DIGITAL COMPRESSION *for* MULTIMEDIA

Principles and Standards

Jerry D. Gibson — *Southern Methodist University*
Toby Berger — *Cornell University*
Tom Lookabaugh — *DiviCom*
Dave Lindbergh — *PictureTel Corporation*
Richard L. Baker — *PictureTel Corporation*

Morgan Kaufmann Publishers, Inc.
San Francisco, California

1 Introduction to Data Compression

CHAPTER

1.1 Why Compress?

Entertainment, telecommunications, the Internet—all are part of our daily lives. We enjoy them, use them in our businesses, surf them. We read about them in magazines and newspapers. We hear about them on television. We invest in them. But we've had radio, TV, music, and telephones for decades. What's new now? Is it all hype? Not really—at least, not all of it! The new word is "digital." Today we are talking about digital communications systems and networks and digital representations of movies, TV, music, images, and voice. Why digital? Digital signals are easy to store and easy to transmit over long distances without accumulating distortion, and stored digital representations (for example, of music) are highly resistant to minor degradations.

But there is a downside. Digital versions of important signals, like voice, music, TV, and movies, require more bits per second of signal to store or transmit, which translates into higher costs. For example, Table 1.1 presents the raw (uncompressed) data rates of several important source signals (Jayant, Johnston, and Safranek 1993). Certainly, many of the numbers seem large, but these numbers are only significant in comparison to the storage capacity available or the rate that can be sent over a chosen communications link. To get some idea of the significance of the rates in Table 1.1, note that the current common telephone modem rate is 28.8 kbps and the bit rate allocated to voice in North American digital cellular is 8 kbps, so the uncompressed 96-kbps requirement for telephone bandwidth voice is too high by about 12:1. Further, CD-ROM capacity is roughly 650 megabytes, and the capacity of one version of the evolving digital versatile (or video) disc (DVD-5) is roughly 40 gigabits.

TABLE 1.1 Approximate Bit Rates for Uncompressed Sources

Telephony (200–3400 Hz):	8000 samples/second × 12 bits/sample = 96 kbps
Wideband speech (50–7000 Hz):	16,000 samples/second × 14 bits/sample = 224 kbps
Wideband audio (20–20,000 Hz):	44,100 samples/second × 2 channels × 16 bits/sample = 1.412 Mbps
Images:	512 × 512 pixel color image × 24 bits/pixel = 6.3 Mbits/image
Video:	640 × 480 pixel color image × 24 bits/pixel × 30 images/second = 221 Mbps
HDTV:	1280 × 720 pixel color image × 60 images/second × 24 bits/pixel = 1.3 Gbps

Thus, for uncompressed video, the CD-ROM could store 23.5 seconds, and the DVD-5 could store about 3 minutes.

If the numbers in Table 1.1 turn out to be too large, and often they do, what can we do to improve the situation and still retain the advantages of digital transmission and storage? The answer is compression. Very generally, compression is the efficient digital representation of a source signal, such as speech, still images, music, or video; that is, we use as few bits as possible to represent the source signal while still having an adequate reproduction of the original (Berger 1971). Hence, the role of compression is to minimize the number of bits needed to retain an acceptable version of the original source signal, thus reducing storage and transmission costs (Sayood 1996).

Although you may now be convinced that compression is needed, it is natural to ask, What is enabling the current proliferation of compression applications? That is, computers, digital communications systems, and telecommunications networks have all been around for decades, so why are we able to implement these compression techniques now? Five events have made this possible. First, we are picking the fruits of over a quarter of a century of extraordinary research in compression methods. Since few people anticipated the amazing array of high-technology consumer products that we have today, much of this research was conducted without these products in mind. However, without this solid foundation of basic research, the rapid progress that we have seen would have been impossible. Second, signal processing capabilities are unparalleled, from VLSI implementations through today's powerful digital signal processors (DSPs) to, more recently, high-speed microprocessors in commonly available personal computers. With such computing power, seemingly complicated equations and algorithms can be implemented without excessive expense. Third, the introduction of perceptually based distortion measures has provided the final giant step toward producing the quality needed in applications—especially important since so many of the applications involve voice, music, images, and video, where the end user is the human eye or ear. Fourth, standards-setting activities have reduced the risk of offering products and services that incorporate compression by guaranteeing product and system interoperability. Finally, evolving technological advances in networks, computers, and telecommunications all continue to open new opportunities and to increase the pace of acceptance of compression methods.

Thus, this book is about compression—the compression of all sources, including data, voice, video, still images, audio, and movies. We discuss basic principles, common algorithms, and important standards. The goal is to help you understand current compression techniques and standards and why the various design choices and trade-offs were made. We also hope to prepare you to evaluate and design new compression algorithms and standards.

In the remainder of this chapter, we pose the data compression problem more carefully and describe the main issues involved in designing, implementing, selecting, and evaluating data compression methods, algorithms, standards, and services.

1.2 The Data Compression Problem

Data compression is simply the efficient digital representation of a source. However, this definition can be embellished slightly to make it more explicit: data compression is the representation of a source in digital form with as few bits as possible while maintaining an acceptable loss in fidelity. The source can be data, still images, speech, audio, video, or whatever signal needs to be stored or transmitted. In this book we use the term "data compression" to encompass both lossless and lossy compression of sources—*lossless* means perfect reconstruction of the source and *lossy* means that the source is not perfectly preserved in the representation. Over the years, there are a host of terms that have been used as synonyms for data compression—and still are—so it is appropriate to review them briefly.

1.2.1 Synonyms for Data Compression

The two terms most often used today as synonyms for data compression are *signal compression* and *signal coding* (Jayant, Johnston, and Safranek 1993).

These terms avoid the confusion that may arise from using the word *data,* since to some people, the word *data* can be limiting and exclude the possibility that we are considering speech, audio, or video compression, for example. The adjective *signal* seems a little bit too inclusionary, perhaps, but these terms are gaining some acceptance. In the information theory literature, the terms *source coding* and *source coding with a fidelity criterion* are common, although the latter is found much less often today than 15 to 20 years ago. *Source coding* can mean both lossless and lossy compression, but it is sometimes reserved by authors to indicate lossless coding only. *Source coding with a fidelity criterion* is long and unwieldy, but absolutely explicit in specifying that the compression being considered is lossy (Shannon 1959; Berger 1971; Gray 1990).

Some researchers in speech and audio compression use the term *source coding* to mean source model based coding, which is at some variance with the original information theoretic usage (Flanagan et al. 1979). However, once aware of this ambiguity, confusion can be avoided with careful reading. Other terms used are *noiseless* and *noisy (source) coding,* which refer, respectively, to lossless and lossy coding. The "noise" referred to is reconstruction error, or reconstruction noise; the use of "noise" in this context probably evolved since channel coding for noisy channels was developed and accepted earlier than source coding. A term coined to avoid the confusion that comes with "compression" is *data compaction,* which is used to indicate lossless encoding of a source. This term is a good one but has not received widespread acceptance (Blahut 1987). More dated synonyms for data compression are *bandwidth compression* and *redundancy removal*. Bandwidth compression is interchangeable with *signal compression* and *data compression,* in general; redundancy removal refers to one step in a lossy coding process, such as prediction in speech coding.

Throughout this book, we will employ "data compression" in the most inclusive sense.

1.2.2 Components of a Data Compression Problem

The major components of a data compression problem are the source, the rate, and the fidelity criterion or distortion measure. We are interested in compressing all sources, but different approaches to compression can be taken depending on whether a good source model is available or not. For example, telephone bandwidth speech is accurately represented for many applications by what is called the *linear prediction model;* hence this structure appears explicitly in many speech coders. On the other hand, wideband audio, still images, and video lack a very structured model; hence, compression of these sources tends not to rely on a particular source model.

TABLE 1.2 Audio Sampling Rates

Application	Bandwidth (kHz)	Sampling Rate (kHz)
Voice telephony	3.2	8
Teleconferencing (audio)	7.0	16
Compact disc (CD) audio	20.0	44.1
Digital audiotape (DAT)	20.0	48

TABLE 1.3 Video Sampling Rates

Format	Lines/Frame × Pixels/Line × Frames/Second =	Sampling Rate (million pixels per second)
CIF (videoconferencing)	360 × 288 × 30 =	3
CCIR (TV)	720 × 576 × 30 =	12
HDTV	1280 × 720 × 60 =	60

The second component, rate (in bits/second), is important. Bit rate can be considered to be made up of sampling rate (in samples/second) times accuracy (in bits/sample). Tables 1.2 and 1.3 show sampling rates for various audio and video sources (Jayant, Johnston, and Safranek 1993). The rates in Table 1.2 roughly represent sampling at the Nyquist rate (rate equal to twice the bandwidth of the source), although various other issues cause the sampling rate to be set above the minimum possible. Table 1.3 illustrates some important differences in audio and video. For video, there is a specified number of horizontal and vertical samples per frame, and then there is a temporal sampling in terms of frames per second. In voice applications, the sampling rates specified are not varied widely. However, in video, considerable license may be taken. Specifically, common routes to lower data rates for video are to reduce the frame size and/or lower the frame rate, both with attendant loss in quality.

The third component of a data compression problem is the fidelity criterion or distortion measure. For mathematical tractability in lossy coding, a preferred choice is the squared error distortion measure, but this criterion has a well-known mismatch with both audio and video subjective performance. As noted, one of the principal areas leading to recent performance leaps in compression algorithms is the use of more perceptually based distortion measures—for voice coding, high-quality audio coding, and for quantization table design for still images and video. As it turns out, how perceptual measures are incorporated is different in each of these cases, but the overriding performance advantage that perceptually based coding provides is undebatable. Subsequent chapters will bear this out.

1.2.3 Types of Compression Problems

There are two types of compression problems of interest (Davisson and Gray 1976). One problem is specified in terms of a constraint on transmitted data rate or storage capacity, and the problem is to compress the source at or below this rate but at the highest fidelity possible. This problem is sometimes called the *distortion-rate problem*. Examples of this problem are voice mail, digital cellular mobile radio, and videoconferencing. A second type of problem consists of the requirement to achieve a certain prespecified fidelity, and to satisfy this constraint with as few bits per second as possible. This problem is called the *rate-distortion problem*. Examples of this approach are CD-quality audio and motion-picture-quality video.

1.3 Input Source Formats

We are interested in the compression of data, voice, audio, still images, and video, and the specification of the specific source format, such as sampling rate, image size, color or black-and-white, and so on, has a tremendous bearing on what rate, distortion, and complexity combination is necessary. Of course, some applications may set these specifications automatically, such as in the compression of movies and high-quality audio. However, other applications, such as forms of teleconferencing, are wide open in terms of the source formats that we desire to employ, at least until a standard is set, as in H.320 and H.324.

For facsimile, the ITU-T standard has two resolutions, as shown in Table 1.4, and expects the image to be bi-level black on white. This type of specification is clearly a tremendous aid in designing successful data compression (data compaction) systems. Tables 1.1 through 1.3 presented other typical speech, audio, still-image, and video formats; Tables 1.5 and 1.6 provide additional examples for image and video applications.

For telephony, a sampling rate of 8000 samples/second is fairly universally accepted. For wideband speech, the sampling rate is usually 16,000 samples/second. For high-quality audio, there are some differences in sampling rate, as indicated for CD audio and DAT audio in Table 1.2. For still images and video, however, there are a host of formats and applications, as is abundantly

TABLE 1.4 ITU-T Facsimile Standards

Size	Vertical Resolution (lines/mm)	Horizontal Resolution (pixels/mm)	Lines/Frame	Pixels/Line
Normal resolution 20.7 cm (8.27 inches) by 29.2 cm (11.7 inches)	3.85	8	1188	1728
High resolution 20.7 cm (8.27 inches) by 29.2 cm (11.7 inches)	7.7	8	2376	1728

TABLE 1.5 H.324 Video Formats

Format	Pixels	H.261	H.263
SQCIF	128 × 96	optional	required
QCIF	176 × 144	required	required
CIF	352 × 288	optional	optional
4 CIF	704 × 576	n/a	optional
16 CIF	1408 × 1152	n/a	optional

clear from Table 1.6. Not only is the image size important, but the frame rate is critical as well, and again, it can have variations. In fact, it can have a greater variation than implied by Table 1.6, since this is one of the primary parameters that is adjusted in videoconferencing applications to stay within the required bit rate.

For color images, there are also the issues of color space and subsampling ratios of the components. There are at least two very common color representation formats, one consisting of the luminance and two chrominance components, designated YCbCr, and the other consisting of the three colors red-green-blue, designated RGB. The subsampling ratios between the components also vary. For example, the CCIR-601 format has an option for NTSC that specifies a spatial resolution of 720 × 480 pixels for the luminance component and 360 × 480 pixels for each of the two chrominance components. This subsampling format is denoted as 4:2:2. Alternatively, there is the SIF (source input format) specified by MPEG-1, where the luminance spatial resolution is 360 × 240 and the two chrominance components have a resolution of 180 × 120, which is denoted as 4:2:0 subsampling.

TABLE 1.6 Image and Video Formats

Formats	Usable Horizontal Lines*	Pixels per Line	Total Pixels per Frame	Frames per Second	Required Bandwidth/ Transmission Rate
Analog video					
NTSC (Americas, Asia)	338	426	150,000	29.97	4 MHz
PAL (Europe)	411	420	172,000	25.00	5 MHz
VHS	338	280	95,000	29.97	<4 MHz
Computer image					
SVGA	1024	768	786,500	60	—
VGA	640	480	307,000	60	—
Motion picture film					
35mm	(not a raster-		500,000	24	—
16mm	scanned image)		125,000	24	—
Digital video					
QCIF (H.261)	144	176	25,000	15–30	56 kbps–2 Mbps
CIF (H.261)	288	352	100,000	15–30	56 kbps–2 Mbps
HDTV	806	1920	1,550,000	50	140 Mbps
MPEG (constrained set)	345	360	124,000	30	1.5 Mbps and higher

*Eliminates retrace lines and includes the utilization ratio.

All in all, as this short section makes clear, there is quite a variety of source formats. In later sections and chapters discussing algorithms and standards, the details of the source format being employed are clearly stated. Matching the source format to the particular application is critical to the design of an efficient data compression system and to the success of the final product for that application.

1.4 Reconstructed Source Quality

Throughout the discussion thus far, and for almost the entire range of sources, we have emphasized that quality is an important requirement. Since these are real sources and there is a vast array of standards and applications, you might certainly wonder how the quality of the reconstructed source is measured and, further, how data compression systems are designed to achieve sufficiently high quality. These are two separate, but related, issues.

1.4.1 Performance Measurement

The measurement of coder performance is a long-standing and difficult problem, since what is desired is an objective or numerical indicator of performance that translates into what the user agrees is good quality. After a great deal of research and experience, some usable indicators of performance for telephone bandwidth speech coders have been developed. These are described in additional detail in Appendix A, but we mention a few here to orient you to their importance and to the limitations inherent in such indicators. The most widely quoted measure of speech coder performance is the Mean Opinion Score (MOS). An MOS for a coder is a value from 1 to 5, indicating the listeners' assessment of the reconstructed speech quality when presented with coded speech samples and asked to rate them with respect to the following scale: excellent (5), good (4), fair (3), poor (2), bad (1). The MOS values for a coder can vary across languages and between tests, but for G.711 log-PCM, an MOS of 4.0 to 4.2 is often quoted. Thus, this value would be associated with toll quality.

Another measure used to assess the performance of speech coders is the Diagnostic Rhyme Test (DRT). This test is primarily designed to determine speech intelligibility and consists of using trained listeners to evaluate coded pairs of words that differ in beginning or ending consonant sounds. The DRT score for a "good" speech coder is usually in the 85–90 range. There is also an assessment test for quality called the Diagnostic Acceptability Measure (DAM), which is a multidimensional test intended to evaluate medium- to high-quality speech. While the DAM scores can be useful, DAM has not been as widely accepted as the MOS and DRT in speech coder evaluations. Typical values of the MOS, DRT, and DAM for common speech coders are shown in Table 1.7 (Jayant 1992). You should only regard the values in Table 1.7 as approximate, since different implementations and test conditions can cause variations.

1.4.2 Perceptual Distortion Measures

We turn now to how perceptual effects can be incorporated into the design of compression systems for speech, audio, still images, and video. The specifics of the approaches are different in each case, but some of the concepts are the same. The broad idea is that inaccuracies in the reconstructed source can be

TABLE 1.7 DRT, DAM, and MOS Scores for Common Speech Coders

Coder	DRT	DAM	MOS
64-kbps PCM (Pulse Code Modulation)	95	73	4.2
32-kbps ADPCM (Adaptive Differential PCM)	94	68	4.0
16-kbps LD-CELP (Low-Delay Code Excited Linear Predictive Coding)	94	70	4.0
4.8-kbps CELP (Code Excited Linear Predictive Coding)	91	65	3.2
2.4-kbps LPC (Linear Predictive Coder—vocoder)	87	54	2.2

covered up, or *masked,* by the components in the source signal itself. The result is that by shaping the reconstruction error spectrum or spatial distribution in relation to the source, much larger errors can be absorbed without being objectionable to the user. In early telephony-based speech coding, this approach was called *noise spectral shaping,* but more recently and in high-quality audio, this approach is called *auditory masking.* The real power of the analysis-by-synthesis speech coders lies in the perceptually based distortion measure that chooses the excitation by minimizing a distortion measure that weights errors differently according to their relationship to the input spectral content.

For high-quality audio, where real-time operation may not be required and additional encoder complexity may be acceptable, even more effort is placed on exploiting the masking effects. In particular, the input audio spectrum is accurately calculated and a masking threshold is determined using known results from auditory masking experiments. The result is transparent-sounding audio at surprisingly low bit rates.

For still images and video, the incorporation of perceptual effects tends to come during the design of the encoder quantization tables. Extensive perceptual experiments are performed over a representative set of input image or video sources, and the quantizer characteristics are adjusted to minimize visual distortion. The final result can be that very simple scalar quantizers may achieve acceptable visual quality at bit rates often associated with vector quantization.

The importance of perceptually based distortion measures on data compression system performance cannot be overemphasized. Virtually all successful data compression systems today have this principle at their core.

1.5 System Issues and Performance Comparisons

In Section 1.2, we noted that the three major components of a data compression problem are the source, the rate, and the distortion measure. Further, the two types of problems of interest in compression applications, rate distortion and distortion rate, were also outlined in that section. Section 1.3 introduced several possible source formats and pointed out the importance of a clear specification of the source to the success of a compression algorithm and product. We have also mentioned in previous sections that the introduction of perceptually based distortion measures has been instrumental in the widespread utilization of compression in applications.

If the only consideration in designing and evaluating data compression systems is achieving acceptable performance at a specified rate, the problem may be very easy indeed. For example, part of the ITU-T G.726 specification is to achieve toll-quality speech coding at 32 kbps. It turns out that these two requirements alone on rate and quality are very simple to achieve. However, that is not all that is required by the specification. Other requirements include low delay, less than 5 ms (not too difficult); moderate to low complexity (still doable); tandem connections with other speech coders (probably not too hard at this rate); acceptable performance with an independent bit error rate up to 0.01 (more challenging); and pass signaling tones and some voiceband data modem signals with good performance (nontrivial). So, for most problems, if we are lucky enough only to have to address rate and distortion performance, the problem is much simplified. However, this is seldom the case, and these system and network issues, as we call them, can completely dominate the compression system approach and design. In this section, we call attention to several of the surrounding issues that become important in the subsequently considered compression standards and applications.

The voice telephony issues that have just been touched upon continue to be important today. In these telephony applications, the usual desire is toll quality, which is considered to be equivalent to 64 kbps G.711, and as low a complexity implementation as possible, although implementations on fixed or floating-point DSPs become acceptable as the rates are pushed down to 16 kbps and below. The low-delay requirement follows from the desire to eliminate or reduce the use of echo cancelers, but again, as rates are pushed below 16 kbps, the ITU-T has been more forgiving on this specification. For telephony applications, tandeming (or interoperability) is an issue that will not go away, since it will always be a requirement to operate with voice coders in other parts of the wired network and, of course, to interoperate with speech coders in various digital cellular mobile radio systems. The necessity for speech coders to perform adequately over channels that introduce bit errors is also a common requirement, whether one expects independent bit errors at rates of 10^{-9} or 10^{-2} or a fading channel that can produce burst errors.

For digital cellular mobile radio systems, the low-delay requirement is not present because burstiness of the channel requires interleaving, which already inserts substantial delay. A requirement that is implicit in telephony—and that takes on additional importance in digital cellular and, for that matter, in developing personal communications systems (PCS) applications—is the need for good speech coder performance in the presence of background noise and competing sounds such as road noise. Also, in the mobile environment, complexity is an issue in terms of battery power usage.

Not all voice coding applications have such stringent requirements. Voice messaging needs good quality, often called *communications quality,* but not toll quality; low-to-moderate complexity; and usually has no bit error rate requirements—although we might imagine the need to send digitized voice mail over a noisy channel. Additionally, low delay is not an issue, although excessive delay might be a problem because of intermediate storage needs.

One of the principal applications for wideband speech is telephony-based videoconferencing, hence many of the telephone-band voice coding requirements may carry forward. In general, however, low delay is not imposed because additional delay is often added to the coded speech to synchronize with the compressed video, which takes longer to code. Channel errors can still be a problem, but quality, rate, and complexity are dominant issues in such applications.

High-quality audio for storage applications usually involves non-real-time encoding and very low error rate channels, so the principal issues are quality, rate, and decoder complexity. Of course, the requirement might be added in some applications for remote encoding and broadcastlike transmission to users.

Still-image compression is often applied to image storage and retrieval, so the issues are quality, bit rate, and complexity, especially decoder complexity. However, as is evident from applications of the JPEG standard, compressed imagery may be transmitted over a variety of nonideal channels, and so the effects of bit errors introduced by a channel could become important.

Video compression for video telephony and videoconferencing applications needs to be the highest possible quality for the several stated bit rates, not too complex (but most coders are fairly complicated), with as low a delay as possible (but certainly not what is called "low delay" for speech coders, 5 ms or less), and able to operate with bit errors.

For MPEG-1 and MPEG-2 stored video applications, quality, rate, and decoder complexity are paramount. For most of their envisioned applications, bit errors during transmission are not a major issue, but innovative applications are already employing these standards in transmission environments. For stored video playback, issues derived from VCR-type features, such as fast forward, rewind, pause, and fast search, are critical.

This is but a brief view of systems and operational issues that may dominate a particular application of compression techniques or be part of a standards specification. The goal is not to be exhaustive, but suggestive of the types of requirements that arise in current applications and that may arise as engineers and entrepreneurs develop new products and services.

1.6 Applications and Standards

There are a host of known applications for data compression techniques and systems, and new applications are being proposed each day, creating a demand for modifications of existing compression methods or entirely new approaches. Existing applications of compression methods include facsimile, voice mail, telephony, digital cellular mobile radio, personal communications systems, CD-quality audio, still-image archival, videoconferencing, and video and movie distribution, just to name a few. These applications have led to a plethora of compression standards—agreed-upon techniques and systems-level specifications that are adopted by interested industry representatives to allow the manufacture of compatible equipment.

There are numerous active standards-setting bodies—for example, the International Telecommunication Union (ITU), the U.S. ANSI Committee T1 on Telecommunications, the Telecommunications Industry Association (TIA), the European Telecommunications Standards Institute (ETSI), the Japanese Telecommunications Technology Committee (TIC), the Institute of Electrical and Electronics Engineers (IEEE), and the International Standards Organization (ISO). Each of these has had a role in setting standards for data compression. The standards efforts have been phenomenally successful. Important standards have been produced, and in general, most of the standards development efforts have moved along at a relatively rapid pace. Additionally, the standards bodies and the particular standards committees have been responsive to perceived needs and requested modifications.

Tables 1.8 and 1.9 contain abridged lists of voice, still-image, and video coding standards. The sheer variety of standards being set and the dates of their adoption are indicative of the level of activity in source compression. Plus, most

TABLE 1.8 Speech Coder Standards

Description	Year of Introduction	Bit Rates (kbps)	MOS
PCM (for PSTN)	1972	64	4.4
LPC-10 (U.S. Fed. Std. 1015)	1976	2.4	2.7
G.721 ADCPM (for PSTN)	1984	32	4.1
INMARSAT (satellite)	1990	4.15	≈3.2
GSM (European cellular)	1991	13	3.6
CELP (U.S. Fed. Std. 1016)	1991	4.8	3.2
G.728 (low-delay CELP)	1992	16	4.0
VSELP (NA cellular)	1992	8	3.5
QCELP (NA CDMA)	1993	1–8	≈3.4
VSELP (Japanese cellular)	1993	6.8	≈3.3
G.729 (new toll-quality)	1995	8	≈4.2
G.723.1 (in H.323 and H.324)	1995	6.3	3.98
Half-rate GSM	1995	5–6	≈3.4
New low-rate U.S. Fed. Std.	1996	2.4	≈3.3

TABLE 1.9 Image/Video Compression Standards

Source	Standard	Rates
Video telephone Px64	ITU-T H.261	56 kbps–2 Mbps
Black-and-white, color, multispectral images	JPEG	0.25–2 bits/pixel
Moving pictures and audio	MPEG-1	1.5 Mbps
Broadcast-quality pictures and audio	MPEG-2	6–10 Mbps
High-quality audio for MPEG	HDTV	64/128/192 kbps per channel
Video	H.263	≤28.8 kbps

of these standards are set in response to a perceived need for a service or product. There have been some fears that setting a standard in a particular area would inhibit research and somehow stifle the field. In general, this has not been the case. The community at large is very responsive to improvements, and engineers, managers, and entrepreneurs are always finding new applications for standards that were not previously envisioned and that lead to further refinements and research. The principal challenge is the exceedingly fast pace of activity.

1.7 Outline of the Book

The major steps in data compression are summarized in Figure 1.1. The source may first go through a redundancy removal stage, which often consists of time domain prediction or frequency domain transforms or their equivalent. The parameters generated by the redundancy removal step are then passed to an entropy reduction step, which is some version of quantization. Finally, the quantizer outputs are losslessly encoded and sent to the storage device or transmitted over the channel. Reconstruction consists of decoding in the reverse direction.

The development of the basic principles in this book starts with lossless coding, then considers entropy reduction, and then finally redundancy removal. Note that if we begin with what is called a discrete memoryless source, such as independent samples from a data string, the redundancy removal and entropy reduction steps are not needed, and only lossless coding is involved. If we have discrete-time, continuous-amplitude samples of a memoryless source, then we need both the entropy reduction and lossless coding steps, but not redundancy removal. For sources like voice, video, and audio, we need all three steps, including redundancy removal, for efficient compression.

Chapter 2 covers basic details of lossless coding, along with a treatment of Huffman coding. Chapter 3 follows with a development of universal lossless coding, including Lempel-Ziv and arithmetic coding. Chapter 4 presents the essential ideas and techniques for scalar, adaptive, and vector quantization. Predictive coding, particularly as employed in the numerous telephony and digital cellular speech coding standards, is introduced in Chapter 5. Chapter 6 describes the existing and evolving speech coding standards based upon predictive coding, giving performance comparisons, system-level assumptions,

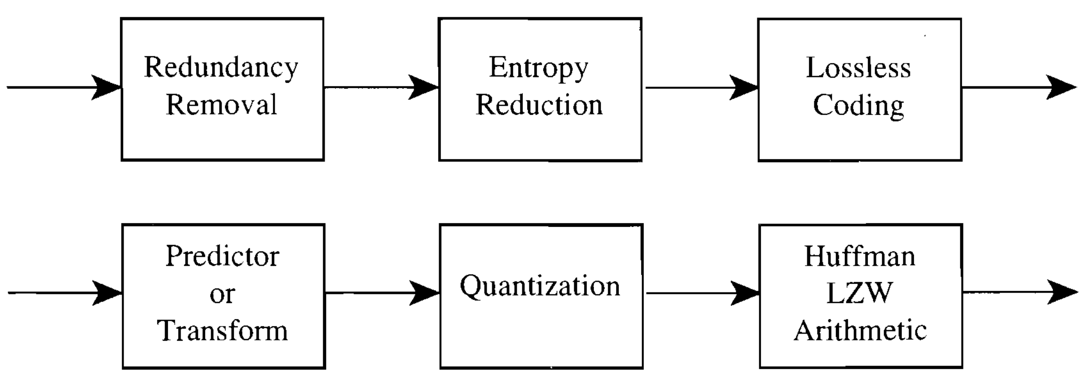

FIGURE 1.1 Major Steps in Data Compression

and their relative complexity. The fundamentals of frequency domain coding, including subband, discrete transform, wavelet, and fractal decompositions, are presented in Chapter 7. Chapter 8 then describes how these frequency domain ideas are incorporated in speech and high-quality audio compression algorithms and systems. The JPEG still-image compression standard is covered in Chapter 9 with discussions of baseline JPEG, the progressive transmission mode, and the lossless mode. The H.320, H.323, and H.324 standards for videoconferencing are presented in Chapter 10, which also contains video and audio compression capabilities as well as important system operational details. Chapter 11 discusses the MPEG-1, MPEG-2, and MPEG-4 standards, including compression details, bitstream syntax, and system-level features.

The general approach of this book is to provide the requisite background details concerning lossless coding, quantization, and redundancy removal and then go over the standards that utilize these techniques. This book is not intended to be a research monograph or an introductory textbook, so the material presented was selected based upon what is being used today, what are the fundamental underpinnings of today's methods, and what knowledge is needed to track evolving capabilities.