SANDIA REPORT
SAND97–8274
Unlimited Release
Printed November 1999

Digital Video Technologies and their Network Requirements

R. P. Tsang, H. Y. Chen, J. M. Brandt, J. A. Hutchins

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of
Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (703) 605-6000
Web site: http://www.ntis.gov/ordering.htm

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A01

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

SAND97-8274
Unlimited Release
Printed November 1999

Digital Video Technologies and their Network Requirements

Rose P. Tsang, Helen Y. Chen, James Brandt, Jim Hutchins
Security and Networking Research Department
Sandia National Laboratories,
Livermore, CA, USA

Abstract

Coded digital video signals are considered to be one of the most difficult data types to transport due to their real-time requirements and high bit rate variability. In this study, we discuss the coding mechanisms incorporated by the major compression standards bodies, i.e., JPEG and MPEG, as well as more advanced coding mechanisms such as wavelet and fractal techniques. The relationship between the applications which use these coding schemes and their network requirements are the major focus of this study. Specifically, we relate network latency, channel transmission reliability, random access speed, buffering and network bandwidth with the various coding techniques as a function of the applications which use them. Such applications include High-Definition Television, Video Conferencing, Computer-Supported Collaborative Work (CSCW), and Medical Imaging.

3/4

Table of Contents

1. Introduction 7

2. Video Encoding Techniques, and Display and Compression Methodologies and Standards 7

2.1 Video Encoding Techniques 8

2.2 Video Display Methodologies and Standards 9

2.2.1 Current Broadcast-Quality Television 10

2.2.2 Studio-Quality Digital Television 11

2.2.3 High-Definition Television (HDTV) 11

2.3 Video Compression Techniques and Standards 12

2.3.1 Human Perception Properties 13

2.3.2 JPEG 13

2.3.3 H.261 and H.263 15

2.3.4 MPEG-1 and MPEG-2 16

2.3.5 Wavelets 19

2.3.6 Fractals 20

3. Relationship Between Applications, Coding Techniques, and Network Requirements 21

3.1 Applications 22

3.1.1 Television 22

3.1.2 Video Conferencing and Computer-Supported Cooperative Work (CSCW) 24

3.1.3 Medical Imaging 25

4. Bibliography 27

1. Introduction

The past decade has seen the emergence of a wide variety of distributed multimedia applications. Such applications include collaborative systems, video teleconferencing, video on demand, telemedicine and distance learning. These applications produce a variety of data types, each with their own distinct Quality of Service (QoS) requirements necessary for their appropriate transport, storage, display and manipulation.

The most common classification of multimedia data types is discrete versus continuous. Discrete data types occur sporadically, oftentimes initiated by a user. Examples of discrete media include textual data, still images and graphics. In this study, we will concentrate on continuous media data types. Forms of continuous media data include digital speech, digital video and computer animation. Such data types require continuous periodic network delivery in order for the user not to perceive a degradation in the QoS, e.g., jerkiness due to the erratic delivery of video frames.

Full-motion, high-quality motion images are perhaps the most demanding in terms of transmission speeds, processing overheads and storage requirements. Digital video data is considered to be one of the most difficult data types to transport due to its real-time requirements and extremely high bit rates. A single uncompressed digital TV stream requires 140 to 270 Mbps; a single uncompressed High Definition Television (HDTV) stream requires up to several gigabits per second. Moreover, many multimedia applications will be generating a substantial number of these digital video streams simultaneously. Clearly, compression schemes are critical to the viability of digital video services.

Fortunately, video signals contain a substantial amount of intrinsic redundance, and the human visual system is less sensitive to certain properties found in video signals, thus compression techniques are almost always employed. Several compression standards have been developed the past decade. The major ones have been developed to conform to various display quality standards such as HDTV, NTSC, PAL, SECAM and various ITU video-conferencing formats. In this paper we will discuss the mechanisms for still and motion picture compression concentrating on two standards which serve as typical examples of current compression schemes; they illustrate the successive steps and resulting complexity of the image compression process. These two standards are the Joint Photographic Expert Group (JPEG) for still image compression and the Moving Pictures Expert Group (MPEG) for moving image compression.

With the maturing of these standards and the development of more advanced second generation digital compression schemes, software and hardware products conforming to these standards as well as utilizing the more advanced compression techniques are readily available. These products will be widely used in the next few years and should facilitate the wide-spread deployment of many distributed multimedia applications. Understanding the characteristics of these digital compressed video streams is crucial to the efficient design and development of networking systems.

In the following section, section 2, we will discuss the major types of digital encoding techniques, basic television display mechanics, several display standards, and several important compression standards - JPEG, MPEG, H.XXX, wavelets and fractals. In section 3, we will discuss the selection of appropriate coding techniques as a function of various applications' requirements and network capabilities. We will also discuss the requirements of several major video-based applications.

2. Video Encoding Techniques, and Display and Compression Methodologies and Standards

The process of taking an analog signal, say from a camera, and transforming it into a digital image of a particular quality, say of conventional broadcast TV-quality, is accomplished through encoding or digitizing the analog signal, and then compressing the digitized signal according to the desired display quality. In this section, we will discuss basic video encoding techniques, several widely-used display standards, and the compression standards themselves.

2.1 Video Encoding Techniques

Encoding refers to the manner in which an analog signal is captured and digitized. *Pulse Code Modulation (PCM)* is the most widely implemented encoding system. PCM works in three phases. In the first phase, the analog signal is sampled at a certain rate. Then each sample is independently quantized; the quantum level of each individual analog sample is determined without taking into account the value of the preceding samples. In the last step, each quantized value is replaced by a binary code or code-word.

The following is a brief description of the more widely used encoding techniques

- *Differential or predictive encoding.* The principle of this method of encoding is that only the difference between the actual value of a sample and a prediction of that value is encoded. Differential encoding is best suited for signals in which successive values are significantly different from zero but do not differ too much from each other. This method is particularly well suited for video and audio signals where there is a large probability of redundancy between successive frames as well as redundancy within a frame.

The major categories of differential encoding schemes are *differential PCM (DPCM), delta modulation*, and *adaptive DPCM (ADPCM).* Different differential encoding schemes usually differ in the method used to derive the predicted value. For instance, in simple DPCM, the predicted value is simply the last sampled value. So at time *t* the difference between the value of the sample at time *t* (the current sample value) and the value of the sample at time *t-1* (the predicted value) is transmitted. Figure 2.1 depicts an example of DPCM encoding. The resulting differentially coded signal samples require fewer number of bits.

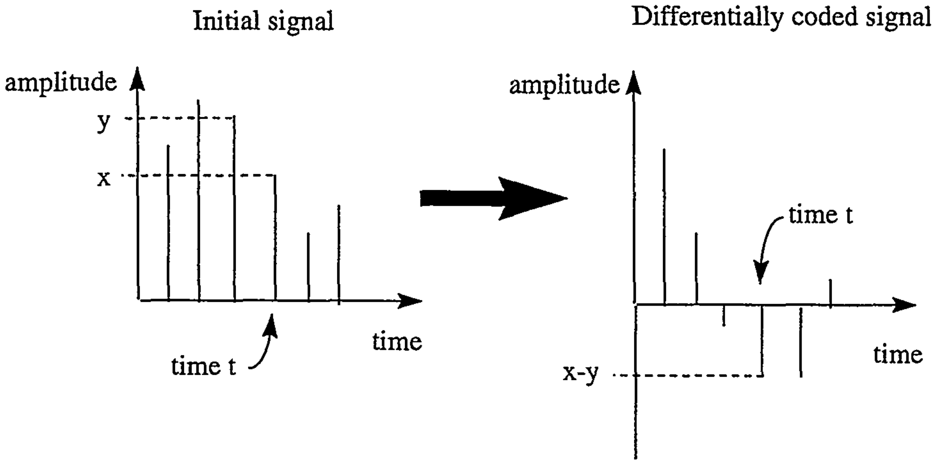

Figure 2.1 Example of differential pulse code modulation encoding

Delta modulation is based upon DPCM encoding. It varies from simple DPCM in that the difference between the predicted value and the current value of a sample is coded with a single bit. Obviously, this technique is most suitable for encoding signals whose successive sampled values do not frequently change. ADPCM is another form of DPCM. Its difference lies in using a variable estimation function based upon the short term characteristics of the sampled signal.

- *Transform encoding.* In this technique, the initial data is submitted as input to a mathematical transformation (e.g., Fourier transform or Discrete Cosine Transform (DCT)); the image is transformed from the initial spatial or temporal domain into an abstract domain which is more suitable for compression. The transformation to be used should be the one which provides the greatest amount of compression for the particular data type. For real-life (i.e., continuous-tone) images, that transformation is usually to the frequency domain via the DCT. Transform encoding, with respect to the DCT, is discussed in depth in section 2.3.2.

So far we have discussed some basic principles about the capture of analog data and their subsequent digitization. In the remainder of this subsection, we discuss the components of a digital television signal.

Analog capture devices, i.e., cameras, produce three distinct continuous signals. Each signal represents the intensity of either the red, blue or green component. Together these three signals are often referred to as the RGB signal set. In television [Fluckiger], the RGB signal set is usually translated into another set of three signals: a luminance signal and two chrominance (color) signals. The luminance signal carries information about the total amount of light energy. The two chrominance signals are derived from three computed color difference signals. Each element of a color difference signal is the color signal minus the luminance signal.

The chrominance signals and the luminance signal are produced through a linear transformation of the RGB signal set. For instance, the NTSC standard for analog broadcast television refers to the luminance as the Y component and the two chrominance signals as the I and Q components.

The primary motivation for transforming the RGB signals into a luminance signal and two chrominance signals is that the human visual system is less sensitive to color than to luminance. Thus the color signal may be represented with less accuracy resulting in reduced bandwidth and storage requirements. This results in the color components being subsampled; fewer samples per line are sampled, and fewer lines per frame are used. The standard notation for indicating the sampling ratios of all three components is:

$$Y\ sampling\ frequency\!:\ C_1\ sampling\ frequency\!:\ C_2\ sampling\ frequency$$

Different video display standards have chosen different subsampling ratios. These will be discussed in the next sub-section.

2.2 Video Display Methodologies and Standards

This subsection covers, in abbreviated form, background information on the electronic image-scanning techniques used today in television. This information is necessary because the development of conventional television to HDTV was an evolutionary process where each stage grew from the requirements and advances of the previous stages.

In conventional television transmission, the picture information must be carried as a one-dimensional signal. The process of taking picture information and converting it to serial form is called *picture scanning*. At the sender, the picture is scanned in a series of horizontal sweeps that encode the instantaneous luminance at each pixel. The receiver regenerates the picture by varying the intensity of the light beam projected onto a screen in proportion to the received signal.

Synchronization pulses, which are used to mark the beginning of each picture frame, and *blanking pulses*, which are used to turn off the beam until the scanning beam returns to the beginning of the next scanned line, are combined with the luminance signal. Usually when the term luminance signal is used, it is understood to refer to a composite video signal consisting of the actual luminance signal combined with the synchronization pulse and blocking pulse signals. In addition to the *horizontal blanking interval*, there is a *vertical blanking interval* corresponding to the interval when the beam is turned off while it moves from the end of one frame to the beginning of the next frame. The transmission of color information, i.e., chrominance, takes place via a color carrier in the video band, modulated simultaneously in frequency and amplitude, and added to the luminance signal. The audio signal is frequency modulated on another carrier. The signal is broadcast by amplitude modulating the composite signal on an appropriate carrier wave.

For reconstruction of the picture, an electron beam of a cathode ray tube (CRT) shoots electrons at the coated surface of the tube. The surface is covered with phosphor which glows when radiated by electrons. The image of the electron beam is usually called the dot. Once the dot reaches the right side of the CRT screen, the horizontal sync pulse is received by the TV set instructing the dot to travel back to the left side of the CRT screen, i.e., horizontal retrace, and print the next line. The horizontal retrace lasts for approximately 10.2 to 11.5 microseconds. When the dot reaches the bottom of the CRT screen, a vertical sync pulse is transmitted, instructing the dot to return to the top of the CRT screen. Based on this process, it is obvious that the absence of accurate system timing causes the video picture to degrade considerably .

Video display quality is a topic which immediately brings to mind a few well-known international display standards such as HDTV, PAL, SECAM, and NTSC. National Television Standard Code (NTSC), Phase Alternating Line (PAL), and SEquential Couleur Avec Memoire (SECAM) are the three major standards for existing analog broadcasting systems. High-Definition Television (HDTV) provides the major all-digital improvement in video quality from the above three analog standards. Although, in this paper, we discuss standards with regard to motion images these standards typically also define how audio is coded and transmitted, as well as the synchronization between the audio streams and video streams.

The Federal Communications Commission (FCC) has allocated a transmission spectrum of 6 MHz for a broadcast television signal. The bandwidth directly impacts the resolution of the video signal. For instance the 4.2 MHz NTSC signal supports a maximum of 336 lines of horizontal resolution. The NTSC compliant signal limits the quality of the received signal; even if high resolution images are created, say via a high resolution camera at the sender, the receiver will receive video quality equivalent to that of any other broadcast NTSC signal.

2.2.1 Current Broadcast-Quality Television

Conventional broadcast television services are based upon analog transmission. This means that transmission relies on the modulation of the amplitude of the frequency of carrier signals. The modulated frequencies and bandwidth are confined due to restrictions on occupation of the frequency spectrum and the attenuation of the signal.

As mentioned before, there exist three international standards for current analog television:

- *NTSC*, used mostly in the USA and Japan, is the oldest and most widely used television standard. It is based upon quadrature amplitude modulation. 4.2 MHz can be used for the luminance signal and 0.5 MHz for each of the two chrominance signals. NTSC specifies a motion frequency of 30 frames per second and a vertical resolution of 525 lines/frame.

- *PAL*, used mostly in Europe, Australia Africa and Asia is also based upon quadrature amplitude modulation. The frequency of the subcarrier is 4.43 MHz. PAL specifies a motion frequency of 25 Hz and a vertical resolution of 625 lines/frame.

- *SECAM*, used in France and Eastern Europe, is based upon frequency modulation, as opposed to NTSC and PAL. It uses two subcarriers: one at 4.25 MHz and one at 4.40 MHz. SECAM specifies a motion frequency of 25 Hz and a vertical resolution of 625 lines/frame (like PAL).

All three standards specify interlaced scanning (see section 2.2.3 for a discussion on interlaced vs progressive scanning), and aspect ratios, i.e., the ratio of the width to the height of a frame, of 4:3.

VCR-quality television is the quality observed when viewing a typical videocassette recorder of VHS quality which contains a recording of a broadcast-quality program. The resolution is approximately one half that of a PAL or SECAM broadcast-quality television program.

2.2.2 Studio-quality digital television.

The *ITU-CCIR-601* is a family of compatible standards which serves as a reference for the transmission of digital television signals. The parameters, such as resolution and frame rate, defined by the CCIR-601 standard, are compatible with the existing analog television standards, however, the perceived quality of studio-quality video is considerably higher than that of current analog broadcast television.

The CCIR-601 standard can essentially be considered a digital component version of the NTSC and PAL video display standards. It departs from the two analog video standards in two major ways (1) the signals are entirely digital, and (2) the signals are component signals rather than composite signals; the color information is represented by individual digital RGB component signals or by two chrominance component signals.

CCIR-601 images have the format of 525 lines and 858 samples per line, to accommodate the NTSC standard, or, 625 lines and 864 samples per line, to accommodate the PAL and SECAM standards. The number of samples per line only applies to the luminance component of each pixel. The sampling frequency is 13.5 MHz for the luminance signal. The two color components are typically sampled at half the frequency of the luminance signal, i.e., approximately 430 samples per line; the subsampling ratios are 4:2:2.

The frame rate for digital television is the same as that of current analog broadcast television - 25 or 30 frames per second. However, the CCIR-601 standard defines an interlace scan format, so the field frequency is 50 or 60 fields per second. CCIR-601 is intended for use within the commercial broadcast television industry for video transport between studios, control rooms and transmission facilities.

It is of interest to note that upon a cursory inspection of the parameters of digital television and analog television there should be no difference in perceived quality; the number of lines and the frame rate of both are identical. However, analog television is perceived to be of significant lower visual quality. One reason is that not all TV broadcast lines contribute to the user perceived image quality; some lines are used for control information, captions, etc. For example in an NTSC 525 line system, only 483 lines contain direct video information. Another reason is that the analog signal of each individual line does not have a resolution equivalent to that defined in the digital studio standard (CCIR-601).

2.2.3 High-Definition Television (HDTV)

HDTV [Steinmetz2] is forecasted to set the standard for the next generation of television applications. Its service is predicted to supersede the current NTSC (current broadcast television) display standard by the year 2008 [Minoli]. The following items distinguish HDTV from conventional television.

- *Resolution.* The HDTV standard defines several different image scanning formats. The *high resolution/high frame rate* format specifies images with 1920 pixels per line and 1080 lines per frame, at 60 frames per second. The *high resolution/conventional frame rate* specifies images with the same pixel dimensions but with a frame rate close to that of current broadcast television: 24 to 30 frames per second. The *improved resolution/conventional frame rate* format specifies images with 1280 pixels per line and 720 lines per frame, at 24 to 30 frames per second. This mode provides an intermediary format for a viewing quality which is between current broadcast television and actual *high resolution/high frame rate* video quality. These are only three of the HDTV formats which have been defined; many other HDTV formats have been defined and adopted in different countries. Hereafter, when we use the term *HDTV quality*, we will be referring to the *high resolution/high frame rate* format.

- *Aspect ratio.* Another important parameter of motion video formats is the aspect ratio. The aspect ratio is defined as the ratio of the width to the height of a frame. HDTV has adopted a 16:9 aspect ratio - conventional broadcast TV has a 4:3 aspect ratio.

- *Scanning techniques.* An issue which impacts the quality of video is whether or not images are scanned in the progressive mode or interlaced mode. Interlacing is used in current broadcast television. In interlaced mode, each frame is divided into two fields where each field contains alternating lines of the image. In conventional TV, where the frame rate is 30 frames per second, interlacing is used; 60 fields per second are being transmitted. Interlacing mode was used to reduce bandwidth requirements, however, the tradeoff is that the perceived resolution degrades by about one third [Fluckiger,1995]. Most computer display units use progressive scan techniques. In progressive mode, each frame is displayed, or refreshed, line by line where each line is scanned from left to right. In this case, there is only one field per frame. Given equal bit rates, progressive scan results in significantly better perceived quality [Fluckiger]. Progressive scan is usually assumed for HDTV quality.

2.3 Video Compression Techniques and Standards

Compression techniques fall into two categories: *lossless* compression and *lossy* compression. In lossless compression, data is compressed with a technique which allows the decompression process to reconstruct the identical original form of the data. Lossless compression is most suitable for applications which require the complete reconstruction of the original data, e.g., computer data, medical imaging. *Entropy coding,* a form of lossless compression, refers to techniques which do not consider the semantics of the data to be compressed. These techniques are usually based upon either the suppression of repetitive sequences or statistical information about the frequency of occurrences of pre-specified strings of bits. Typical compression ratios of computer data are 2:1 to 4:1. For image data, since the level of redundancy is usually higher, compression ratios may be as high as 8:1. Well-known examples of entropy compression techniques include run-length coding, Huffman coding, and arithmetic coding.

In lossy compression, the compression process alters the data in such a way that the decompression process can never reconstruct the data to its original form. Ideally, in lossy compression, the user should not notice that the received image has been altered. *Source coding* techniques, the most widely-used lossy compression method, exploit the semantics of the data in order to achieve compression. The degree of compression that can be achieved is a function of the content of the data stream. For instance, a particular content prediction technique may take advantage of the spatial redundancy found in still images. The Discrete Cosine Transformation (DCT) [Rao] is an example of a source encoding technique. In this case, the image is transformed from the spatial domain to the frequency domain. This transformation is used because it is known that the human visual system is more sensitive to certain ranges of frequencies, and depending upon the application, frequently it may not be necessary to display more information than the human eye or ear can detect. The DCT technique is used in the JPEG and MPEG standards and will be discussed in much greater detail in section 2.3.2. Table 2.1 shows the characteristics and examples of the different encoding techniques.

Encoding Technique	Coding Strategy	Loss Characteristics	Sub-categories and Examples	
Entropy coding	suppression of repetitive sequences, statistical encoding	lossless	run-length coding, Huffman coding, arithmetic coding	
Source coding	considers semantics of data stream	lossless or lossy	differential or predictive transformation layered coding vector quantization	DPCM, DM, ADPCM FFT, DCT, wavelets bit position, sub-sampling, sub-band coding fractal transforms
Hybrid coding	both of above strategies	lossless or lossy	JPEG, MPEG, H.261, DVI RTV, DVI PLV	

Table 2.1 Examples and characteristics of coding techniques

In the next subsection, we discuss human perception properties which are used in the compression schemes described in the later subsections. In the later subsections, we will discuss two standards for still image and moving image compression. They serve to illustrate the successive steps and resulting complexity of the image encoding and compression process.

2.3.1 Human Perception Properties

To increase compression ratios, human perception properties are often used to send only the information most visible to a human observer. There are a number of phenomena that affect our assessment of the quality of received images. Three such vision properties which are incorporated in many source-based compression schemes are the following. (1) The human eye is less sensitive to chrominance components. Thus less data needs to be used to represent the chrominance components. (2) The human visual system is more sensitive to middle spatial frequencies and less sensitive to low and high spatial frequencies; the fidelity of the edge contours is not as important and other aspects of the image. Thus we can segment images into different blocks according to spatial frequencies and encode them accordingly to achieve higher compression performance. (3) The human eye's sensitivity to quantizing distortion decreases as the luminance level increases, i.e., noise masking property. Thus high intensity samples can be quantized more coarsely.

2.3.2 JPEG

The Joint Photographic Experts Group (JPEG) standard [Wallace] was developed jointly by the ISO and International Telecommunication Union (ITU-T) for the compression of still images. JPEG is the first international digital image compression standard for multilevel continuous-tone gray-scale or color images. Depending upon the image and JPEG parameters, compression ratios may reach 25:1 without noticeable degradation.

The JPEG standard defines the following four modes of operation . *Sequential encoding* - single scan from left to right and from top to bottom - results in lossy images. *Progressive encoding* - multiple scan for applications in which the transmission bandwidth is low - the viewer can watch the image resolution increase at every pass from course grain to finer grain resolution. This technique results in lossy images. *Lossless encoding* - this technique results in lower compression ratios but is completely reversible at the decoder. *Hierarchical encoding* - the image is encoded at multiple resolutions which can be decompressed separately. Since sequential encoding is sufficient for many applications and platforms, most current implementations on the market implement it. In the following discussion on JPEG, we concentrate on the sequential encoding processes.

The JPEG compression process consists of the following.

1. *Preprocessing.* The preprocessing stage consists of the color space transformation and the block preparation. As mentioned before, the color space transformation maps the original representation, RGB color space, onto another in which the color information, or chrominance components, are separated from the intensity information, the luminance component. Each component is individually sampled according to its own relevance to the final visual display. In the block preparation phase, the image is divided into individual 8 pixel by 8 pixel blocks. The number of blocks depends on the subsampling format. Assuming an image contains P pixels per line and L lines: if a 4:1:1 chrominance format is specified, the luminance component consists of a matrix of P x L luminance values and each of the two chrominance components consists of a matrix of P/2 by L/2 values. Thus the input to the DCT phase will be P*L/16 blocks of luminance and P*L/(4*16) blocks for each of the two chrominance components.

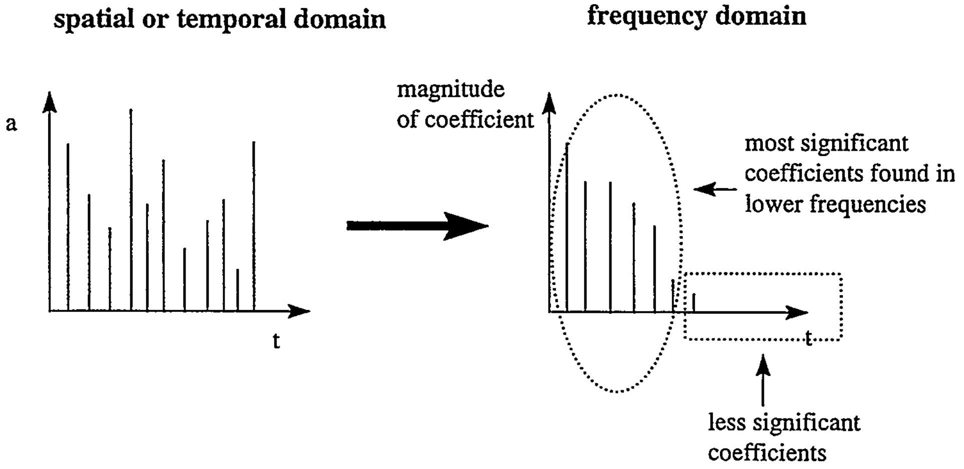
spatial or temporal domain
frequency domain
a
t
magnitude of coefficient
most significant coefficients found in lower frequencies
t
less significant coefficients

Figure 2.2 Principle of transform coding

2. *Discrete Cosine Transform* [Rao]. Each 8 x 8 block of source image samples is actually a 64 point discrete signal that is a function of the x and y spatial dimensions, i.e., f(x,y),where 0<=x<8, 0<=y<8.

The Forward DCT (FDCT) takes the input signal and decomposes it into 64 orthogonal basis vector signals. The output of the FDCT is a set of 64 basis signal amplitudes known as the *DCT coefficients*. The coefficient for the (0,0) vector is called the *DC component*. All other coefficients are called *AC components*. The DC component usually contains the most significant fraction of the total image block energy. Because sample values typically vary slowly point to point across an image, many of the coefficients will have values near zero or equal to zero.

Figure 2.2 depicts the basic principle of transform encoding. After the transformation into the frequency domain, the most significant coefficients may be coded with better accuracy than the less significant ones. Certain less significant coefficients may also be discarded.

3. *Quantization and post-processing of DCT coefficients.* Each of the 64 DCT coefficients obtained at the output of the FDCT is then uniformly quantized by using a 64-element quantization table that is specified by the application. Each element of the table specifies the step size of the quantizer for its corresponding DCT coefficient. The purpose of quantization is to achieve further compression by discarding information that is not visually significant. Quantization is a lossy process and is the principal source of lossiness in DCT-based encoders. After the quantization process the DC components are encoded separately. Since there is usually a high correlation between the DC coefficients of adjacent 8 x 8 blocks, the quantized DC

coefficients are encoded using DPCM. To facilitate entropy encoding, the quantized AC components are ordered into a "zigzag" sequence. This ordering helps the entropy coding process be more efficient; the low coordinate coefficients, which are more likely to be non-zero, are ordered before the high coordinate coefficients.

4. *Entropy Coding.* The DCT coefficients are then entropy-coded. Since the pixel information of each block was scanned in zigzag order, the runs of zero coefficients are increased allowing the entropy encoding process to more efficiently compress the data. JPEG uses either Huffman coding or arithmetic coding as the entropy encoding scheme in this final stage.

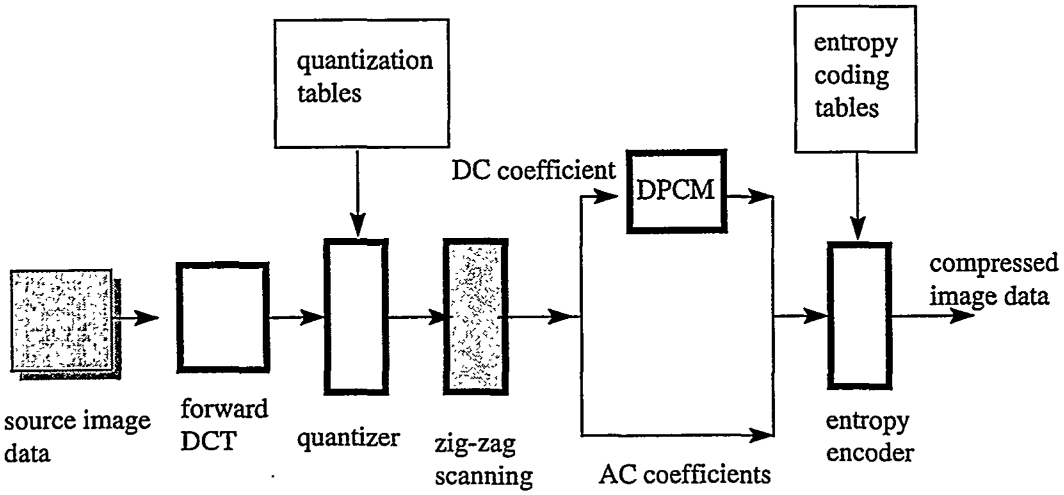

Figure 2.3 JPEG image compression process

2.3.3 H.261 and H.263

The first completed standardization effort with respect to motion compression was the ITU-T recommendation H.261 standard also known as the px64 standard [Liou, Turletti]. The H.261 standard can work at 64 kbps, or at multiples of 64 kbps when several ISDN connections are aggregated. Its primary use is in real-time video conferencing applications over low speed links.

H.263 is an ITU-T standard for low bit rate video and audio teleconferencing. Since H.263 was defined after H.261 and after the formative stage of MPEG-1, it contains many of the advanced methods discovered during the formation of those prior standards. Actually, the syntax of H.263 is much more similar to that of MPEG-1 than that of H.261. H.263 is designed to be more efficient than H.261 for bit rates below 64 kbps (ISDN B channel). The primary target bit rate, approximately 27 kbps, is the payload rate of the V.34 modem standard. For this bit rate, typically, 20 kbps would be allocated for the video portion and 6.5 kbps for the speech section. The H.261 and H.263 video compression schemes are based on transform coding, via the DCT, quantization, and entropy coding of the remaining coefficients using Huffman coding. This scheme is so similar to the ones used in JPEG and MPEG that the reader should go to those sections for details of the specific workings of the scheme.

H.320 is the ITU standard for visual telephony coding, compression and framing. More specific standards include H.323, the international audio, video and data conferencing standard for IP-based networks, and H.324, the standard for video conferencing over ordinary POTS (plain-old telephone system) lines. H.261 is the video coding component of H.323. H.263 is the video coding component of H.324.

2.3.4 MPEG-1 and MPEG-2

The Moving Pictures Expert Group (MPEG) [LeGall] was formed in 1988 to establish an international standard for the coded representation of moving pictures and associated audio stored on digital storage media. The fundamental compression technique, DCT-based transformation and entropy coding, is also incorporated in MPEG. Thus we will discuss only the major departures of MPEG from the JPEG standard.

As mentioned before, compression consists of removing redundant information. In video, there are two types of redundancies which may be exploited: *spatial redundancies* and *temporal redundancies.* Compression techniques which use spatial correlations eliminate or reduce redundancies within each frame. The JPEG standard for still images is an example of a compression scheme which employs spatial correlation techniques. Temporal correlation techniques identify redundancies between successive frames. Reduction by way of temporal correlation usually employs DPCM techniques. This process is also called motion compensation. It is based on the observation that over certain time intervals, frames contain redundancies and hence may be partially constructed from other frames. MPEG-1 and MPEG-2 video streams may employ both spatial correlation and temporal correlation techniques.

Motion JPEG is an example of a coding scheme which uses only spatial correlation techniques (also called intraframe compression). Since Motion JPEG only performs intraframe compression, redundancies between adjacent frames may not be eliminated in order to increase the compression ratio. Thus a motion JPEG encoding of a stream will be of higher bandwidth than an identical stream encoded using MPEG's intraframe as well as temporal correlation (interframe) techniques. The advantage of pure intraframe encoding is that refresh or reference frames are accessable at any frame time, i.e., "fast" random access to any frame. If interframe encoding is used, refresh access may be significantly greater than one frame time.

MPEG-1 is the first version of the standard. The part of the standard which addresses the coding and compression of video has defined several frame and sampling formats. The most common format, which nearly corresponds to VCR-quality, is called the Standard Interchange Format (SIF). SIF is derived from the ITU-R 601 format. The luminance of SIF has 352 samples per line, nearly half that of ITU-R 601, and either 240 lines, for NTSC, or 288 lines, for PAL/SECAM. The two color difference signals are sampled at one half the frequency of the luminance; 176 samples per line and either 120 or 144 lines per frame. Thus the MPEG-1 subsampling ratio is 4:1:1.

Before delving further into the MPEG standard, we will briefly discuss a technique for reducing the temporal redundancies called *motion compensation*. Consider three successive frames, *frame1*, *frame2* and *frame3*, taken several milliseconds apart from a real world scene. If only spatial redundancy techniques are used, a block of data which was already transferred in *frame1* will be transferred again in *frame2*. In such cases it is more efficient to transfer only the vector which indicates the spatial translation of the block from *frame1* to *frame2*. This vector is called the motion vector (see Figure 2.4).

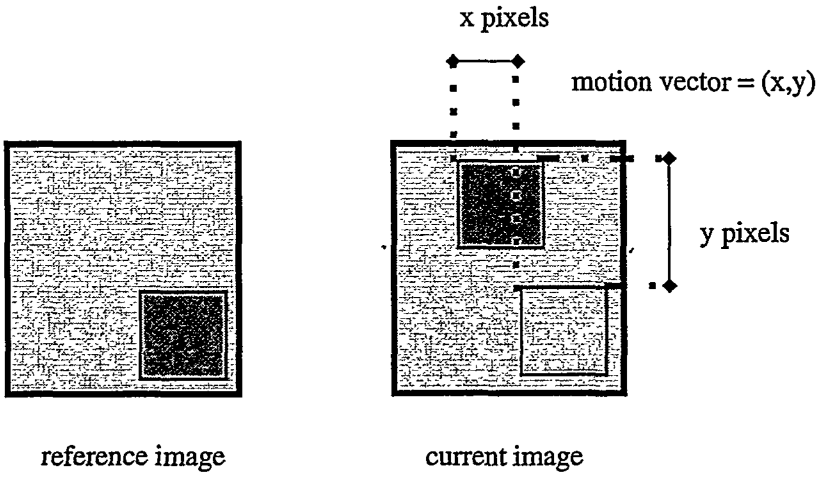

Figure 2.4 Motion compensation: using part of a reference image to construct the current image.

The blocks on which the motion vector are applied are called matching blocks. In MPEG-1, the matching blocks are squares of 16 x 16 pixles on the luminance plane and squares of 8 x 8 pixels on each of the color difference planes. The combination of these squares is called a macroblock. It cannot be assumed that for each macroblock in a predicted or interpolated frame, we can find a macroblock that matches it in the reference frame. Thus the algorithm searches for the best matching block within a certain area. If the match is not exact, the arithmetic difference between the actual block and the best matching macroblock is calculated. This difference, which is also a macroblock, is called the error term. If certain macroblocks do not find a satisfactory matching block, they will be coded with no dependencies to prior frames, i.e., in the same way as macroblocks in I-frames.

MPEG distinguishes frames into the following three categories.

1. *Intra-coded frames (I frames).* An I frame is a still image frame. It is independent of all other frames in the stream, so it can serve as a reference frame for the remaining two types of frames. An I frame is coded using very similar techniques used to code JPEG still images.
2. *Predictive-coded frames (P frames).* P frames exploit the notion of temporal redundancy. They require information from the previous I frame and/or P frame for encoding and decoding. A P frame can contain macro blocks which are intraframe encoded or predictively encoded. Predictively encoded macro blocks contain vector and difference DCT coefficients from the best matching macro block from the previous I and/or P frame. Since the motion vectors of adjacent macro blocks often differ only slightly, DPCM coding is used. The result is then transformed to a variable-length encoded block
3. *Bidirectional-coded frames (B frames).* A B frame is defined as the difference of a prediction of the past frame and the following P or I frame. So B frames are dependent upon the previous and following I and/or P frame for encoding and decoding. The B frame is created by a very similar process, i.e., DCT, quantization and entropy encoding, as P frames. The major difference is that both the forward vector and the backward vector must be "averaged" , and then it must be subtracted from the current macro block.

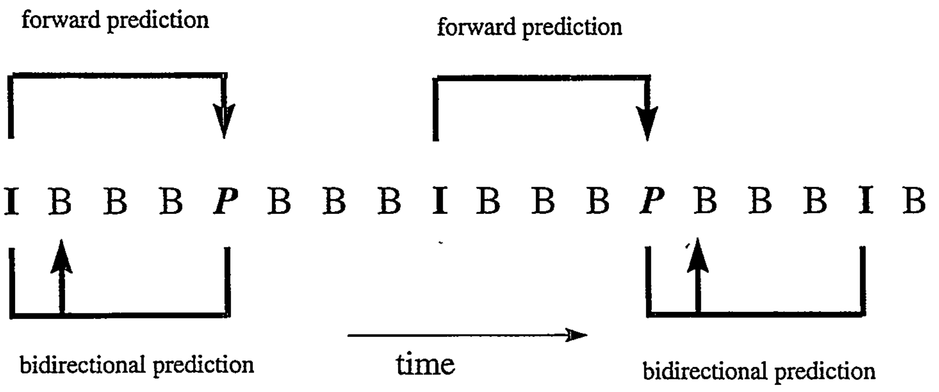

Figure 2.5 A sequence of I, B and P frames

The usage of B frames has been the subject of much controversy. There are two major advantages of using B frames. The first is that at low bit rates, under 1.5 Mbps, the noise level measured by the signal to noise ratio (SNR) is reduced. B frames also result in higher compression ratios which lead to lower average bit rates for equivalent quality video streams. The disadvantages of using B frames are the additional computational complexity, delay and image buffer size requirements. B frames require the averaging process of macro blocks from two other frames, thus resulting in additional processing at the encoder and decoder. They also require an extra frame buffer to store the future prediction reference. Finally, an extra delay is introduced while encoding since the frame used for the backwards prediction needs to be transmitted to the decoder before the intermediate B frame can be decoded and displayed.

MPEG-2 [LeGall, Steinmetz2] was initially targeted for the digital transmission of broadcast-quality TV at bit rates between 4 to 9 Mbps. Later, during its development it absorbed the compression format for HDTV (initially covered by the failed MPEG-3 standard). The standard allows compression of both progressive and interlaced video at various bit rates ranging from about 1.5 Mbps to more than 60 Mbps, enabling applications ranging from home entertainment quality video up to HDTV.

MPEG-2 video compression uses the same principles as MPEG-1 video compression with some noticeable extensions and improvements to support high quality video. The major extensions are the following.

1. MPEG-2 supports both interlaced and non-interlaced video.
2. The video stream syntax allows picture sizes as large as 16,383 x 16,383 pixels.
3. To support a possible heterogenous environment, MPEG-2 provides scaleable video. Four scalable modes encode a MPEG-2 video stream into different layers (base, middle and high layers) mostly for prioritizing data. This has two main purposes: (1) important parts of the video data are specified as high priority to protect them from transmission errors or overflow conditions, (2) scaleable video allows the decoder to selectively decode part of a video stream. For example, if an HDTV video stream is coded in multiple layers, and one layer corresponds to conventional TV resolution, then an end system capable of only processing conventional TV quality signals may decode only the part of the stream it requires. The four scaleable modes are the following.

- *Spatial scalability* allows a video signal to be carried in a two part format which allows inexpensive decoders to extract a low-resolution signal. More powerful decoders use the entire signal.
- Data partitioning. This is a frequency domain method that breaks the block of 64 quantized transform coefficients into two bit streams. The first higher priority channel contains the more critical lower frequency coefficients. The second lower priority channel carries the higher frequency AC data. This mode is similar to JPEG's frequency progressive mode.
- *Temporal scalability* allows one signal to be transmitted and displayed at different frame rates.

- *Signal-to-noise ratio (SNR) scalability.* This is a spatial domain method where layers are encoded at identical sample rates but at different picture qualities. The higher priority layer contains base layer data to which a lower priority refinement layer can be added to construct a higher quality image.
- Hybrid scalability. This is the combination of two different types of scalability. The types of scalability that can be combined are SNR scalability, spatial scalability and temporal scalability.

Subsets of the standards are known as profiles (which delimit syntax, i.e., algorithms) and levels (which delimit signal parameters). Table 2.1 depicts the MPEG-2 defined levels,,and Table 2.2 depicts the MPEG-2 defined profiles.

Level	Description
Simple	same as Main Level, but with no B frames; intended for software decoders; 4:2:0; not scaleable
Main	low-cost single chip implementation for cable TV and satellite uplink compression; 4:2:0; not scaleable
Spatial	same as Main Level but with spatial scalability; e.g., HDTV; 4:2:0
SNR	same as Spatial Level but with SNR scalability; 4:2:0
High	same as SNR Level but with 4:4:4, 4:2:0 and 4:2:2 formats

Table 2.1 Main features of MPEG-2 levels

Profile	Target Bit Rate	Resolution	Display Quality
Low	4 Mbps	352 x 240 x 30	VHS equivalent
Main	15 Mbps	720x480x30	Broadcast quality
High-1440	60 Mbps	1440x1152x30	HDTV quality
High	80 Mbps	1920x1080x30	Film production quality

Table 2.2 MPEG-2 Profiles

2.3.5 Wavelets

Wavelet compression [Rioul] is a type of transformation encoding method. It works similarly to the DCT. As described in section 2.3.2, a DCT represents an input signal using a series of cosine functions as basis functions. The output of a DCT is a series of coefficients which represent the amplitudes of the frequencies. The DCT and Fourier transform work best for continuous and repetitive inputs. For applications involving graphics, or bitonal images, e.g., any image with only a few distinct lines (e.g., line drawings, cartoons), a DCT or Fourier transform would produce too many large coefficients for the high frequency components.

Wavelets also represent input images with coefficients of basis functions. However, the basis functions for wavelets are much more complicated than sines or cosines. These wavelet basis functions can more effectively represent non-continuous and non-repetitive images. Thus wavelets can represent these images using only a small set of coefficients resulting in higher compression ratios.

2.3.6 Fractals

Fractal encoding [Bogdan] is based upon *vector quantization.* In vector quantization, the data stream is divided into blocks of data bits or vectors. There is a previously constructed table which contains sets of patterns of bits. This table, called a code-book, may also be dynamically constructed or/and modified. For each vector, the code-book is used to find the entry which most closely matches the current vector. The index of the table with the best match is transmitted. Both the source and destination must have a copy of the code-book. Differences between the code-book entries and the vector may occur. In that case, the difference between the actual vector and the pattern may be transmitted. This difference may be exact, for lossless coding, or approximate, for lossy coding. Vector quantization encoding is particularly suitable for data which contains well known patterns, such as speech.

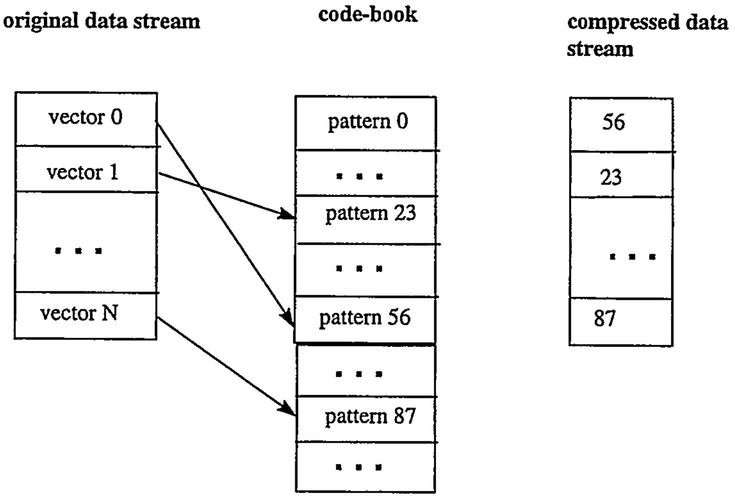

Figure 2.3 Principle of vector quantization encoding

Fractal encoding consists of searching for fractals in an existing digitized image. A fractal is an irregular geometric form which consists of parts which can be repeated at different scales and angles. Initially images are partitioned into small square areas. Each individual square area is "compared" to other parts of the same image. Since in most cases there is little chance that the square exactly matches another area, it is actually "compared" to modified versions of other parts - that is parts which have been shrunk, slanted, rotated or mirrored.

The transformations that the found fractal had to undergo to match the initial square may be expressed with coded formulas, i.e., fractal transform codes. These formulas are used to reconstruct the image at the decoding stage. The set of coded formulas is much more compact than the initial image. It is obvious from the description of this encoding/decoding process that fractal transform requires significantly more processing power for encoding than for decoding.

3. Relationship between applications, coding techniques and network requirements

As seen on the prior section, there are many existing coding techniques for performing video compression. In this section, we present the various compression techniques in relationship to the applications which will use them as well as in relationship to the underlying network transmission media.

The selection of an appropriate coding technique for a particular application is dependent upon that application's requirements and network environment. The following factors must be considered.

1. *Degree of user interactivness.* Compression speed is particularly important for real-time applications, such as video conferencing, where the degree of user interactiveness depends upon the speed of the coding and decoding processes. For such real-time applications, either hardware-implemented coding techniques, or fast algorithms based upon predictive coding or transform coding, can be used. For information retrieval and presentation type applications, where the images may be coded entirely off-line, the compression and decompression speeds must be considered separately. Since the compression process is not subject to real-time bounds, its speed is not very important. However, the decompression speed is very important in order to provide the user with an adequate response. For non-real-time applications, coding schemes which support more desirable compression ratios and higher image quality may be chosen.

   With most compression schemes, the compression process is more complex than the decompression process. Algorithms where the compression process requires significantly more processing power than the decompression phase are said to be *asymmetric.* Vector quantization is a typical asymmetric method; the decompression is extremely fast because it consists almost entirely of direct access to a table. Fractal transformation coding is one of the most highly asymmetric compression methods.

2. *Reliability of transmission channels.* If the channel is not very reliable, non-predictive coding techniques such as transform coding are preferable. Predictive coding techniques are very sensitive to transmission errors. This is because prediction is based upon image statistics and past history of image data. The variation of image statistics may cause variation in the quality of the reconstructed image. A single bit error may cause distortion of a noticeable area of an image and may affect several successive image frames.

3. *Random access speed.* Some applications, such as interactive VOD, allow a user to interactively randomly select a particular frame in a video sequence. Intraframe coding techniques are most appropriate for these types of applications.

4. *Available channel capacity.* In the conversion from analog video to digital video, samples are taken at regular intervals and each sample is represented using the same number of bits. Thus an uncompressed video stream is a constant bit rate (CBR) stream. Given unlimited channel capacity, this uncompressed CBR stream may be transmitted. Compression schemes use either CBR coding techniques or variable bit rate (VBR) coding techniques. In most cases, VBR techniques are more efficient and produce higher quality as well as constant image quality video than CBR techniques at the same compression ratio. All of the compression techniques we have discussed produce VBR traffic.

   In the previous section, we discussed several widely used standards - JPEG and MPEG. Even though the compression ratio provided by these standards is a function of various parameters, e.g., quantization factors, GOP pattern, video image type, etc., the genralization that MPEG provides higher compression ratios than JPEG is reasonable. Basically, MPEG provides inter-frame compression via motion compensation as well as the intra-frame compression techniques found in JPEG. It is also important to remember that high compression ratios may be sought to reduce network resource usage, however, in general, higher compression ratios result in lower re-constructed media quality.

5. *Image type.* Images may be categorized as either bitonal images or continuous-tone images. Bitonal images, or bi-level images, do not use grey scale characters. They do not contain graduations in shading values. Examples of bitonal images are a page of text and cartoons. Continuous-tone images contain gradations in shading values. Images from real-life, say captured in a photograph or moving image camera, are continuous-tone images. Compression schemes based upon the DCT such as JPEG and MPEG are not efficient for compressing bitonal images. This is because the rapid shift between tones in adjacent blocks of a bitonal image entails a high coefficient for one of the highest frequencies. Usually continuous-tone images from real-life do not contain many sharp lines or zones. Thus their information is mainly contained in the low frequencies. Wavelet and various vector quantization approaches such as fractal encoding are much more suited for the compression of bitonal images.

Table 3.1 Types of compression techniques

compression technique/standard	**suitability**	**unsuitability**	**applications**
entropy encoding	loss-intolerant apps	data with content-based redundancy	computer data
JPEG	continuous-tone still and moving images	1) non-realistic images: text, line drawings, cartoons and 2) high quality video	video conferencing, VCR-quality applications
MPEG-1	continuous-tone still and moving images	same as above	VCR-quality applications
MPEG-2	continuous-tone still and moving images	same as above	broadcast-quality TV applications to HDTV applications
wavelets	images which contain sparse distinct lines, e.g., cartoons, line drawings		graphical imaging, WWW, very low bit rate applications
fractals	1) images which contain irregular geometric forms and 2) applications which require very high compression ratios	interactive applications which require fast encoding	WWW, very low bit rate applications

3.1 Applications

3.1.1 Television

Television is undeniably the most important application that has driven the development of motion video. In section 2.2, we discussed the various grades of television quality. In this section, we will discuss the network requirements of transmitting these various grades of television as well as their relationship to various compression technologies.

In the following, we compute the bandwidth requirements for various types of uncompressed video:

1. *Analog Video.* The bandwidth requirements of analog video can be derived from the scanning parameters. Assuming that the frame rate is F, the number of scanning lines per frame is N, the horizontal resolution is H, the fraction of he horizontal scanning interval devoted to signal transmission is C (the remainder is used for horizontal blanking), and the aspect ratio is A, then the system bandwidth B can be derived as follows:

$$B = F * N * \text{(cycles per line)}$$
$$\text{cycles per line} = 0.5 * A * H/C,$$

where 0.5 is the ratio of the number of cycles to the number of lines distinguishable. Thus we obtain: B = 0.5AFHN/C. For example, for the PAL system, A = 4/3, F = 25, H = 409, N = 625, and C = 0.80, thus B = 5.3 MHz.

2. *Studio-Quality Digital Video.* There are two components necessary to compute the overall bit rate for a digital video (CCIR-601) signal: the luminance component and the two chrominance components. The sampling frequency adopted is 13.5 MHz for the luminance component and 6.75 MHz for the color component; the subsampling ratio is 4:2:2. In other words, the number of samples per line for each chrominance component is half that of the luminance component, but the number of lines per frame is the same. As a result, there is a 50% reduction of the bit rate of each chrominance component. The overall reduction of the bit stream is 33%.

   Assuming the vertical resolution is N lines per frame and the frame rate is F frames per second, the number of luminance samples per line is equivalent to 13,500,000/(N*F). Thus, for a 525/30 NTSC signal or a 625/25 PAL/SECAM signal, the resulting bit rate of the luminance component would be 13,500,000 Hz * 8 bits = 108 Mbps, where every luminance sample is coded in 8 bits. The resulting bit rate of the chrominance signal would be 6,750,000 Hz * 8 bits = 54 Mbps. Thus the total bit rate would be 108 Mbps + 2 * 54 Mbps = 216 Mbps.

3. *HDTV.* The total bit rate for a digital HDTV signal is computed similarly to the digital studio-quality signal in the preceding item. Assuming the *high resolution/high frame rate* HDTV quality grade, the resulting bit rate for the luminance component would be 1920 samples per line * 1080 lines * 8 bits * 60 frames/sec = 995 Mbps. A subsampling ratio of 4:2:2 would result in a total bit rate of 1.99 Gbps, a 4:4:4 subsampling ratio would result in a total bit rate of nearly 3 Gbps.

The bandwidth requirements for a generic video stream using various compression schemes are given below.

1. *Broadcast-quality Video.* Existing implementations of the MPEG-2 compression standard operate at about 6 Mbps. It is expected to achieve in the order of 2 to 3 Mbps for a quality equivalent to that of NTSC broadcast and 4 Mbps for a quality equivalent to that of PAL/SECAM broadcast.
2. *VCR quality Video.* Compression schemes such as MPEG-1 and DVI provide off-line compression at 1.2 Mbps
3. *HDTV.* The achieved bit rate is highly dependent on whether the compression takes place off line to produce a stored version of the compressed digital stream, or whether it is performed in real-time. It is expected that a compression ratio in the order of 100 to 1 may be achieved. The high resolution/high frame rate HDTV should be compressed down to a bit rate ranging from 20 to 34 Mbps. The high resolution/conventional frame rate HDTV should require between 15 and 25 Mbps.]

Table 3.2 provides a summary of the compressed and uncompressed bandwidth requirements for the various grades of video quality.

Table 3.2 Bit rates for motion video streams

Quality	Compression Standard	Uncompressed Mbps	Compressed Mbps
HDTV (1920x1080)/60 fps	MPEG-2	1990	25 to 34
Studio-quality digital TV	ITU-R 601 MPEG-2	216	3 to 6
Broadcast-quality TV	MPEG-2	NA	2 to 4
VCR-quality	MPEG-1	4 to 5	1.2

Delay jitter requirements. In general, a real-time motion video stream is transmitted simultaneoulsy with an audio stream for synchronous presentation. In such cases, the requirements on the transit delay and the delay jitter will usually be dictated by the audio stream. Course synchronization, known as "lip-synchronization", will be required for all grades of quality. [Fluckiger,1995] sets the delay jitter bound to not exceed 50 ms for HDTV quality, 100 ms for broadcast quality, and 400 ms for videoconference quality.

3.1.2 Video Conferencing and Computer-Supported Cooperative Work (CSCW)

Videoconferencing is the real-time two-way transmission of digitized video images and audio, between two or more locations. Computer-based video teleconferencing usually implies the low speed transmission of digital images. Low speed implies 64 kbps over a basic rate ISDN circuit (128 kbps for bi-directional video) to a few hundred of kilobits per second, e.g., 384 kbps. There are many different resolution levels for video conferencing. The most well known is the ITU-TS standard called the H.261 for video encoding and compression, which was discussed in section 3.3.3. In general, the resolution of current video teleconferencing images is about one quarter that of broadcast quality television. The low speed also limits the motion frequency to between 5Hz to 10 Hz.

Desktop applications emerged naturally from videoconferencing participants' desire to collaborate in other mediums. One resulting application, document conferencing, or more commonly called "white boarding", involves participants sharing ideas through an electronic blackboard. There exist other similar applications where participants work jointly on word processing, spreadsheet manipulation, and other traditional office tasks. Since these applications don't involve complex images, such as human images, the network requirements tend to be very low. However to maintain the interactive nature of these collaborative applications, low latency must be guaranteed.

An important attribute of videoconferencing/CSCW applications is synchronization [Prab, Steinmetz1]. In these applications, there are inherent dependencies between the various multimedia streams. For instance in a slide show, there are two streams - the moving images which represent the slides, and the audio commentary stream. The presentation of the images must be synchronized with the appropriate audio units; the synchronization skew between the two streams must be bounded in time.

Different media combinations have different skew requirements. The synchronization of two channels of stereo audio have the most stringent requirements. Table 3.3 shows the results of a comprehensive study on intermedia skew tolerances [Steinmetz1]. As the tolerable intermedia skew is subjective and application dependent, different experiments may find different values. The values in Table 3.3 just serve as a rough indication of intermedia synchronization requirements.

Table 3.3 **Intermedia Skew Tolerances [Steinmetz1]**

Media Type		Mode or Application	allowable intermedia skew ranges
Video	Animation	correlated	+/- 120 ms
	Audio	lip sync	+/- 80 ms
	Image	overlay	+/- 240 ms
	Text	overlay	+/- 240 ms
Audio	Animation	correlated	+/- 80 ms
	Audio	stereo	+/- 11ms
		dialogue between multiple people	+/- 120 ms
		background music	+/- 500 ms
	Image	slide show	+/- 500 ms
	Text	text annotation	+/- 240 ms
	Pointer	audio pointer	+/- 500 ms

There are many possible causes for losing synchronization between multiple media types. Some include: different media may require different amounts of coding, decoding and processing times, network transmission times may vary for different packets, depacketization and protocol processing times may vary, user interactions may create anomalies, etc. Preventive and corrective measures to some of these problems have been proposed in [Prab]. The discussion of these measures is beyond the scope of this study.

3.1.3 Medical Imaging

Recent years have seen an increasing use of imaging technologies such as magnetic resonance imaging (MRI) and computerized tomography (CT). One fundamental difficulty in working with digital medical images is the large size of the images. For instance, typical image sizes are 0.5 Mb for a CT image, 0.13 Mb for a MR image of blood vessels in the chest, 8 Mb form a digitized x-ray, 24 Mb for a 2-dimensional brain scan image and 384 Mb for a 3-dimensional brain scan image.

Due to the nature of the data, image quality has always been of the highest importance in medical imaging applications; compression schemes for encoding medical images have traditionally used lossless techniques. Using the standard Lempel-Ziv algorithm applied to MR and CT scans, typical compression ratios of approximately 2:1 are achieved. Recent studies have shown that with more complex lossless compression techniques, compression ratios of 3:1 or 4:1 are possible. In general, lossless compresssion techniques applied to medical images are incapable of providing compression ratios higher than 4:1.

In medical applications which rely on lossy compression techniques, the primary concern is that the diagnostic accuracy of the lossy compressed images remain not less than that of the original images. Signal-to-noise ratios (SNR) or mean squared errors (MSE) may or may not be a good indication of diagnostic accuracy, but the actual accuracy must be demonstrated. In the next paragraph, we discuss a lossy compression scheme where the effects of the lossy compression process on the accuracy of MR images of blood vessels in the chest were actually quantified.

As seen in the previous section, many approaches to digital compression systems have been proposed in the literature and incorporated into standards. These differ primarily by the different choices made for the three basic components: signal encoding or decomposition, quantization and lossless coding. Many algorithms and systems have been proposed for the lossy encoding of various types of medical images. We choose to discuss one developed at Stanford University [Adams] beacause they have also developed methods for quantifying the effects of compression on their target set of images. The algorithm used was predictive pruned tree-structured vector quantization. This algorithm does not perform a transform encoding such as a DCT or wavelet. There are several advantages to this approach. Firstly, there is no need to compute transforms and inverse transforms or to do separate entropy coding. This leads to a simple decompression algorithm, which depends mostly on table lookups for resolving code words. Another advantage is that the tree-structured algorithms inherently provide a natural progressive structure to the code, which allows for the ability of progressive reconstruction of an improved image as bits arrive. Using this algorithm, this study showed that for the purpose of measuring blood vessels in the chest, there is no significant difference in measurement accuracy when images are compressed up to 16:1; real-time MR imaging results in uncompressed streams of nearly 4 Mbps - compressed streams use nearly 250 kbps.

Some types of medical images may be too large to compress in order to support real-time transmission frame rates of 30 frames per second. This was shown in a study undertaken at the University of Minnesota [Claypool], which sought to determine the networking requirements for three-dimensional high-field MR images of the human brain. This application provided the ability to visualize high quality, high-resolution micro-graphs montaged together into three-dimensional structures as they were in the living brain. Estimated network requirements are 720 Mbps for 2-dimensional navigation through the brain and 11.5 Gbps for 3-dimensional navigation through the brain. If there are simultaneous users, terabit per second network throughputs would be required.

4. Bibliography

[Adams] Adams, C.N., Aiyer, A., et.al., "Evaluating Quality and Utility of Digital Mammograms and Lossy Compressed Digital Mammograms", Technical Report, Dept of Electrical Engineering, Stanford University, CA.

[Bogdan] Bogdan, A., "Multiscale Fractal Video Coding", *Proceedings of IEEE International Conference on Image Processing*, vol 1, pp 760-764, Austin, TX, 1994.

[Claypool] Claypool, M., Riedl, J., "Network Zooming Requirements for the Visualization of 3-Dimensional Brain Images", *IEEE Journal on Selected Areas of Communications.*

[Colaitis] Colaitis, F., Bertrand, F., (1994) "The MHEG Standard, Principles, and Examples of Applications", *Multimedia/Hypermedia in Open Distributed Environments*, Proceedings of the Eurographics Symposium, Graz 1994, Springer-Verlag, Vienna,

[Fluckiger] Fluckiger, F., (1995) "Understanding Networked Multimedia", Prentice Hall, New Jersey, USA.

[LeGall] Le Gall, D., (1991) "MPEG: A Video Compression Standard for Multimedia Applications", *Communications of the ACM*, vol. 34, no. 4.

[Liou] Liou, M., (1991) "Overview of the px64 kbps Video Coding Standard", *Communications of the ACM*, vol. 34, no. 4.

[Prab] Prabhakaran, B., Raghavan, S.V., (1993) "Synchronization Models for Multimedia Presentation with User Participation", *ACM Multimedia 1993*, Proceedings, Anaheim, California.

[Rao] Rao. K.R., "Discrete Cosine Transform, Algorithms, Advantages, Applications", Academic Press, San Diego, CA, 1990.

[Rioul] Rioul, O, Vetterli, M., "Wavelets and Signal Processing", *IEEE Signal Processing Magazine,* vol. 8, no. 4, pp 14-38, Oct 1991.

[Steinmetz1] Steinmetz, R., Meyer, T., (1992) "Multimedia Synchronization Techniques: Experiences Based on Different System Structures", *Proceedings of IEEE Multimedia 1992,* Monterey, CA.

[Steinmetz2] Steinmetz, R., Nahrstedt, K., (1995) "Multimedia: Computing Communications and Applications", Prentice Hall, New Jersey, USA.

[Turletti] Turletti, T., (1993) "The INRIA Videoconferencing System (IVS)", *ConneXions*, InterOp Company, Foster City, California.

[Wallace] Wallace, G., (1991) "The JPEG Still Image Compression Standard", *Communications of the ACM*, vol. 34, no. 4.

UNLIMITED RELEASE

INITIAL DISTRIBUTION:

1	MS 9003	K. E. Washington, 8900	
1	MS 9011	P. W. Dean, 8903	
1	MS 9011	B. V. Hess, 8910	
2	MS 9011	J. M. Brandt, 8910	
2	MS 9011	H. Y. Chen, 8910	
2	MS 9011	J. A. Hutchins, 8910	
5	MS 9011	R. Tsang, 8910	
1	MS 9011	P. E. Nielan, 8920	
1	MS 9012	R. Trechter (actg), 8930	
1	MS 9037	J. C. Berry, 8930-1	
1	MS 9019	B. A. Maxwell, 8940	
1	MS 9217	J. C. Meza, 8950	
1	MS 9019	J. A. Larson, 8970	
1	MS 9012	K. R. Hughes, 8990	
1	MS 9001	M. E. John, 8000	
		Attn:	R. C. Wayne, 2200
			J. Vitko, 8100
			W. J. McLean, 8300
			D. Henson, 8400
			P. N. Smith, 8500
			T. M. Dyer, 8700
3	MS 9018	Central Technical Files, 8940-2	
1	MS 0899	Technical Library, 4916	
1	MS 9021	Technical Communications Dept. 8815/Technical Library, 4916	
1	MS 9021	Technical Communications Dept. 8815 for DOE/OSTI	