Digital Video Processing

PRENTICE HALL SIGNAL PROCESSING SERIES

Alan V. Oppenheim, Series *Editor*

ANDREWS & **HUNT** *Digital Image Restoration*
BRACEWELL *Two Dimensional Imaging*
BRIGHAM *The Fast Fourier Transform and Its Applications*
BURDIC *Underwater Acoustic System Analysis 2/E*
CASTLEMAN *Digital Image Processing*
CROCHIERE & RABINER *Multirate Digital Signal Processing*
DUDGEON & MERSEREAU *Multidimensional Digital Signal Processing*
HAYKIN *Advances in Spectrum Analysis and Array Processing. Vols. I, II & III*
HAYKIN, **ED.** *Array Signal Processing*
JOHNSON & DUDGEON *Array Signal Processing*
KAY *Fundamentals of Statistical Signal Processing*
KAY *Modern Spectral Estimation*
KINO *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*
LIM *Two-Dimensional Signal and Image Processing*
LIM, ED. *Speech Enhancement*
LIM & OPPENHEIM, EDS. *Advanced Topics in Signal Processing*
MARPLE *Digital Spectral Analysis with Applications*
MCCLELLAN & RADER *Number Theory in Digital Signal Processing*
MENDEL *Lessons in Estimation Theory for Signal Processing Communications and Control 2/E*
NIKIAS *Higher Order Spectra Analysis*
OPPENHEIM & NAWAB *Symbolic and Knowledge-Based Signal Processing*
OPPENHEIM, WILLSKY, **WITH YOUNG** *Signals and Systems*
OPPENHEIM & SCHAFER *Digital Signal Processing*
OPPENHEIM & SCHAFER *Discrete-Time Signal Processing*
PHILLIPS & NAGLE *Digital Control Systems Analysis and Design, 3/E*
PICINBONO *Random Signals and Systems*
RABINER & GOLD *Theory and Applications of Digital Signal Processing*
RABINER & SCHAFER *Digital Processing* ***of*** *Speech Signals*
RABINER & JUANG *Fundamentals* of *Speech* Recognition
ROBINSON & TREITEL *Geophysical Signal Analysis*
STEARNS & DAVID *Signal Processing Algorithms in Fortran and C*
TEKALP *Digital Video Processing*
THERRIEN *Discrete Random Signals and Statistical Signal Processing*
TRIBOLET *Seismic Applications* of *Homomorphic Signal Processing*
VIADYANATHAN *Multirate Systems and Filter Banks*
WIDROW & STEARNS *Adaptive Signal Processing*

Digital Video Processing

A. Murat Tekalp
University of Rochester

Prentice Hall PTR
Upper Saddle River, NJ 07458

Tekalp, A. Murat.
Digital video processing / A. Murat Tekalp.
p. cm. -- (Prentice-Hall signal processing series)
ISBN 0-13-190075-7 (alk. paper)
1. Digital video. I. Title. II. Series.
TK6680.5.T45 1995
621.388'33--dc20 95-16650
CIP

Editorial/production supervision: **Ann Sullivan**
Cover design: **Design Source**
Manufacturing manager: **Alexis R. Heydt**
Acquisitions editor: **Karen Gettman**
Editorial assistant: **Barbara Alfieri**

Printed on Recycled Paper

Printed in the United States of America

1 0 9 8 7 6 5 4 3 2 1

ISBN: 0-13-190075-7

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., Mexico
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

To *Sevim* and *Kaya Tekalp, my* **mom and** dad
and *to Özge, my beloved wife*

nal directly using one A/D converter with a clock high enough to leave the **color** subcarrier components free from aliasing, and then perform digital decoding to obtain the desired RGB or YIQ component signals. This requires sampling at a rate three or four times the color subcarrier frequency, which can be accomplished by special-purpose chip sets. Such chips do exist in some advanced TV sets for digital processing of the received signal for enhanced image quality.

The horizontal and vertical resolution of digital video is related to the number of pixels per line and the number of lines per frame. The artifacts in digital video due to lack of resolution are quite different than those in analog video. In analog video the lack of spatial resolution results in blurring of the image in the respective direction. In digital video, we have pixellation (aliasing) artifacts due to lack of sufficient spatial resolution. It manifests itself as jagged edges resulting from individual pixels becoming visible. The visibility of the pixellation artifacts depends on the size of the display and the viewing distance [Lut 88].

The arrangement of pixels and lines in a contiguous region of the memory is called a bitmap. There are five key parameters of a bitmap: the starting address in memory, the number of pixels per line, the pitch value, the number of lines, and number of bits per pixel. The pitch value specifies the distance in memory from the start of one line to the next. The most common use of pitch different from the number of pixels per line is to set pitch to the next highest power of 2, which may help certain applications run faster. Also, when dealing with interlaced inputs, setting the pitch to double the number of pixels per line facilitates writing lines from each field alternately in memory. This will form a "composite frame" in a contiguous region of the memory after two vertical scans. Each component signal is usually represented with 8 bits per pixel to avoid "contouring artifacts." Contouring results in slowly varying regions of image intensity due to insufficient bit resolution. Color mapping techniques exist to map 2^{24} distinct colors to 256 colors for display on 8-bit color monitors without noticeable loss of color resolution. Note that display devices are driven by analog inputs; therefore, D/A converters are used to generate component analog video signals from the bitmap for display purposes.

The major bottleneck preventing the widespread use of digital video today has been the huge storage and transmission bandwidth requirements. For example, digital video requires much higher data rates and transmission bandwidths as compared to digital audio. CD-quality digital audio is represented with 16 bits/sample, and the required sampling rate is 44kHz. Thus, the resulting data rate is approximately 700 kbits/sec (kbps). In comparison, a high-definition TV signal (e.g., the AD-HDTV proposal) requires 1440 pixels/line and 1050 lines for each luminance frame, and 720 pixels/line and 525 lines for each chrominance frame. Since we have 30 frames/s and 8 bits/pixel per channel, the resulting data rate is approximately 545 Mbps, which testifies that a picture is indeed worth 1000 words! Thus, the viability of digital video hinges upon image compression technology [Ang 91]. Some digital video format and compression standards will be introduced in the next subsection.

1.2.2 Digital Video Standards

Exchange of digital video between different applications and products requires digital video format standards. Video data needs to be exchanged in compressed form, which leads to compression standards. In the computer industry, standard display resolutions; in the TV industry, digital studio standards; and in the communications industry, standard network protocols have already been established. Because the advent of digital video is bringing these three industries ever closer, recently standardization across the industries has also started. This section briefly introduces some of these standards and standardization efforts.

Table 1.1: Digital video studio standards

Parameter	CCIR601 525/60 NTSC	CCIR601 625/50 PAL/SECAM	CIF
Number of active pels/line			
Lum (Y)	720	720	360
Chroma (U,V)	360	360	180
Number of active lines/pic			
Lum (Y)	480	576	288
Chroma (U,V)	480	576	144
Interlacing	2:1	2:1	1:1
Temporal rate	60	50	30
Aspect ratio	4:3	4:3	4:3

Digital video is not new in the broadcast TV studios, where editing and special effects are performed on digital video because it is easier to manipulate digital images. Working with digital video also avoids artifacts that would be otherwise caused by repeated analog recording of video on tapes during various production stages. Another application for digitization of analog video is conversion between different analog standards, such as from PAL to NTSC. CCIR (International Consultative Committee for Radio) Recommendation 601 defines a digital video format for TV studios for 525-line and 625-line TV systems. This standard is intended to permit international exchange of production-quality programs. It is based on component video with one luminance (Y) and two color difference (Cr and Cb) signals. The sampling frequency is selected to be an integer multiple of the horizontal sweep frequencies in both the 525- and 625-line systems. Thus, for the luminance component,

$$f_{s,lum} = 858 f_{h,525} = 864 f_{h,625} = 13.5\text{MHz}, \tag{1.10}$$

and for the chrominance,

$$f_{s,chr} = f_{s,lum}/2 = 6.75\text{MHz}. \quad (1.11)$$

The parameters of the CCIR 601 standards are tabulated in Table 1.1. Note that the raw data rate for the CCIR 601 formats is 165 Mbps. Because this rate is too high for most applications, the CCITT (International Consultative Committee for Telephone and Telegraph) Specialist Group (SGXV) has proposed a new digital video format, called the Common Intermediate Format (CIF). The parameters of the CIF format are also shown in Table 1.1. Note that the CIF format is progressive (noninterlaced), and requires approximately 37 Mbps. In some cases, the number of pixels in a line is reduced to 352 and 176 for the luminance and chrominance channels, respectively, to provide an integer number of 16 x 16 blocks.

In the computer industry, standards for video display resolutions are set by the Video Electronics Standards Association (VESA). The older personal computer (PC) standards are the VGA with 640 pixels/line x 480 lines, and TARGA with 512 pixels/line x 480 lines. Many high-resolution workstations conform with the S-VGA standard, which supports two main modes, 1280 pixels/line x 1024 lines or 1024 pixels/line x 768 lines. The refresh rate for these modes is 72 frames/sec. Recognizing that the present resolution of TV images is well behind today's technology, several proposals have been submitted to the Federal Communications Commission (FCC) for a high-definition TV standard. Although no such standard has been formally approved yet, all proposals involve doubling the resolution of the CCIR 601 standards in both directions.

Table 1.2: Some network protocols and their bitrate regimes

Network	Bitrate
Conventional Telephone	0.3-56 kbps
Fundamental BW Unit of Telephone (DS-0)	56 kbps
ISDN (Integrated Services Digital Network)	64-144 kbps (px64)
Personal Computer LAN (Local Area Network)	30 kbps
T-1	1.5 Mbps
Ethernet (Packet-Based LAN)	10 Mbps
Broadband ISDN	100-200 Mbps

Various digital video applications, e.g., all-digital HDTV, multimedia services, videoconferencing, and videophone, have different spatio-temporal resolution requirements, which translate into different bitrate requirements. These applications will most probably reach potential users over a communications network [Sut 92]. Some of the available network options and their bitrate regimes are listed in Table 1.2. The main feature of the ISDN is to support a wide range of applications

over the same network. Two interfaces are defined: basic access at 144 kbps, and primary rate access at 1.544 Mbps and 2.048 Mbps. As audio-visual telecommunication services expand in the future, the broadband integrated services digital network (B-ISDN), which will provide higher bitrates, is envisioned to be the universal information "highway" [Spe 91]. Asynchronous transfer mode (ATM) is the target transfer mode for the B-ISDN [Onv 94].

Investigation of the available bitrates on these networks and the bitrate requirements of the applications indicates that the feasibility of digital video depends on how well we can compress video images. Fortunately, it has been observed that the quality of reconstructed CCIR 601 images after compression by a factor of 100 is comparable to analog videotape (VHS) quality. Since video compression is an important enabling technology for development of various digital video products, three video compression standards have been developed for various target bitrates, and efforts for a new standard for very-low-bitrate applications are underway. Standardization of video compression methods ensures compatibility of digital video equipment by different vendors, and facilitates market growth. Recall that the boom in the fax market came after binary image compression standards. Major world standards for image and video compression are listed in Table 1.3.

Table 1.3: World standards for image compression.

Standard	Application
CCITT G3/G4	Binary images (nonadaptive)
J B I G	Binary images
JPEG	Still-frame gray-scale and color images
H.261	p x 64 kbps
MPEG-1	1.5 Mbps
MPEG-2	10-20 Mbps
MPEG-4	4.8-32 kbps (underway)

CCITT Group 3 and 4 codes are developed for fax image transmission, and are presently being used in all fax machines. JBIG has been developed to fix some of the problems with the CCITT Group 3 and 4 codes, mainly in the transmission of halftone images. JPEG is a still-image (monochrome and color) compression standard, but it also finds use in frame-by-frame video compression, mostly because of its wide availability in VLSI hardware. CCITT Recommendation H.261 is concerned with the compression of video for videoconferencing applications over ISDN lines. The target bitrates are p x 64 kbps, which are the ISDN rates. Typically, videoconferencing using the CIF format requires 384 kbps, which corresponds to p = 6. MPEG-1 targets 1.5 Mbps for storage of CIF format digital video on CD-ROM and hard disk. MPEG-2 is developed for the compression of higher-definition video at 10-20 Mbps with HDTV as one of the intended applications. We will discuss digital video compression standards in detail in Chapter 23.

1.3 Digital Video Processing

Digital video processing refers to manipulation of the digital video bitstream. All known applications of digital video today require digital processing for data compression, In addition, some applications may benefit from additional processing for motion analysis, standards conversion, enhancement, and restoration in order to obtain better-quality images or extract some specific information.

Digital processing of still images has found use in military, commercial, and consumer applications since the early 1960s. Space missions, surveillance imaging, night vision, computed tomography, magnetic resonance imaging, and fax machines are just some examples. What makes digital video processing different from still image processing is that video imagery contains a significant amount of temporal correlation (redundancy) between the frames. One may attempt to process video imagery as a sequence of still images, where each frame is processed independently. However, utilization of existing temporal redundancy by means of multiframe processing techniques enables us to develop more effective algorithms, such as motion-compensated filtering and motion-compensated prediction. In addition, some tasks, such as motion estimation or the analysis of a time-varying scene, obviously cannot be performed on the basis of a single image. It is the goal of this book to provide the reader with the mathematical basis of multiframe and motion-compensated video processing. Leading algorithms for important applications are also included.

Part 1 is devoted to the representation of full-motion digital video as a form of computer data. In Chapter 2, we model the formation of time-varying images as perspective or orthographic projection of 3-D scenes with moving objects. We are mostly concerned with 3-D rigid motion; however, models can be readily extended to include 3-D deformable motion. Photometric effects of motion are also discussed. Chapter 3 addresses spatio-temporal sampling on 3-D lattices, which covers several practical sampling structures including progressive, interlaced, and quincunx sampling. Conversion between sampling structures without making use of motion information is the subject of Chapter 4.

Part 2 covers nonparametric 2-D motion estimation methods. Since motion compensation is one of the most effective ways to utilize temporal redundancy, 2-D motion estimation is at the heart of digital video processing. 2-D motion estimation, which refers to optical flow estimation or the correspondence problem, aims to estimate motion projected onto the image plane in terms of instantaneous pixel velocities or frame-to-frame pixel correspondences. We can classify nonparametric 2-D motion estimation techniques as methods based on the optical flow equation, block-based methods, pel-recursive methods, and Bayesian methods, which are presented in Chapters 5-8, respectively.

Part 3 deals with 3-D motion/structure estimation, segmentation, and tracking. 3-D motion estimation methods are based on parametric modeling of the 2-D optical flow field in terms of rigid motion and structure parameters. These parametric models can be used for either 3-D image analysis, such as in object-based image compression and passive navigation, or improved 2-D motion estimation. Methods

that use discrete point correspondences are treated in Chapter 9, whereas optical-flow-based or direct estimation methods are introduced in Chapter 10. Chapter 11 discusses segmentation of the motion field in the presence of multiple motion, using direct methods, optical flow methods, and simultaneous motion estimation and segmentation. Two-view motion estimation techniques, discussed in Chapters 9-11, have been found to be highly sensitive to small inaccuracies in the estimates of point correspondences or optical flow. To this effect, motion and structure from stereo pairs and motion tracking over long monocular or stereo sequences are addressed in Chapter 12 for more robust estimation.

Filtering of digital video for such applications as standards conversion, noise reduction, and enhancement and restoration is addressed in Part 4. Video filtering differs from still-image filtering in that it generally employs motion information. To this effect, the basics of motion-compensated filtering are introduced in Chapter 13. Video images often suffer from graininess, especially when viewed in freeze-frame mode. Intraframe, motion-adaptive, and motion-compensated filtering for noise suppression are discussed in Chapter 14. Restoration of blurred video frames is the subject of Chapter 15. Here, motion information can be used in the estimation of the spatial extent of the blurring function. Different digital video applications have different spatio-temporal resolution requirements. Appropriate standards conversion is required to ensure interoperability of various applications by decoupling the spatio-temporal resolution requirements of the source from that of the display. Standards conversion problems, including frame rate conversion and de-interlacing (interlaced to progressive conversion), are covered in Chapter 16. One of the limitations of CCIR 601, CIF, or smaller-format video is the lack of sufficient spatial resolution. In Chapter 17, a comprehensive model for low-resolution video acquisition is presented as well as a novel framework for superresolution which unifies most video filtering problems.

Compression is fundamental for all digital video applications. Parts 5 and 6 are devoted to image and video compression methods, respectively. It is the emergence of video compression standards, such as JPEG, H.261, and MPEG and their VLSI implementations, that makes applications such as all-digital TV, multimedia, and videophone a reality. Chapters 18-21 cover still-image compression methods, which form the basis for the discussion of video compression in Chapters 22-24, In particular, we discuss lossless compression in Chapter 18, DPCM and transform coding in Chapter 19, still-frame compression standards, including binary and gray-scale/color image compression standards, in Chapter 20, and vector quantization and subband coding in Chapter 21. Chapter 22 provides a brief overview of interframe compression methods. International video compression standards such as H.261, MPEG-1, and MPEG-2 are explained in Chapter 23. Chapter 24 addresses very-low-bitrate coding using object-based methods. Finally, several applications of digital video are introduced in Chapter 25.

Bibliography

[Ang 91] P. H. Ang, P. A. Ruetz, and D. Auld, "Video compression makes big gains," *IEEE Spectrum,* pp. 16-19, Oct. 1991.

[Byt 92] "Practical desktop video," *Byte,* Apr. 1992.

[Hsi 93] T. R. Hsing, C.-T. Chen, and J. A. Bellisio, "Video communications and services in the copper loop," *IEEE* Comm. Mag., pp. 63-68, Jan. 1993.

[Lat 89] B. P. Lathi, Modern *Digital* and Analog Communication *Systems,* Second Edition, HRW Saunders, 1989.

[Lip 90] A. Lippman, "HDTV sparks a digital revolution," *Byte,* Dec. 1990.

[Lut 88] A. C. Luther, *Digital Video* in *the PC* Environment, New York, NY: McGraw-Hill, 1988.

[Mil 92] G. M. Miller, *Modern Electronic* Communication, Fourth Edition, Regents, Prentice Hall, 1992.

[Net 89] A. N. Netravali and B. G. Haskell, *Digital* Pictures - Representation *and Compression,* New York, NY: Plenum Press, 1989.

[Onv 94] R. 0. Onvural, Asynchronous Transfer *Mode Networks:* Performance *Issues,* Norwood, MA: Artech House, 1994.

[Pro 94] J. G. Proakis and M. Salehi, Communication *Systems Engineering,* Englewood Cliffs, NJ: Prentice Hall, 1994.

[Spe 91] "B-1SDN and how it works," *IEEE Spectrum,* pp. 39-44, Aug. 1991.

[Spe 92] "Digital video," *IEEE Spectrum,* pp. 24-30, Mar. 1992.

[Spe 93] "Special report: Interactive multimedia," *IEEE* Spectrum, pp. 22-39, Mar. 1993.

[Spe 95] "Digital Television," *IEEE* Spectrum, pp. 34-80, Apr. 1995.

[Sut 92] J. Sutherland and L. Litteral, "Residential video services," *IEEE Comm. Mag.*, pp. 37–41, July 1992.

[Tru 93] H. J. Trussell, "DSP solutions run the gamut for color systems," *IEEE* Signal *Processing Mag.,* pp. 8-23, Apr. 1993.

Chapter 2

TIME-VARYING IMAGE FORMATION MODELS

In this chapter, we present models (in most cases simplistic ones) for temporal variations of the spatial intensity pattern in the image plane. We represent a time-varying image by a function of three continuous variables, $s_c(x_1, x_2, t)$, which is formed by projecting a time-varying three-dimensional (3-D) spatial scene into the two-dimensional (2-D) image plane. The temporal variations in the 3-D scene are usually due to movements of objects in the scene. Thus, time-varying images reflect a projection of 3-D moving objects into the 2-D image plane as a function of time. Digital video corresponds to a spatio-temporally sampled version of this time-varying image. A block diagram representation of the time-varying image formation model is depicted in Figure 2.1.

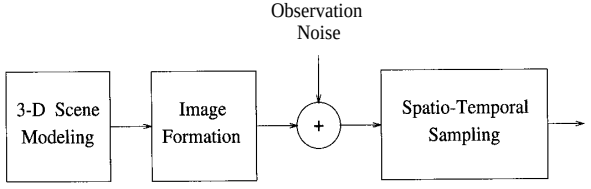

Figure 2.1: Digital video formation

In Figure 2.1, "3-D scene modeling" refers to modeling the motion and structure of objects in 3-D, which is addressed in Section 2.1. "Image formation," which includes geometric and photometric image formation, refers to mapping the 3-D scene into an image plane intensity distribution. Geometric image formation, discussed

Chapter 23

VIDEO COMPRESSION STANDARDS

Standardization of compressed digital video formats facilitates manipulation and storage of full-motion video as a form of computer data, and its transmission over existing and future computer networks, or over terrestrial broadcast channels. Envisioned areas of application for digital video compression standards include all-digital TV, videoconferencing, videophone, video mail, multimediastations, digital movies, video games, other forms of entertainment, and education. In this chapter, we discuss the international video compression standards for videoconferencing (H.261), multimedia (MPEG-l), and all-digital TV (MPEG-2) applications.

23.1 The H.261 Standard

ITU (CCITT) Recommendation H.261 is a video compression standard developed to facilitate videoconferencing and videophone services over the integrated services digital network (ISDN) at p x 64 kbps, p = 1, . , 30. For example, 64 kbps (p = 1) may be appropriate for a low quality videophone service, where the video signal can be transmitted at a rate of 48 kbps, and the remaining 16 kbps is used for the audio signal. Videoconferencing services generally require higher image quality, which can be achieved with $p \geq 6$, i.e., at 384 kbps or higher. Note that the maximum available bitrate over an ISDN channel is 1.92 Mbps ($p = 30$), which is sufficeint to obtain VHS-quality (or better) images.

ITU (CCITT) Recommendation H.261 has emerged as a result of studies performed within the European Project COST (Cooperation in the field of Scientific and Technical research) 2llbis during the period 1983-1990. In 1985, the COST 211bis videoconference hardware subgroup developed an initial codec operating at bit rates of n x 384 kbps, n = 1,. . . , 5, which was adopted in 1987 as ITU (CCITT) Recommendation H.120. Later, it became clear that a single standard can cover

all ISDN rates, p x 64 kbps, $p = 1, , 30$. The specifications of such a codec were completed in 1989, and the corresponding Recommendation H.261 was adopted by ITU (CCITT) in 1990.

In addition to forming a basis for the later video compression standards such as MPEG-1 and MPEG-2, the H.261 standard offers two important features: i) It specifies a maximum coding delay of 150 msec. because it is mainly intended for bidirectional video communication. It has been determined that delays exceeding 150 msec. do not give the viewer the impression of direct visual feedback. ii) It is amenable to low-cost VLSI implementation, which is rather important for widespread commercialization of videophone and teleconferencing equipment. The important aspects of the H.261 standard are summarized below. Further details can be found in [Lio 90, Lio 91, CCI 90].

23.1.1 Input Image Formats

To permit a single recommendation for use in and between regions using 625- and 525-line TV standards, the H.261 input picture format is specified as the so-called Common Intermediate Format (CIF). For lower-bitrate applications, a smaller format, QCIF, which is one-quarter of the CIF, has been adopted. The specifications of the CIF and QCIF formats are listed in Table 23.1, where the numbers in the parenthesis denote the modified specifications so that all four numbers are integer multiples of 8. Note that, at 30 frames/s, the raw data rate for the CIF is 37.3 Mbps, and for QCIF it is 9.35 Mbps. Even with QCIF images at 10 frames/s, 48:1 compression is required for videophone services over a 64 kbps channel. CIF images may be used when $p \geq 6$, that is, for videoconferencing applications. Methods for conversion to and from CIF/QCIF are not subject to recommendation.

Table 23.1: H.261 input image formats.

	CIF	QCIF
Number of active pels/line		
Lum (Y)	360 (352)	180 (176)
Chroma (U,V)	180 (176)	90 (88)
Number of active lines/pic		
Lum (Y)	288	144
Chroma (U,V)	144	72
Interlacing	1:1	1:1
Temporal rate	30, 15, 10, or 7.5	30, 15, 10 or 7.5
Aspect ratio	4:3	4:3

23.1.2 Video Multiplex

The video multiplex defines a data structure so that a decoder can interpret the received bit stream without any ambiguity. The video data is arranged in a hierarchical structure consisting of a picture layer, which is divided into several group-of-blocks (GOB) layers. Each GOB layer in turn consists of macroblocks (MB), which are made of blocks of pixels. Each layer has a header identifying a set of parameters used by the encoder in generating the bitstream that follows.

A macroblock is the smallest unit of data for selecting a compression mode. (The choices for the modes of compression are described below.) It consists of four 8 x8 (i.e., 16 pixels by 16 lines) of Y (luminance) and the spatially corresponding 8 x8 U and V (chrominance) blocks. Since the chrominance channels are subsampled in both directions, there is only one U and one V block for every four luminance blocks. The composition of a macroblock is shown in Figure 23.1.

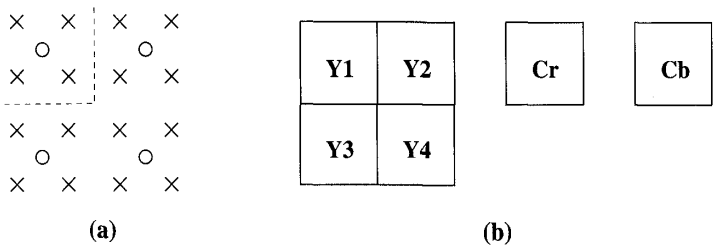

Figure 23.1: a) Positioning of luminance and chrominance pixels; b) the composition of a macroblock.

The GOB layer is always composed of 33 macroblocks, arranged as a 3 x 11 matrix, as depicted in Figure 23.2. Note that each MB has a header, which contains a MB address and the compression mode, followed by the data for the blocks.

Finally, the picture layer consists of a picture header followed by the data for GOBs. The picture header contains data such as the picture format (CIF or QCIF)

Table 23.2: The composition of the picture layer.

Format	Number of GOB in a frame	Number of MB in a GOB	Total number of MB in a frame
CIF	12	33	396
QCIF	3	33	99

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

(a)

MBA	MTYPE	MQUANT	MVD	CBP	Block data

(b)

Figure 23.2: a) Arrangement of macroblocks (MBs) in a GOB; b) structure of the MB layer [CCI 90].

and the frame number. Note that each CIF image has 12 GOBs, while the QCIF has only 3 GOBs, as shown in Table 23.2.

23.1.3 Video Compression Algorithm

The video compression scheme chosen for the H.261 standard has two main modes: the intra and inter modes. The *intra* mode is similar to JPEG still-image compression in that it is based on block-by-block DCT coding. In the *inter* mode, first a temporal prediction is employed with or without motion compensation (MC). Then the interframe prediction error is DCT encoded. Each mode offers several options, such as changing the quantization scale parameter, using a filter with MC, and so on. The algorithm can be summarized through the following steps:

1) Estimate a motion (displacement) vector for each MB. The standard does not specify a motion estimation method (in fact, MC is optional), but block matching based on 16 x 16 luminance blocks is generally used. The decoder accepts one integer-valued motion vector per MB whose components do not exceed $\pm$ 15.

2) Select a compression mode for each MB, based on a criterion involving the displaced block difference (dbd), which is defined by

$$dbd(\mathbf{x}, k) = b(x, k) - b(x - \mathbf{d}, k - 1) \quad (23.1)$$

where $b(.,.)$ denotes a block, x and k are pixel coordinates and time index, respectively, and $\mathbf{d}$ is the displacement vector defined for the kth frame relative to the $(k-1)$st frame. If $\mathbf{d}$ is set equal to zero, then

frames in a GOP are application-dependent, e.g., dependent on access time and bitrate requirements. The composition of a GOP is illustrated by an example.

Example

A GOP is shown in Figure 23.5 which is composed of nine pictures. Note that the first frame of each GOP is always an I-picture. In MPEG, the order in which the pictures are processed is not necessarily the same as their time sequential order. The pictures in Figure 23.5 can be encoded in one of the following orders:

0, 4, 1, 2, 3, 8, 5, 6, 7
or
0, 1, 4, 2, 3, 8, 5, 6, 7

since the prediction for P- and B-pictures should be based on pictures that are already transmitted.

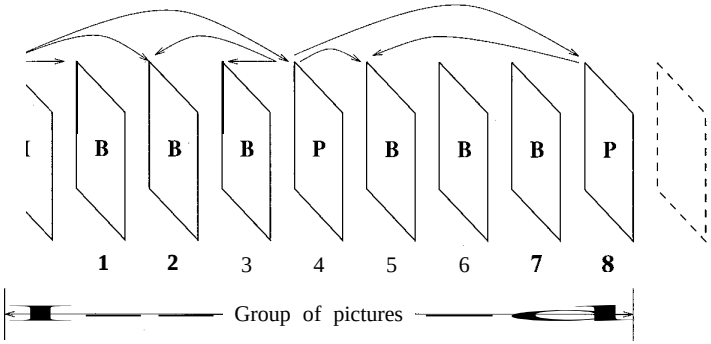

Figure 23.5: Group of pictures in MPEG-1.

4) *Slices* are made up of macroblocks. They are introduced mainly for error recovery.

5) The composition of *macroblocks* (MB) are the same as in the H.261 standard. Some compression parameters can be varied on a MB basis. The MB types depending on the choice of these parameters are listed in Table 23.5. We will take a closer look at each of these MB types in the following when we discuss the video compression algorithm.

6) *Blocks* are 8 x 8 pixel arrays. They are the smallest DCT unit.

Headers are defined for sequences, GOPs, pictures, slices, and MBs to uniquely specify the data that follows. For an extensive discussion of the MPEG-1 standard, the reader is referred to [ISO 91].

Table 23.5: Macroblock types in MPEG-1.

I-pictures	*P-pictures*	*B-pictures*
Intra	Intra	Intra
Intra-A	Intra-A	Intra-A
	Inter-D	Inter-F
	Inter-DA	Inter-FD
	Inter-F	Inter-FDA
	Inter-FD	Inter-B
	Inter-FDA	Inter-BD
	Skipped	Inter-BDA
		Inter-I
		Inter-ID
		Inter-IDA
		Skinned

23.2.4 Intraframe Compression Mode

The pixel intensity values are DCT encoded in a manner similar to JPEG and the intra mode of H.261. Compression is achieved by a combination of quantization and run-length coding of the zero coefficients.

Quantization

Assuming R-bit input images, the DC coefficient can take values in the range [0,2040], and the AC coefficients are in the range [-1024,1023]. These coefficients are quantized with a uniform quantizer. The quantized coefficient is obtained by dividing the DCT coefficient value by the quantization step size and then rounding the result to the nearest integer. The quantizer step size varies by the frequency, ac-

Table 23.6: MPEG default intra quantization matrix.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

corresponds to bidirectional prediction. This is illustrated in Figure 23.7. Note that in the bidirectional prediction mode, two displacement vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ and the corresponding prediction error **b-b** need to be encoded for each macroblock **b.**

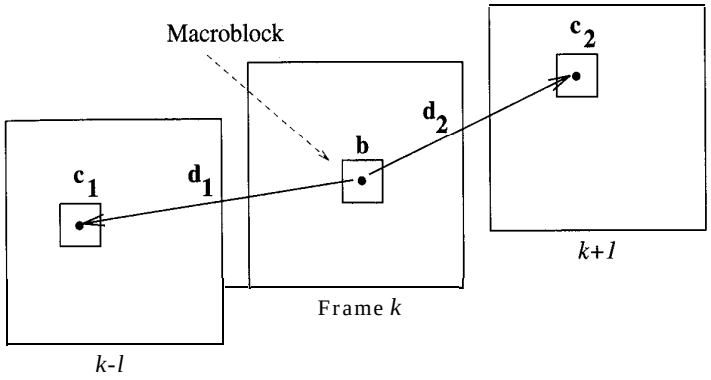

Figure 23.7: MPEG-1 bi-directional prediction.

The concept of bidirectional prediction or interpolative coding can be considered as a temporal multiresolution technique, where we first encode only the I- and P-pictures (typically 1/3 of all frames). Then the remaining frames can be interpolated from the reconstructed I and P frames, and the resulting interpolation error is DCT encoded. The use of B-pictures provides several advantages:

. They allow effective handling of problems associated with covered/uncovered background. If an object is going to be covered in the next frame, it can still be predicted from the previous frame or vice versa.

. MC averaging over two frames may provide better SNR compared to prediction from just one frame.

• Since B-pictures are not used in predicting any future pictures, they can be encoded with fewer bits without causing error propagation.

The trade-offs associated with using B-pictures are:

• Two frame-stores are needed at the encoder and decoder, since at least two reference (P and/or I) frames should be decoded first.

• If too many B-pictures are used, then i) the distance between the two reference frames increases, resulting in lesser temporal correlation between them, and hence more bits are required to encode the reference frames, and ii) we have longer coding delay.

The mode of compression for each MB in a B-picture is selected independently from the list of allowable modes shown in Table 23.5. Again, **"Intra"** and **"Intra-A"** MBs are coded independently of any reference frame. MBs classified as **"Inter"** have the following options: "D" indicates that the DCT of the prediction error will be coded, "F" indicates forward prediction with motion compensation, "B" indicates backward prediction with motion compensation, "I" indicates interpolated prediction with motion compensation, and "A" indicates adaptive quantization. A macroblock may be **"Skipped"** if the block from the previous frame is good enough as is; that is, no information needs to be sent.

Quantization and Coding

In the interframe mode, the inputs to the DCT are in the range [-255,255]; thus, all DCT coefficients have the dynamic range [-2048,2047]. The quantization matrix is such that the effective quantization is relatively coarser compared to those used for I-pictures. All quantized DCT coefficients, including the DC coefficient, are zigzag scanned to form [*run*, level] pairs, which are then coded using VLC. Displacement vectors are DPCM encoded with respect to the motion vectors of the previous blocks. VLC tables are specified for the type of MB, the differential motion vector, and the MC prediction error. Different Huffman tables are defined for encoding the macroblock types for P- and B-pictures, whereas the tables for motion vectors and the DCT coefficients are the same for both picture types.

23.2.6 MPEG-1 Encoder and Decoder

An MPEG-1 encoder includes modules for motion estimation, selection of compression mode (MTYPE) per MB, setting the value of MQUANT, motion-compensated prediction, quantizer and dequantizer, DCT and IDCT, variable-length coding (VLC), a multiplexer, a buffer, and a buffer regulator. The dequantizer and the IDCT are needed in the encoder because the predictions are based on reconstructed data. The IDCT module at the encoder should match within a prespecified tolerance the IDCT module at the decoder to avoid propogation of errors in the prediction process. This tolerance is specified in IEEE Standard 1180-1990 for 64-bit floating-point IDCT implementations.

The relative number of I-, P- or B-pictures in a GOP is application-dependent. The standard specifies that one out of every 132 pictures must be an I-picture to avoid error propagation due to IDCT mismatch between the encoder and decoder. The use of B-pictures is optional. Neither the motion estimation algorithm nor the criterion to select MYTPE and MQUANT are part of the standard. In general, motion estimation is performed using the luminance data only. A single displacement vector is estimated for each MB. One-half (0.5) pixel accuracy is allowed for motion estimates. The maximum length of the vectors that may be represented can be changed on a picture-by-picture basis to allow maximum flexibility. Motion vectors that refer to pixels outside the picture are not allowed.

In summary, a typical MPEG encoder performs the following steps:

1. Decide on the labeling of I-, P- and B-pictures in a GOP.
2. Estimate a motion vector for each MB in the P- and B-pictures,
3. Determine the compression mode MTYPE for each MB from Table 23.5.
4. Set the quantization scale, MQUANT, if adaptive quantization is selected

An MPEG-1 decoder reverses the operations of the encoder. The incoming bit stream (with a standard syntax) is demultiplexed into DCT coefficients and side information such as MTYPE, motion vectors, MQUANT, and so on. The decoder employs two frame-stores, since two reference frames are used to decode the B-pictures.

We conclude this section by summarizing the main differences between the H.261 and MPEG-1 standards in Table 23.7.

Table 23.7: Comparison of H.261 and MPEG-1 Standards

H.261	MPEG-1
Sequential access	Random access
One basic frame rate	Flexible frame rate
CIF and QCIF images only	Flexible image size
I and P frames only	I, P and B frames
MC over 1 frame	MC over 1 or more frames
1 pixel MV accuracy	1/2 pixel MV accuracy
121 filter in the loop	No filter
Variable threshold + uniform quantization	Quantization matrix
No GOF structure	GOF structure
GOB structure	Slice structure

23.3 The MPEG-2 Standard

The quality of MPEG-1 compressed video at 1.2 Mbps has been found unacceptable for most entertainment applications. Subjective tests indicate that CCIR 601 video can be compressed with excellent quality at 4-6 Mbps. MPEG-2 is intended as a compatible extension of MPEG-1 to serve a wide range of applications at various bitrates (2-20 Mbps) and resolutions. Main features of the MPEG-2 syntax are: i) it allows for interlaced inputs, higher-definition inputs, and alternative subsampling of the chroma channels, ii) it offers a scalable bitstream, and iii) it provides improved quantization and coding options.

Considering the practical difficulties with the implementation of the full syntax on a single chip, subsets of the full syntax have been specified under five "profiles," simple profile, main profile, SNR scalable profile, spatially scalable profile, and high profile. Furthermore, a number of "levels" have been introduced within these profiles to impose constraints on some of the video parameters [ISO 93]. It is important to note that the MPEG-2 standard has not yet been finalized. The syntax of the Main Profile was frozen in March 1993. However, work in other profiles is still ongoing. It is likely that the all-digital HDTV compression algorithm will conform with one of the profiles in MPEG-2.

In the following we discuss the MB structure in MPEG-2, how MPEG-2 handles interlaced video, the concepts related to scalability, and some extensions for encoding of higher-definition video along with a brief overview of the profiles and levels.

23.3.1 MPEG-2 Macroblocks

A macroblock (MB) refers to four 8 x 8 luminance blocks and the spatially associated chroma blocks. MPEG-2 allows for three chroma subsampling formats, 4:2:0 (same as MPEG-1), 4:2:2 (chroma subsampled in the horizontal direction only), and 4:4:4 (no chroma subsampling). The spatial locations of luma and chroma pixels for the 4:2:0 and 4:2:2 formats are depicted in Figure 23.8. Therefore, in MPEG-2 a MB may contain 6 (4 luma, 1 Cr, and 1 Cb), 8 (4 luma, 2 Cr, and 2 Cb), or 12 (4 luma, 4 Cr, and 4 Cb) 8 x 8 blocks.

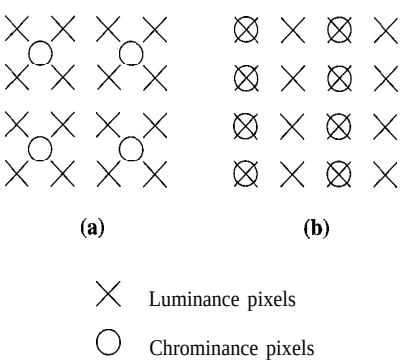

Figure 23.8: Chrominance subsampling options: a) 4:2:0 format; b) 4:2:2 format.

23.3.2 Coding Interlaced Video

MPEG-2 accepts both progressive and interlaced inputs. If the input is interlaced, the output of the encoder consists of a sequence of fields that are separated by the field period, There are two options in coding interlaced video: i) every field can be encoded independently (field pictures), or ii) two fields may be encoded together as a composite frame (frame pictures). It is possible to switch between frame pictures and field pictures on a frame-to-frame basis. Frame encoding is preferred for relatively still images; field encoding may give better results when there is significant motion. In order to deal with interlaced inputs effectively, MPEG-2 supports:

. two new picture formats: frame-picture and field-picture,
. field/frame DCT option per MB for frame pictures, and
. new MC prediction modes for interlaced video,

which are described in the following.

New Picture Types for Interlaced Video

Interlaced video is composed of a sequence of even and odd fields separated by a field period. MPEG-2 defines two new picture types for interlaced video. They are:

i) *Frame pictures,* which are obtained by interleaving lines of even and odd fields to form composite frames. Frame pictures can be I-, P-, or B-type. An MB of the luminance frame picture is depicted in Figure 23.9.

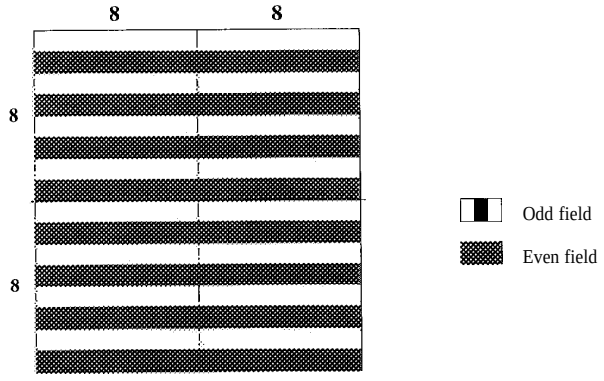

Figure 23.9: The luminance component of a MB of a frame picture.

ii) *Field pictures* are simply the even and odd fields treated as separate pictures. Each field picture can be I-, P- or B-type.

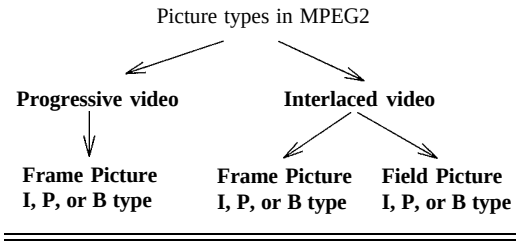

Figure 23.10: Summary of picture types in MPEG-2.

A summary of all picture types is shown in Figure 23.10. Clearly, in progressive video all pictures are frame pictures. A group of pictures can be composed of an arbitrary mixture of field and frame pictures. Field pictures always appear in pairs (called the top field and bottom field) which together constitute a frame. If the top field is a P- (B-) picture, then the bottom field must also be a P- (B-) picture. If the top field is an I-picture, then the bottom field can be an I- or a P-picture. A pair of field pictures are encoded in the order in which they should appear at the output. An example of a GOP for an interlaced video is shown in Figure 23.11.

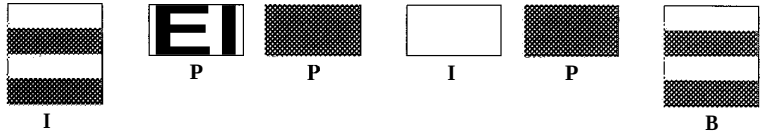

Figure 23.11: A GOP for an interlaced video

Field/Frame DCT Option for Frame Pictures

MPEG-2 allows a field- or frame-DCT option for each MB in a frame picture. This allows computing DCT on a field-by-field basis for specific parts of a frame picture. For example, field-DCT may be chosen for macroblocks containing high motion, whereas frame-DCT may be appropriate for macroblocks with little or no motion but containing high spatial activity. The internal organization of a MB for frame (on the left) and field (on the right) DCT is shown in Figure 23.12. Note that in 4:2:0 sampling, only frame DCT can be used for the chroma blocks to avoid 8 x 4 IDCT.

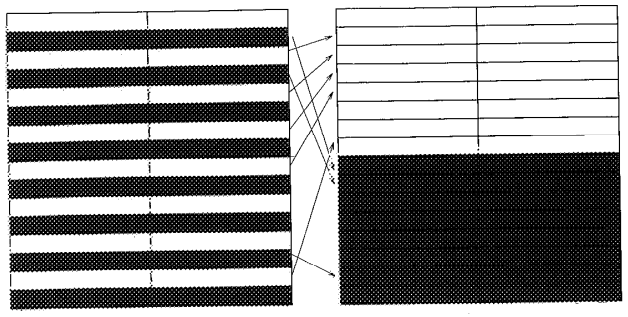

Figure 23.12: DCT options for interlaced frame pictures: a) Frame DCT and b) field DCT.

MC Prediction Modes for Interlaced Video

There are two main types of predictions: simple field and simple frame prediction. In simple field prediction, each field is predicted independently using data from one or more previously decoded fields. Simple frame prediction forms a prediction for an entire frame based on one or more previously decoded frames. Within a field picture only field predictions can be used. However, in a frame picture either field or frame prediction may be employed on an MB-by-MB basis. Selection of the best prediction mode depends on presence/absence of motion in an MB, since in the presence of motion, frame prediction suffers from strong motion artifacts, while in the absence of motion, field prediction does not utilize all available information.

There are also two other prediction modes: 16 x 8 MC mode and dual-prime mode. 16 x 8 MC mode is only used in field pictures, where two motion vectors are used per MB, one for the upper and the other for the lower 16 x 8 region, which belong to the top and bottom fields, respectively. In the case of bidirectional prediction, four motion vectors will be needed. In dual-prime mode, one motion vector and a small differential vector are encoded. In the case of field pictures, two motion vectors are derived from this information and used to form predictions from two reference fields, which are averaged to form the final prediction [ISO 93]. Dual-prime mode is used only for P-pictures.

23.3.3 Scalable Extensions

Scalability refers to ability to decode only a certain part of the bit-stream to obtain video at the desired resolution. It is assumed that decoders with different complexities can decode and display video at different spatio-temporal resolutions from the same bitstream. The minimum decodable subset of the bitstream is called the base layer. All other layers are enhancement layers, which improve the resolution of the base layer video. MPEG-2 syntax allows for two or three layers of video. There are different forms of scalability:

Spatial *(pixel* resolution) *scalability* provides the ability to decode video at different spatial resolutions without first decoding the entire frame and decimating it. The base layer is a low spatial resolution version of the video. Enhancement layers contain successively higher-frequency information. MPEG-2 employs a pyramidal coding approach. The base layer video is obtained by decimating the original input video. The enhancement layer is the difference of the actual input video and the interpolated version of the base-layer video.

SNR scalability offers decodability using different quantizer step sizes for the **DCT** coefficients. The base-layer video is obtained by using a coarse quantization of the DCT coefficients. It is at the same spatio-temporal resolution with the input video. The enhancement layer simply refers to the difference of the base layer and the original input video.

Temporal scalability refers to decodability at different frame rates without first decoding every single frame. *Hybrid scalability* refers to some combination of the above. An important advantage of scalability is that it provides higher resilience to transmission errors as the base-layer video is usually transmitted with better error correction capabilities.

23.3.4 Other Improvements

MPEG-2 also features some extensions in the quantization and coding steps for improved image quality in exchange to slightly higher bitrate. In particular, it allows for i) a new scanning scheme (alternate scan) in addition to the zigzag scanning of the DCT coefficients, ii) finer quantization of the DCT coefficients, iii) finer adjustment of the quantizer scale factor, and iv) a separate VLC table for the DCT coefficients for the intra macroblocks, which are explained in the following.

Alternate Scan

In addition to the zigzag scanning, **MPEG-2** allows for an optional scanning pattern, called the "alternate scan." The alternate scan pattern, which is said to fit interlaced video better, is depicted in Figure 23.13.

Finer Quantization of the DCT Coefficients

In intra macroblocks, the quantization weight for the DC coefficient can be 8, 4, 2 or 1. That is, 11 bits (full) resolution is allowed for the DC coefficient. Recall that this weight is fixed to 8 in MPEG-1. AC coefficients are quantized in the range [-2048,2047], as opposed to [-256,255] in MPEG-1. In non-intra macroblocks, all coefficients are quantized into the range [-2048,2047]. This range was [-256,255] in MPEG-1.

Besides videoconferencing and desktop videophone, video communication using an ISDN basic rate interface (BRI), operating at a total of 128 kbps, may be used for such applications as distance learning and access to multimedia information services. In distance learning, an instructor teaches students who are at remote locations, The students can interact with the instructor, by direct talking or sharing a whiteboard [Ald 93]. Multimedia information services include image-based electronic library systems and shopping-catalogs that can be browsed from home.

25.2 Interactive Video and Multimedia

Multimediacan mean different things to different people. Here it refers to the ability to provide full-motion interactive digital video in the personal computer (PC) or desktop workstation environment. Multimedia also involves text, graphics, sound, and still-images, which have long existed in the PC or workstation environment. The main components of a multimedia system are:

. data capture tools, such as video recorders and digitizers,

. data editors and authoring tools, for animation, audio and video editing, user-interface development, etc., and

- database storage and retrieval tools, for searching large databases, archiving, and backup.

The difficulty with full-motion video has been in the large data rates required. Multimedia workstations use CD-ROMs and hard disks for video storage at about 1.5 Mbps, which require approximately 30:1 compression of standard TV resolution images. The latest developments in image compression make this possible in real time. Two of the earliest initiatives in this area have been Intel's digital video interactive (DVI) technology, and compact disc-interactive (CD-I), jointly funded by NV Philips and Sony.

DVI technology is a general-purpose hardware for providing full-motion video in the PC environment, and uses CD-ROM for the storage medium. The design goal is to provide multimedia functionality at a cost suitable for desktop computers. It is based on two custom VLSI chips: the 82750PB pixel processor and the 82750DB display processor. The pixel processor mainly performs compression and decompression of images along with functions like YUV-RGB conversion, bilinear interpolation, and so on. It is able to decompress 640 x 480 JPEG encoded images in less than one second. The display processor retrieves decompressed images from memory, converts them to the format needed for display, and produces timing and control signals to drive various displays. Popular formats such as NTSC, PAL, VGA, and SVGA are supported. This chip set is compatible with 16 or 32 bit microprocessors operating at 6 MHz or higher clock speed, and utilizes 16 Mbyte video RAM (VRAM). The chip-set is programmable; that is, the compression algorithms can be modified according to the application.

The Compact Disc-Interactive (CD-I) system is a CD-based interactive audio/video special-purpose computer. It is based on Motorola's 680X0 processor, and the CD-RTOS, an operating system developed for the CD-I environment. Its original design was capable of providing full-motion video only on part of the screen. However, full-screen, full-motion video has become available with an add-in module [Mee 92]. This module reproduces video from an encoded bitstream with a frame rate of 24 Hz, 25 Hz, or 30 Hz. The frame rate converter transforms this video to 50 Hz or 60 Hz. Finally, the YUV representation of the encoded bitstream is converted into RGB for display. It includes a number of multimedia capabilities, such as a CD file manager for audio, a user communication manager of the video system, and a motion picture file manager. CD-I players currently exist that can be connected to standard TV sets to play back movies that are recorded on CDs. A full-length feature movie usually requires two CDs.

Recently, almost every computer and workstation manufacturer, including Apple, IBM, SUN, and Silicon Graphics, have added full-motion video capabilities. However, no industry-wide standards have been established yet. Recall that the MPEG-1 video compression standard addresses video compression for multimedia stations, but not full compatibility between various multimedia products. As computers and communication equipment move closer towards "grand unification," the line between multimedia systems and videoconferencing and videophone products is getting more and more blurred; that is, most multimedia systems can now be interfaced with LAN (Local Area Networks), WAN (Wide Area Networks), ISDN, or ATM (Asynchronous Transfer Mode) networks for interactive desktop videoconferencing. At present several companies are introducing plug-in cards and software for videoconferencing over LAN, WAN, and ATM networks. A nice feature of these products is that they provide multiplatform support, so that people using different workstations and personal computers may still share a whiteboard, text tools, and full-motion video.

With the newer generation of higher-speed CD-ROMs and faster processors entering the market every day, multimedia remains an active and exciting field. An emerging technology in interactive multimedia is effective human/machine interaction. Stereo vision, talker verification, speech synthesis, tactile interaction, and integration of multiple sensor modalities are active research areas to expand the present capabilities for human/machine interaction.

25.3 Digital Television

TV is arguably the most commonly used image communication system in the world today. However, present TV transmission standards are based on technology that is more than 40 years old. As a result, there has been widespread interest in the consumer electronics industry to develop more advanced TV systems that benefit from recent technological advances. The major advanced TV research programs in the world are tabulated in Table 25.2.

Table 25.2: Major programs for HDTV research.

Japan	NHK
Europe	EUREKA 95
U.S.A.	Grand Alliance
	(AT&T, General Instrument, Mass. Inst. Tech., Philips N. A. David Sarnoff Res. Cen., Thomson Cons. El., Zenith)

Development efforts pioneered by the Japanese in late the 70s and early 80s resulted in hybrid advanced TV systems, with digital processing at the transmitter and receiver but using analog transmission. Later, in the early 90s, studies in the U.S. proved the feasibility of the all-digital TV approach, which was believed to be unrealistic not too long ago. All-digital TV will not only offer better image quality, easy conversion between multiple standards, and more channels within the same bandwidth (thanks to advances in digital data compression technology), but more important, it will unite computers with TV sets and telecommunication services with cable TV services in a revolutionary fashion. In the following, we provide an overview of the developments around the world that led to the advent of all-digital TV.

25.3.1 Digital Studio Standards

Although the present TV transmission standards are analog (see Chapter 1), digital TV signals find routine use in TV studios for image processing and digital storage. Digital techniques are commonly used for such tasks as program editing, generating special effects, and standards conversion. Digital storage is preferred at the production level, because consumer-quality video storage devices, such as VHS recorders, introduce degradations that are objectionable in the production studio environment.

In the following, we first summarize the existing analog TV standards. The corresponding CCIR 601 standards for digitization of the respective signals (also known as digital studio standards) are listed in Table 25.3.

. **NTSC** (National Television Systems Committee) - Accepted for B&W in 1941, and extended to color in 1954
Used in the USA, Canada, Japan, and Latin America
2:1 interlace, 4:3 aspect ratio
525 lines/frame, 29.97 frames/sec (262.5 lines/field, 59.94 fields/sec)
Perceptually 340 lines/frame, 420 resolvable pels/line (Kell factor)
Analog transmission over 6 MHz channel
There are 68 channels assigned in the US: 54-88 MHz (ch 2 to 6), 174-216 MHz (ch 7 to 13), 470-806 MHz (ch14 to 69). However, less than 20 channels are used in a locality to prevent interference.

Table 25.3: CCIR 601 standards.

Format	Sampling Frequency (Y/U,V)	Field rate	Interlace/ Aspect	Active lines/frame	Active pixels/frame
NTSC	13.5/6.75	60	2:1/4:3	488	720
PAL/ SECAM	13.5/6.75	50	2:1/4:3	576	720

- **PAL** (Phase Alternation Line) - Accepted in 1967 for color broadcast
  Used in most of Europe, and Australia
  625 lines/frame, 2:1 interlace, 50 fields/sec, 4:3 aspect ratio
  Analog transmission over 8 MHz channel.

- **SECAM** (Systeme Electronique Color Avec Memoire) - Accepted in 1967 for color broadcast, used in France, Russia, and Eastern Europe
  625 lines/frame, 2:1 interlace, 50 fields/sec, 4:3 aspect ratio
  Analog transmission over 8 MHz channel.

Note that these standards are incompatible with each other, and conversion from one to another requires digital techniques.

25.3.2 Hybrid Advanced TV Systems

There were several proposals, before the advent of all-digital TV, that offered various degrees of improvements on the quality of present TV systems. We refer to them as "hybrid advanced TV (ATV) systems." ATV systems generally feature better spatial and/or temporal resolution, wider screens, improved color rendition, and CD-quality stereo sound. They fall under two broad categories: compatible ATV systems and incompatible ATV systems.

1) Compatible system proposals may be summarized as follows:

- IDTV (improved definition TV) systems which use digital processing at the receiver for picture quality improvement using existing analog TV transmission standards. Picture quality may be improved through: i) spatial and temporal resolution enhancement by nonlinear interpolation to recover super-Nyquist frequencies, including motion-compensated deinterlacing and frame rate conversion, ii) luminance-chrominance separation to eliminate cross-luminance and cross-chrominance artifacts due to imperfect Y/C separation in conventional NTSC demodulators, and iii) ghost elimination to reduce time-delayed, attenuated, and distorted versions of the TV signal.

- EQTV (extended-quality TV) systems which require transmission of an augmentation signal. They include systems that feature 16:9 aspect ratio by