

for complex coding algorithms and error correction, the result is a bit rate of around 120 Mbit/s! If we then take into account the chrominance signals, the basic bit rate would be in the order of 200 Mbit/s.

As we mentioned earlier, differential PCM can often be used with video signals due to the amount of correlation between adjacent samples. An alternative coding method is to transform the picture into a set of "transform coefficients." We could use FFTs, but this results in values which have real and imaginary parts and so produces just as complex a code as the PCM signal.

"Transform" coding makes use of the fact that most of the information is in the low-frequency components. A square picture transforms into a square set of transform coefficients (Figure 1) where the amplitudes at the bottom right hand corner are often so small that they can be ignored. The remainder are then encoded using ADPCM or DPCM, etc. Obviously after decoding, the receiver must perform the inverse transform to reconstruct the signal.

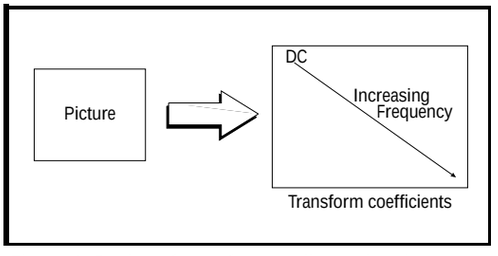

Figure 1: Basics of transform coding—the picture transforms into a set of "transform coefficients"

The most common technique in image coding is to use a discrete cosine transform (DCT) since it generates coefficients that are independent. In practice the picture is split into blocks 88 pixels in size, which are individually transformed. High-amplitude, low-frequency coefficients are transmitted first, producing progressive picture build up which is useful in bandwidth-limited systems. By using transforms and some form of DPCM, we can reduce the bandwidth required to transmit a full-screen image from 200 Mbit/s to around 34 Mbit/s.

Another popular encoding scheme is variable-length encoding or "entropy" coding. Entropy coding can either be used on its own or as an additional bit-rate-reduction method after DPCM. The technique exploits the statistical redundancy in a signal where all the possible types of output do not occur with the same probability.

The entropy coder has to construct output words that the decoder can interpret both in length and content. This is ensured by making the start of all the output words different. Several methods have been devised to do this, the most popular by Huffman [1951]. The entropy of a signal can be used as an indication of what levels of bit rate savings can be achieved using variable-length coding. The entropy of a signal is defined starting from the information content of a codeword, which is related in the following manner to the probability of its occurrence, P_i:

$$information = -\log_2 P_i$$

Therefore the information content of a rare codeword is very small. The entropy of a signal source is then defined as the average information content per codeword:

$$entropy = H = -\sum_i P_i \log 2 P_i$$

It has been shown that this entropy value represents the minimum number of bits per symbol into which the source can be coded assuming that the samples are independent. As we said earlier, cosine transforms generate independent coefficients and this is one of the reasons they are popular in image coding. A variable-length Huffman code gives an average word length that is close to the entropy value, implying a very efficient scheme.

A normal twisted pair of copper wires used for domestic purposes have a bandwidth of only a few kiloHertz. It doesn't take a mathematical genius to work out that this means that it is virtually impossible to transmit high-quality moving pictures using this channel. Nevertheless it is possible to produce some low-quality videophones that work over the telephone line.

Generally these videophones make use of the fact that the people using them are stationary, i.e., standing/sitting in one position. The amount of change in the picture from frame to frame is then proportionally small. By encoding only the difference between consecutive frames we can considerably reduce the required bit rate. In fact,

videophones use a method of encoding called motion vector estimation, which is based on this property.

At the beginning of this section we spoke briefly about the new transmission standards that will be in widespread use by the year 2000, i.e., SONET and SDH. These new standards are based on the use of very high bit rates of 155 MHz, 622 MHz, and 2.4 GHz transmitted over coaxial cable and fiber optics. These will bring the full realization of multimedia systems where images, speech, and data can be transmitted simultaneously over the same channel. The onset of high-bandwidth channels has lead the development of new standards for image and video transmission combined with speech. In fact, there are three standards which we shall briefly look at: H.261 from the CCITT, JPEG (Joint Photography Experts Group) for still images, and MPEG (Moving Pictures Expert Group), both from the International Standards Organization (ISO).

H.261 Video Compression

In the late 1970s, the European telecommunications industry identified the need for international collaboration on the implementation of audio visual services. The research project COST211 (Co-Operation in the field of Scientific and Technical research) and the CEPT (Conference on European Posts and Telecommunications) Working Group TR1 resulted in a European standard videoconferencing codec specification for the transmission of 625-line, 25-pictures/s PAL television at 2 Mbit/s. The demand for the same services in North America also prompted the development of a compatible codec at 1.544 Mbit/s for 525-line, 30-pictures/s NTSC systems. Standard conversion between PAL and NTSC was incorporated in the codec.

This system concept lead to the CCITT recommendations H.120 and H.130. In the 1980s, not all codecs manufactured complied to these standards, partly because only the Europeans were involved in the systems definition. In the USA and Japan, proprietary systems arose that were adopted by corporate users, who then were unable to interface to the standard codecs.

The SGXV/1 Specialist Group was set up to develop a worldwide standard at 384 kbit/s and its multiples. The H.261 standard developed an algorithm for px64 kbit/s which covers bit rates from 64 kbit/s up to 2 Mbit/s. The lowest frequency can be used for videophones running on narrowband ISDN (Integrated Services Digital Network).

The H.261 scheme uses a hybrid DPCM/DCT technique with motion compensation. Figure 2 shows a simplified diagram of the scheme. The luminance signal is first sampled at 6.75 MHz and coded with 8 bits, the chrominance signals are then sampled at 3.375 MHz. The difference between the present frame and the previous frame is then split into 88 blocks on which DCTs are performed. The coefficients are then coded using a variable-length Huffman coding algorithm to reduce the amount of data. An inverse DCT is performed on the quantized coefficients and the previous frame is re-added. This results in a picture very similar to the original which can be stored for subsequent use in the next frame.

The motion vector compensator works on the principle that an object which moves from one frame to the next can simply be represented by the displacement and direction in which it moves, i.e., the vector, and no information about the object needs coding. Obviously this is an over simplification since the object may also change in some way during the translation.

The motion compensator takes each 88-pixel block and searches the previous image by moving the block 15 samples in the horizontal and vertical directions in an attempt to find the best match. The variable store shown in Figure 2 is adjusted to use the best approximation for the subtraction from the present frame and the resultant vector is passed on to a coder for transmission. Figure 3 outlines a simplified diagram of the H.261 decoder.

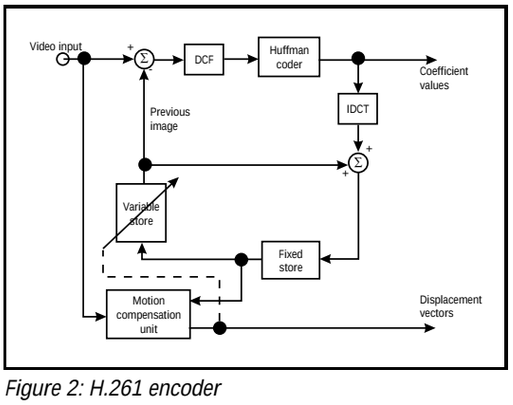

Figure 2: H.261 encoder

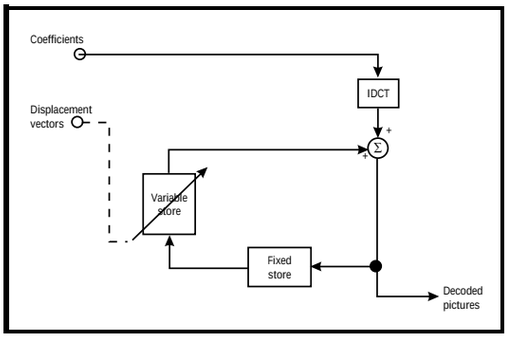

Figure 3: H.261 decoder

H.261 leaves manufacturers the ability to design codecs for various applications, e.g., videoconferencing, videophones, etc. at different bit rates. A videophone will probably only use a small screen and therefore the number of pictures per second required, the number of bits / pixel, the number of pixels, and the coarseness of the motion compensation can be adjusted to give a much lower bit rate than for a videoconferencing system which would use a large screen and require a higher-quality image.

There is equipment available which conforms to this standard—for example British Telecom—have an H.261 videoconferencing system. The algorithm is computationally intensive and hence most applications use a large amount of processing power. Videophones generally use multiple fixed-point DSPs to perform the DCT and DPCM coding for low bit rate, 64-kbit / s systems.

JPEG

The Joint Photographics Expert Group (JPEG) proposed standard is aimed at still picture compression. It is a transform-based coding algorithm which is applicable to any type of composite color system, e.g., the standard television luminance and chrominance signals, or a system that is based on the primary colors: red, green, and blue. Each color component is transformed by DCTs on 88 blocks and the resulting coefficients are quantized using a system dependent on the component and the frequency (Figure 4).

The frequency dependence allows high-frequency components to be encoded with a smaller number of bits than the more important lower-frequency information. JPEG also allows the designer to assign different quantization schemes to the different components ensuring the most important has the highest number of bits—similar in effect to the H.261 case where the chrominance signals were sampled with only half the frequency of the luminance signal.

The coefficients are reordered into a single stream in a zigzag fashion (Figure 5), and the whole stream is Huffman coded to reduce the amount of data to be transmitted. The DC term is differentially encoded with the previous frame before Huffman coding in order to reduce the difference between this value and the subsequent higher-frequency coefficients.

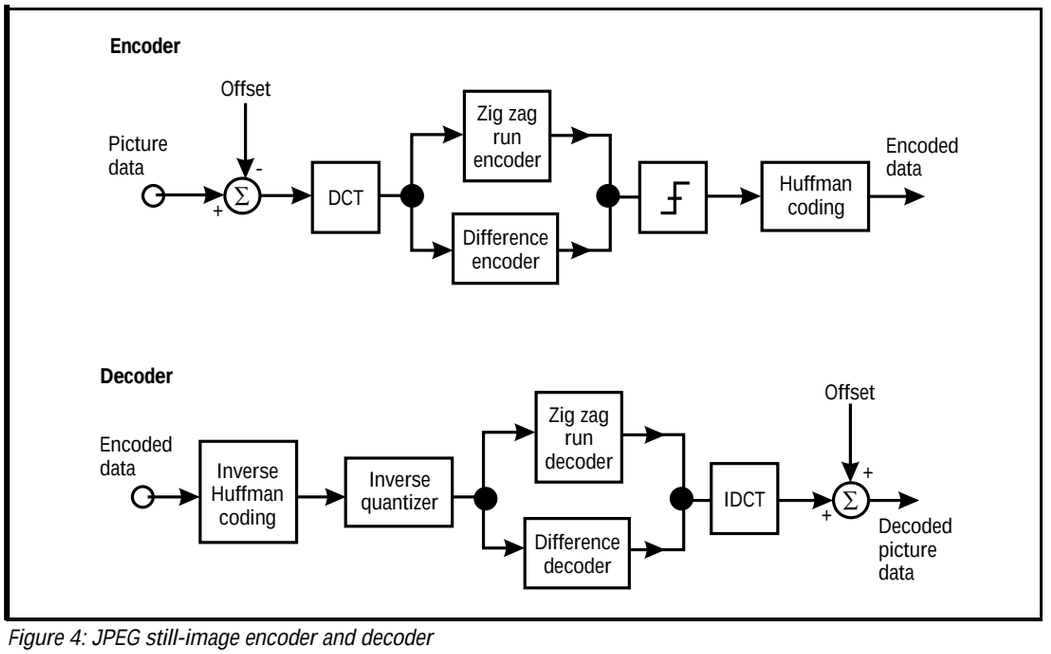

Figure 4: JPEG still-image encoder and decoder

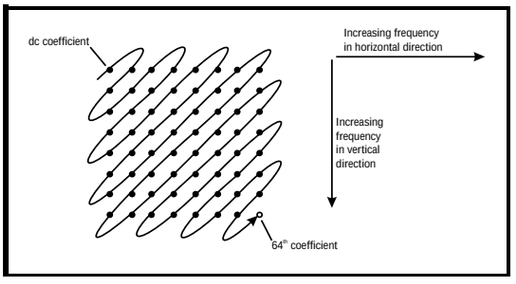

Figure 5: Zigzag coding reorders coefficients into a single stream

The JPEG coder is simpler than the H.261 system although they contain many similar elements. On the other hand, the JPEG decoder is more complicated. In general, the main difference between the two is the use of differential encoding and motion compensation in the H.261, the JPEG system simply resets after each transmission. This is not unreasonable when we think of the aim of this standard—still picture transmission, it does not expect another frame.

MPEG

The Moving Pictures Expert Group (MPEG) proposed standard (MPEG 1) is aimed at full-motion compression on digital storage media. It is similar to H.261 in that it uses inter-frame and intra-frame techniques to reduce the bit rate. As it is aimed at digital storage media such as CD-ROM, it cannot exceed their present limit of 1.5 Mbit / s.

MPEG uses the same type of intra-frame compression as JPEG and taking into account the inter-frame compression, is more than twice as effective at compression than JPEG. On the down side, it is more complex than JPEG, needing significantly more computing power.

Once the picture has been split into blocks and the DCT performed on each of the blocks, the coefficients are coded with either forward or backward predictive coding or a combination of both. After this, the blocks are uniformly quantized using a matrix of quantization steps chosen to achieve the desired output bit rate.

Although JPEG and MPEG were devised to solve two different tasks, JPEG also has some advantages in video compression. First of all it is a symmetrical technique, so that the same method is used to decompress the information as to compress it. In this way we can build a single-compression processor to perform both tasks. In some video teleconferencing systems, JPEG is being used for this reason.

References

Bonomi, M. [1991]. Multimedia and CD ROM: An Overview of MPEG and JPEG, *CD ROM Professional,* November 1991, pp. 38–40.

Gonzales, C.A. and Viscito, E. [1991]. Motion Video Adaptive Quantization in the Transform Domain, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 1, No. 4, December 1991, pp. 374–378.

Huffman, D.A. [1951]. A Method For The Construction of Minimum Redundancy Codes, *Proceedings of the IRE,* No. 40, pp. 1098–1101.

Kenyonm N. and Nightingale, C. [1992], *Audiovisual Telecommunications,* Chapman & Hall.

Sandbank, C.P. [1990]. *Digital Television,* John Wiley and Sons, NY.

Contact:

Excerpted from *A Simple Approach to Digital Signal Processing*
by Craig Marven and Gillian Ewers
Texas Instruments
ISBN 0-04047-00-8