Multimedia Systems

Edward Cheung
email: icec@polyu.edu.hk
28 July, 2003.

Contents

- Multimedia Systems Design
  - Analogue representation vs digital representation
  - Reconstruction
  - File formats
  - Compression
- Multimedia Elements
  - Images
  - Voice commands / synthesis
  - Audio / Video messages
  - GIS
  - Holography
  - Fractal

Contents (cont.)

- Multimedia Applications
  - Image processing and recognition
  - Electronic messaging
  - Video conferencing
- Multimedia Architecture
  - High resolution graphics display
  - Network architecture
    - Data transfer, data duplication and data replication
- Multimedia data interface
  - File formats (e.g. RTF, TIFF, MIDI, JPEG, MPEG, AVI)
  - Compression of data

Contents (cont.)

- Data Compression and Decompression
  - Lossless vs lossy compression
  - 1-D compression
  - 2-D compression
  - JPEG standard
  - MPEG coding methodolgy

What is Multimedia?

- Human behaviour – reacts according to external stimuli
- Five senses of human – vision, hearing, touch, taste and smell
- A medium refers to different type of data representation such as text, images, graphics, speech, audio, video.
- A user manipulates multimedia data with multimedia systems.
- Multimedia communication system enables the integration of more than one of the media for the purpose of transmission, storage, access and content creation.
  - Interestingly, with this definition, reading newspaper is a multimedia experience since it integrates text and halftone images.
  - TV is another example that integrates video and audio signals.

The Need for Multimodal Communication

- Human communication is multimodal in nature.
  - Human speech is bimodal in nature
  - Others; hand gesture, facial expression
- McGurk effect
  - When humans are presented with conflicting audio and video stimuli, the perceived sound may not exist in either modality due to the dominance of visual sense.
  - Examples:-

Audio	Visual	Perception
ba	ga	da
pa	ga	ta

  - Reverse McGurk occurs when the sound produced by an object influenced its visual interpretation.
    - Visit the epsych site of the Mississippi State University
    - http://epsych.msstate.edu/descriptive/crossModal/mcgurk/mcgurk_desc.html

Analog Systems vs Digital Systems

- Naturally audio and video signals are analog continuous signals; they vary continuously in time as the amplitude, pressure, frequency or phase of the speech, audio, or video signal varies.
- In general, an analog multimedia system is relatively simple with less building blocks and fast response when comparing to a digital system.
- An analog system is more sensitive to noise and less robust than its digital counterpart.
- An analog media is more difficult in editing, indexing, search and transmit among systems.
- Analog signals must be digitized by codec before computer manipulation.

Digital Representation of Multimedia Contents

- Text or Data
  - Text or data is a block of characters with each character represented by a fixed number of binary digits knows as a codeword; ASCII code, Unicode, etc. The size of text or data files are usually small.
- Graphics
  - Graphics can be constructed by the composition of primitive objects such as lines, circles, polygons, curves and splines.
  - Each object is stored as an equation and contains a number of parameters or attributes such as start and end coordinates; shape; size; fill colour; border colour; shadow; etc.
  - Graphics file take up less space than image files.
  - For data visualization, illustration and modelling applications.

Digital Representation of Multimedia Contents

- Images
  - Continuous-tone images can be digitized into bitmaps
  - Resolution for print is measured in dots per inch (dpi)
  - Resolution for screen is measured in pixels per inch (ppi)
  - A digitized image can be represented as two-dimensional matrix with individual picture elements.
  - Each pixel is represented by a fixed number of bits or N-bit quantization.
  - 24-bit quantization provides 16,777,216 colours per pixel.
  - Dots are digitized in 1-bit quantization

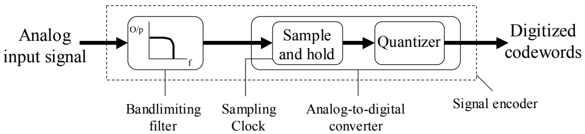
Bitmap Graphics & Vector Graphics

	Bitmap Graphics	Vector Graphics
Display speed	X	
Image quality	X	
Memory usage		X
Ease of editing		X
Display independence		X

- Different method serve different purpose.
- Neither method is better than the other but depending on the application.

Digital Representation of Speech / Audio Signal

- 1-dimensional but need high dynamic range - large file size
  - For telephone lines, the channel bandwidth is 200Hz-3.4kHz for a sampling of 8-bit per channel. The sampling rate is 8kHz for slightly oversampling.
  - Nyquist sampling rate = (3.4-0.2)x2 kHz = 6.8kHz.



An Speech / Audio Encoder

Speech Compression

The digitization process is known as **pulse code modulation** or **PCM.** Typical bit rate is 64kbps - (8 bit x 8kHz) per channel.

This involves sampling the (analog) audio signal/waveform at a minimum rate which is twice that of the maximum frequency component(4kHz) that makes up the signal. – Nyquist Criteria

The sampling rate is limited to conserve the bandwidth of the communications channel.

A compression algorithm is normally used to achieve comparable perceptual quality (as perceived by the ear) to conserve bandwidth.

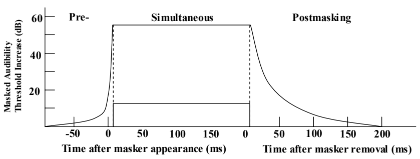
Digital Speech / Audio Representation

- The ear is more sensitive to noise on quite signals than it is on loud signals. In a PCM system the quantization intervals are made non-linear with narrower intervals for low level signals. This is known as companding. In practice, there are 2 different compression-expansion characteristics; A-law for China and Europe and μ-law for Hong Kong and North America.
- The use of companding gives a perceived level of performance with 8-bit PCM data that is comparable with that of 12-bit uniform quantization.
- For music, the audible bandwidth is from 15Hz to 20kHz and is normally sample at 44.1kHz. CD-Digital Audio (CD-DA) standard requires 16-bit ADC.

Data Rate of Digital Audio Applications

Type	Approx. Freq. (Hz)	Sampling Rate (kHz)	Bits per Sample	Uncompressed Bit Rate
Telephone Speech	200-3200	8	8	64 kbps
Wideband Speech	50-10000	16	16	256 kbps
CD Audio	15-20000	44.1	16 x 2	1.41 Mbps

Temporal Masking Properties Of The Human Ear



Digital Video

- Composed of a series of still image frames and produces the illusion of movement by quickly displaying one frame after another.
- The Human Visual System (HVS) accepts anything more than 20 frames per second (fps) as smooth motion.
- The challenges are the massive volume of data involved and the needs to meet real time constraints on retrieval, delivery and display.
- A digitized video comprises both images and audio and need to be synchronize in time.
- The solution is to reduce the image size, use high compression ratios; eliminates spatial and colour similarities of individual images and temporal redundancies between adjacent video frames.

Picture Formats Supported by H.261 & H.263

Parameter	Sub-QCIF	QCIF	CIF	4CIF	16CIF
Number of pixels per line	128	176	352	704	1,408
Number of lines	96	144	288	576	1,152
Uncompressed bit rates (Mb/s)	4.4	9.1	37	146	584

Colour Model

- A system for representing colours is called a colour model.
- Colour models are usually designed to take advantage of a particular type of display device
- The range of colours that can be represented by a colour model is known as a colourspace.
- The RGB colour model is used in computer system because of the CRT colour monitor.
- 8-bit resolution per pixel is used which corresponding to 256 different shades of each primary colour.
- In telecommunication engineering, YUV(YC_BC_R) is used.
- Y is luminance which represent the brightness of the image and C_B C_R are chrominance components.

Color Models in Digital Video

YUV(YC_BC_R) Model

- CCIR 601 standard for digital video
- Y is the luminance (brightness) component
- The Cx known as the colour difference or chrominance components
- Cr+Cb+Cg is a constant, only need to transmit 2 out of 3
- HVS is more sensitive to luminance than chrominance, thus less the chrominance component can be represented with a lower resolution.

Conversion to YUV	Conversion to RGB
$Y = 0.299R + 0.587G + 0.114B$ $C_b = 0.564(B - Y)$ $C_r = 0.713(R - Y)$	$R = Y + 1.402\ C_r$ $G = Y - 0.344\ C_b - 0.714\ C_r$ $B = Y + 1.772\ C_b$

Chroma Subsampling

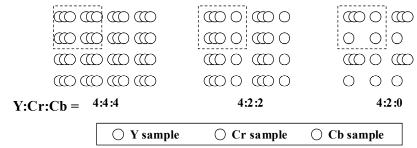
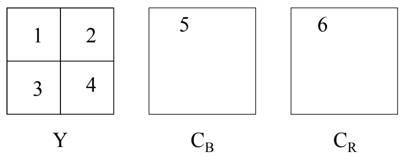


Figure: Chrominance subsampling patterns

- 4:4:4 implies full fidelity
- 4:2:2 is for high quality colour production
- 4:2:0 is for digital television, video conferencing and DVD
- 4:2:0 video requires exactly haf as many samples as RGB or 4:4:4 video

Video Macroblock



GOB Structures

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

GOB for H.261

COB 1	COB 2
COB 3	COB 4
COB 5	COB 6
COB 7	COB 8
COB 9	COB 10
COB 11	COB 12

CIF

COB 1
COB 3
COB 5

QIF

GOB structures in CIF and QCIF for H.261

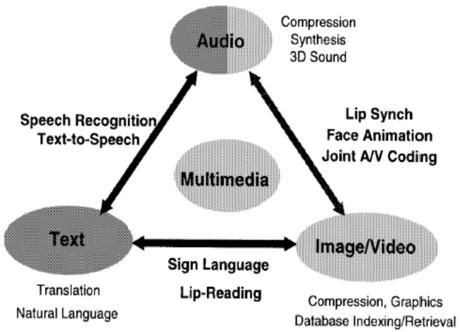
Type of Media and Communication Modes

- Text, data, image and graphics media are known as discrete media
- Audio and video media are known as continuous media
- Most of the time, a multimedia systems can be distinguished by the importance or additional requirements imposed on temporal relationship in particular between different media; synchronization, real-time application.
- In multimedia systems, there are 2 communication modes:-
  - people-to-people
  - people-to-machine
  - An user interface is responsible to integrate various media and allow users to interact with the multimedia signal.

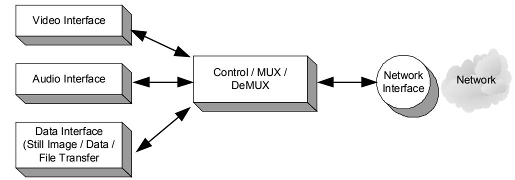
Enabling Technologies for Multimedia Communication

- High speed digital network
  - E.g. optical backbone for network to network service and Ethernet / Digital Subscriber Line (DSL) / Cable broadband services to the public in Hong Kong
- Video compression
  - Algorithms like MPEG 2 and MPEG 4 are available for high and low bit rate applications
- Data security
  - Encryption and authentication technologies are available to protect users against eavesdropping and alternation of data on communication lines.
- Intelligent agents
  - Software programs to perform sophisticated tasks such as seek, manage and retrieve information on the network

Interaction of Media in Communication



A Typical Multimedia Communication System



Analogue Representation Vs Digital Representation

Compression

Compression technique can be easily applied to digitized data.

Since the digitization of an audio and video that much higher bit rates and longer time durations are involved.

In order to reduce the bit rate to a level which various the communication networks can support, so the compression is applied to text, audio, image and video by different types of compression algorithms.

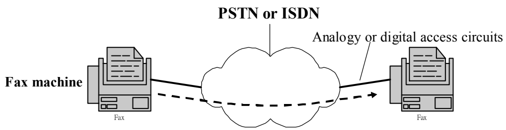
Multimedia Elements - Images

One of the legacy Electronic document images but still very useful communication media is – Facsimile (Fax)

The two fax machines communication with each other to establish operational parameters after which the sending machine starts to **scan** and **digitize** each page of the document in turn.

The process involved scanning and compression.

Resolution of FAX is typically 200 dpi or 8 pels per mm.

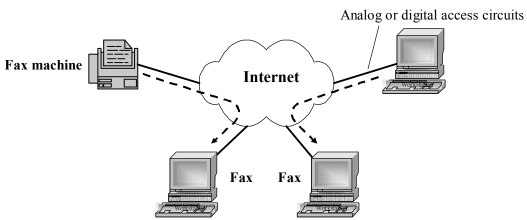


Application - PC Fax

PC fax is use a PC instead PC instead of a normal fax machine to send an electronic version of a document that is stored directly within the PC's memory.

As with telephony, this requires a telephone interface card and associated software.

Email to fax gateway (free international fax):- http://www.tpc.int/



Multimedia Application - Speech

Interpersonal communications involving speech – telephony – a real time application. People using telephones that connected either to a PBX, PSTN, ISDN or cellular phone network. A wide installation base and highly demanded service area.

Using a multimedia PC equipped with a microphone and speakers, the user can take part in telephone calls through the PC.

- This is known as **computer telephony integration (CTI)**
- This requires a modem / telephone interface and associated software for connecting to switching network.

For the public or private packet networks that supports voice services, they are known as **voice over IP (VoIP) service; teleconferencing calls** is one of the applications.

Multimedia Application

Teleconferencing

Tele-conferencing calls or an **audio-conferencing calls** involve multiple interconnected telephones /PCs.

Each person can hear and talk to all of others involved in the calls.

It required a central unit known as an **audio bridge** which provide the support to set up a conferencing calls automatically.

Voice over IP (VoIP)

Internet was designed to support computer-to-computer communications – delay is not a concern. The industry is developing the application of packet network to carry voice / speech where delay and equipment cost is a major problem.

Multimedia Application

PC-to-PC

- In the internet, the Voice over IP (VoIP) operate in a packet mode, both PCs must have the necessary hardware and software to convert the speech signal from the microphone into packets on input and back again prior to output to the speakers.

PC-to-Telephone

- When a PC connected to the internet need to make a call to a telephone that is connected to a PSTN/ISDN, both operate in a circuit mode, an internetworking unit known as a telephony gateway must be used. But who is going to pay for the gateway?

Audio / Video / Multimedia Messages

Internet-based electronic mail – **email** is text based. But the information content of email message does not limited to text. Email can containing other media types such as audio, image and video but the mail server need to handle multimedia messages.

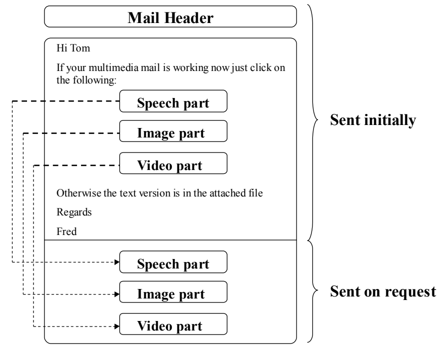
Voice-mail

- With internet-based voice-mail, a voice-mail server must associated with each network.
- The user enters a voice message addressed to the intended recipient and the local voice-mail server then relays this to the server associated with the intended recipient's network.

Multimedia-mail

- With multimedia-mail, the textual information is annotated with a digitized image, a speech message, or a video message.

Multimedia Electronic Mail Structure



Multimedia Elements

- For example, speech-and-video, the annotations can be sent either directly to the mailbox of the interned recipient together with the original textual message --
  - and hence stored and played out in the normal way
  - or they may have to be requested specifically by the recipient when the textual message is being read.
- In this way, the recipient can always receive the basic text-only message but the multimedia annotations can be received only if the terminal being used by the recipient supports voice and/or video.

Multimedia Applications

GIS

Geographic information systems (GIS) refer to those computer systems that are being used in scientific investigations, analysis, management, development and planning of geographically data and resources.

Multimedia application in GIS allows the user to visualize and interact with scientific data more effectively.

GIS is similar to CAD except a continuous and large data set, heterogeneous, GIS geometry is primarily obtained by remote sensing or surveying, lines of fractal nature like the coastlines are common.

GIS

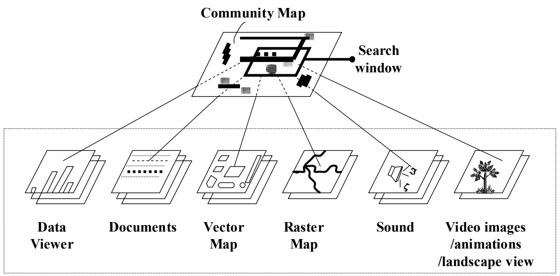
Different kinds of data in map form can be entered into a GIS.

A GIS can also convert existing digital information, which may not yet be in a map form, into forms that it can process and use. For example, digital satellite images can be analyzed to produce a map like layer of digital information about vegetative covers

Census or hydrologic data can be converted to map-like form and presented as layers of thematic information in a GIS.

Some GIS system includes simulation and modelling functions that demands sophisticated display.

Application of Multimedia in GIS



GIS Software and Reference on Web

GeoVISTA Center of The Pennsylvania State University
http://www.geovista.psu.edu/index.jsp

National Spatial Data Infrastructure (NSDI)
http://nsdi.usgs.gov/pages/what_is_gis.html

GIS software at ESRI
http://www.esri.com
http://www.mapinfo.com

Other CAD base GIS software
http://www.autodesk.com
http://www.integraph.com

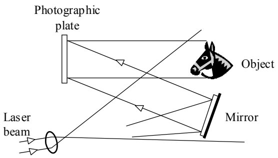
Multimedia Elements - Holography

In 1948, Dennis Gabor recognized that when a suitable coherent reference wave is present simultaneously with the light diffracted by or scattered from an object, then information about both the amplitude and phase of the diffracted or scattered waves can be recorded, in spite of the fact that recording media respond only to light intensity. He demonstrated that from a recorded interference pattern which he called a hologram, meaning “total recording”, an image of the original object can ultimately be obtained. In 1971, Gabor received the Nobel prize in physics for his work.

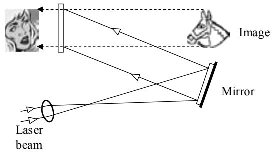
http://www.holography.ru/maineng.htm

Holography

Recording a hologram
The photographic plate records the interference pattern produced by the light waves scattered from the object and a reference laser beam reflected to it by the mirror.



Reading a hologram
The hologram is illuminated with the reference laser. Light diffracted by the hologram appears to come from the original object.



Holography

We feel 3D by estimating the sizes of viewing objects and considering the shape and direction of shadows from these objects, we can create in our mind general representation about volumetric properties of the scene, represented in a photo.

Holography is the only visual recording and playback process that can record our three-dimensional world on a two-dimensional recording medium and playback the original object or scene to the unaided eyes as a three dimensional image.

Hologram image demonstrates complete parallax and depth-of-field and floats in space either behind, in front of, or straddling the recording medium.

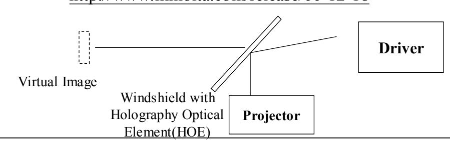
3D vs 2D Images

Holography differs from usual photography because a hologram not only record the amplitude of light intensity on photosensitive material, but also the phase of light waves reflection. The reflected light as scattered by the object carry complete information of the object view.

This difference is why photographs are two dimensional (2-D) images while holograms are three dimensional (3-D) images. We can do this with a stereoscope (for pictures) or with polarized glasses (for movies). The shortcoming of these stereoscopic images is that when we move our head from side to side or up and down, we still only see the same two view points. The image therefore doesn't quite appear to be three dimensional.

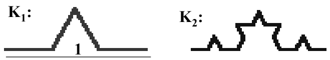
Application of Holographic Display

- Head up displays (HUD)
- One of the most common application is to project an image such as speedometer reading of a vehicle to the windshield. This information appears to be floating in front of the vehicle. The drivers can keep their eyes on the road.
  - Aircraft – projected high at optical infinity
  - Automobile – projected low in front of the vehicle
  - Example of wearable HUD
    - http://www.minolta.com/release/00-12-18



Multimedia Elements - Fractals

What are Fractals?
Fractals are various forms of self-similar curves. They are developed since 1970 by Benoit Mandelbrot of IBM Research.



Koch Curve from Swedish mathematician Helge von Koch in 1904

To form K_{n+1} from K_n:

Subdivide each segment of Kn into 3 equal parts and replace the middle part with a bump in the shape of an equilateral triangle. A Koch snowflake can be formed by joining 3 Koch curve to form a triangle. As n increase to infinity, the area remains bounded.

Image Processing and Recognition

Some curves are exactly self-similar – if a region is enlarged, the enlargement looks exactly like the original (except rotation or shift). Nature provides examples that mimic statistically self-similar – the wiggles and irregularities in the curve are on average the same no matter how many times the picture is enlarged or reduced.

e.g. branches of a tree, cracks in a pavement, coastline, etc.

Application on fractals can be found in the area of bcomputer graphics, compression and recognition algorithms, etc.

http://math.rice.edu/~lanius/frac/

Video Conferencing

Two-party Telephone Call

An application that uses speech and video integrated together is **video telephony** which is supported by all the network types.

- The terminals/PCs incorporate a video camera in addition to the microphone and speaker used for telephony.

The network must provide two-way communication channel of sufficient bandwidth to support the integrated speech-and-video generated by each terminal/PC.

Video Conferencing

Videoconferencing using an MCU

Multiparty videoconferencing call is widely used in large corporations involving multiple geographically distributed sites in order to minimize travel between the various locations.

- The integration of video with speech means that the bandwidth of the access circuits required to support this type of service is higher than that required for speech only.

Moreover, as with telephony, a call may involve not just two persons - and hence terminals/PCs - but several people each located in their own office.

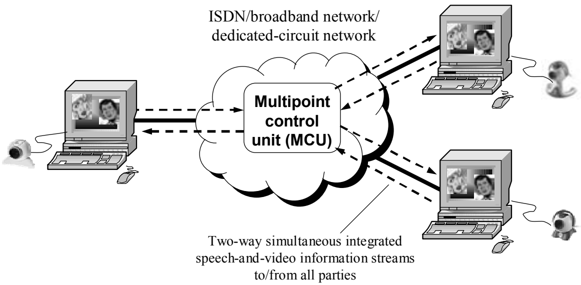
Video Conferencing

Large corporations of this type have an enterprise-wide network to link the sites together and, in order to support videoconferencing, there is a central unit called a **multipoint control unit (MCU)** - or sometimes a **videoconferencing server** - associated with this network.

Network Options

- **IP Based**
- **ISDN Based**
- **Both**

Videoconferencing using an MCU



Multiparty Video Conferencing

In principle, a separate window appears on the screen of each participant's PC/workstation is used to display the video image of all the participants.

This would require multiple integrated speech-and-video communication channels, one for each participant, being sent to each of the other participants.

Normally, this would require more bandwidth than is available.

The integrated speech-and-video information stream from each participant is sent to the MCU which then selects just a single information stream to send to each participant.

Sometimes it is known as always presence.

Other feature:- Chair control / Admission Control

Video Conferencing

Videoconferencing on IP Network

The Internet does not support multicasting but multicasting is supported on LAN.

Multicasting means that all transmissions from any of the PCs/workstations belonging to a predefined **multicast group** are received by all the other members of the group.

Thus with networks that support multicasting, it is possible to hold a conferencing session without an MCU.

Only a limited number of participants are involved owing to the high load it places on the network.

Neither continuous presence nor admission control is available

Common Video Conferencing Standards

Network	PSTN	N-ISDN	LAN
ITU-T Rec.	H.324	H.320	H.323
Video Codec	H.261*, H.263*	H.261*, H.263	H.261*, H.263
Data Conference	T.120	T.120	T.120
System Control	H.245	H.230 H.242	H.245
Multiplex	H.223	H.221	H.225
Security		H.233, H.234	H.235
MCU		H.231, H.243	H.323
Interface	V.34	I.400	TCP/IP
Setup Signaling	SS7	Q.931	Q.931, H.225

ISDN

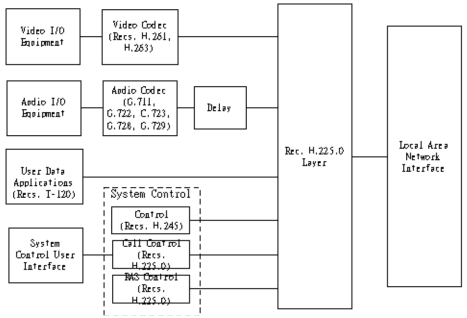
BRI – 2B + 1D Channel

PRI – 23B Channel

B channel Bandwidth 64kbps

D channel Bandwidth 16kbps

H.323 Video Conference Terminal



Examples of Video Coding Standards

Standards organization	Video-coding standard	Typical range of bit rates	Typical applications
ITU-T	H.261	p x 64 Kb/s, p=1, 2,…, 30	ISDN video phone
ISO	IS 11172-2 MPEG-1 Video	1.2 Mb/s	CD-ROM
ISO	IS 13818-2 MPEG-2 Video	4-80 Mb/s	SDTV, HDTV
ITU-T	H.263	64 Kb/s or below	PSTN video phone
ISO	CD 1496-2 MPEG-4 Video	24-1024 Kb/s	Interactive audio/video
ITU-T	H.263 Version 2	< 64 Kb/s	PSTN video phone
ITU-T	H.26L	< 64 Kb/s	Network-friendly packet-based video

High resolution graphics display

- Resolution as dot matrix
  - resolution is given as the size of the dot matrix in pixels per inch
- Resolution as sharpness
  - means the sharpness of an image as measured by the number of dots in a given space normally given as *dpi* (dots per inch)
  - let's consider two cathode ray tube (CRT) monitors both having a resolution of 1,024 by 768,
  - but one of which is listed as 17 " (15.5 ") and the other 19" on the diagonal; the 17" monitor may be displaying 85 dpi for horizontal resolution while the 19" monitor may be displaying 72 dpi. If we are defining the resolution based on dot matrix, the two monitors are the same. But if we define the resolution as sharpness, they are different.

High Resolution Graphics Display

- The dot pitch is a measure of image clarity. The smaller the dot pitch, the crisper the image.
- The dot pitch is the distance between like colour dots usually measured in mm.
- Common dot pitch on monitors are between 0.23mm and 0.27mm (that is between 110 pixels / inch to 94 pixels / inch)
- Normally, a CRT monitor has a better colour fidelity than that of a LCD panel.
- For LCD panel, e.g. Sharp LL-T2020B, 20.1 " panel with a display size of 408mm x 306mm, maximum display resolution 1600x1200 means a dot pitch of 0.255mm which is not as good as CRT monitor

High Resolution Graphics Data

Common Resolution of Graphics

Standard	Horizontal Resolution	Vertical Resolution	Total Dots	Raw File Size
QXGA	2,048	1,536	3,145,728	9.44MB
UXGA	1,600	1,200	1,920,000	5.76MB
SXGA	1,280	1,024	1,310,720	3.94MB
XGA	1,024	768	786,432	2.36MB
SVGA	800	600	480,000	1.44MB
VGA	640	480	307,200	0.93MB

QXGA=Quantum Extended Graphics Array

UXGA=Ultra Extended Graphics Array

- **Need high bandwidth network and compression.**

Multimedia Network Architecture

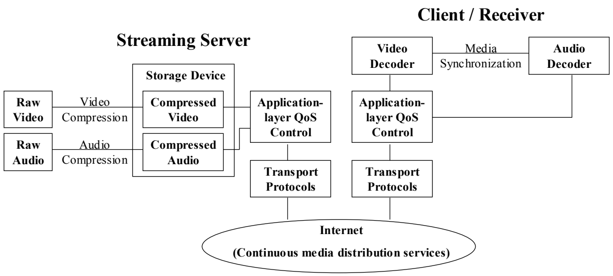
- Last Mile Solution
  - Broadband Network
    - No internationally agreed definition of broadband. In general, broadband refers to internet access service with transmission speed from kbps to Mbps. - OFTA
  - Cable modems, ATM, Ethernet, ADSL and other types of DSL are technologies commonly used to provide broadband services.
- Backbone
  - Optical Network forms the backbone of multimedia delivery because of its high bandwidth
- Wireless
  - High compression, jitter
  - 2.5G, 3G, WLAN, etc.

Multimedia Network Architecture

- Unicasting Network
  - Web TV
- Multicasting Network
  - VoD
- Broadcast Network
  - Current development in Digital Terrestrial Televsion
    - May be DVB-T or from Mainland China
    - http://www.dvb.org
    - http://www.info.gov.hk/info/broadcs.htm
    - http://www.imaschina.com/shownews.asp?newsid=1259
- Packet Network – Ethernet / Frame Relay
- Switching Network – ATM cell
- Soft real time systems

Multimedia Network Architecture

Video-streaming



Multimedia Network Architecture

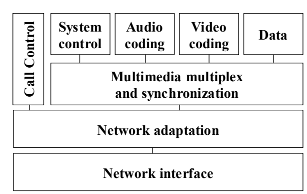
- Quality Parameters
  - Bandwidth -- hundreds of Mbps
  - Latency (maximum frame/cell delay)
    - important for Interactive multimedia delivery, telephone, etc.
  - Frame/Cell Loss Ratio
  - Jitter -- *Playback Jitter* is the standard deviation between frame playback. It is a measure of smoothness.
  - Synchronization
    - Audio signal usually has a higher priority
- Video Server Algorithms
  - Video prioritization algorithms set priorities to minimize playback jitter.
  - Retransmission is not a good method for error correction in multimedia data transmission
    - a late frame in audio/video is a lost frame
  - Admission Control is important for bandwidth management

Multimedia Data Replication

- Duplication vs Replication
- Duplication is to make a copy of the data / contents.
  - CD Duplication
    - Using CD recordable disc and make an exact copy of the source disk
- Replication is the process of creating and managing duplicate versions of a database. Replication not only copies a database but also *synchronizes a set of replicas* so that changes made to one replica are reflected in all the others.
  - CD Replication
    - Using injection molding machines to create the disk
- A mechanism is required for network data replication.

Multimedia Network Standards

- The general protocol stack of H-series audiovisual communication terminal



Multimedia Data Interface – Text and Graphics Files

The Rich Text Format (RTF) Specification provides a method of transferring encoding formatted text and graphics between different output devices, OS and applications.

RTF uses the American National Standards Institute (ANSI), PC-8, Macintosh, or IBM PC character set to control the representation and formatting of a document, both on the screen and in print.

RTF files begins with control word "\rtfN"; where N is the major version number. An entire RTF file is consider a group and must be enclosed in braces. Current version is 1.7, hence the file will begin with

{\rtf1\ansi\ansicpg1252\....} …..

File Format – Image file

- **TIFF**

Tag image file format (TIFF) is typically used to transfer graphics from one computer system to another.

TIFF is an excellent method of transporting images between file systems and software packages.

It is supported by most graphics packages and has a high resolutions that is typically used when scanning images and colours of up to 48 bits.

It can use different compression methods and different file formats, depending on the type of data stored.

File Format - Image

TIFF file structure:

TIFF files have a three-level hierarchy

1. A file header:
   - It contains 8 bytes
2. One or more IFDs (image file directories):
   - These contain codes and their data (or pointers to the data)
3. Data:
   - The data information

File Format – Audio

- WAVE
- Native sound format by Microsoft
  - Based on Interchange File Format (IFF) from Electronic Arts for Amiga
  - Basis of Apple's AIFF (Audio IFF)
  - One of the RIFF file
  - The other is AVI
  - File begins with RIFF
  - But we can use compression in WAVE files

Selected WAVE Format Codes

Code	Description
0	Unknown/illegal
1	PCM
2	Microsoft ADPCM
6	ITU G.711 A-Law
7	ITU G.711 μ-Law
17	IMA ADPCM
20	ITU G.723 ADPCM
49	GSM 6.10
64	ITU G.721 ADPCM
80	MPEG
65535	Experimental

File Format - Audio

- **MIDI**

Musical Instrument Digital Interface (MIDI) provides a relatively simple way for a single keyboard to control a variety of instruments

MIDI is a simple digital protocol and was rapidly incorporated into computer music applications.

The MIDI is used as the foundation for Internet music playback and other new kinds of music publishing.

A MIDI file is a sequence of **chunks**. These chunk have the same general format as the chunks used by AIFF, IFF and WAVE; each chunk has a 4-character type, a 4-byte length code (in MSB format), and some data

MIDI

- Advantages
  - Store a list of notes to be played is more compact than store the entire song
  - The music can be modified conveniently.
  - Isolate a particular instrument in a symphony
  - Great for games – as background music
  - Music can be changed as the game progresses
  - General MIDI specified 175 instruments

File Format - Audio

Three types of MIDI files:

Type zero files contain only one track:

- There are clearly the easiest to play, and preferred when you want to make sure your MIDI files are easy for other people to play.

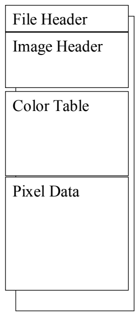
Type one files contain multiple tracks that are played simultaneously:

- A program to play type one files must somehow “flatten” the data into a single event stream before playing.

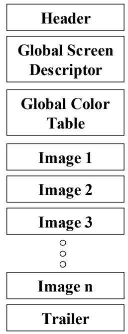
Type two file contain multiple tracks but without assuming any timing relation among the tracks:

- Type two files are relatively uncommon

Image File Structures



Bitmap File Structure



GIF File Structure

File Format – Image file

- **JPEG**

Joint Photographic Expert Group (JPEG) is an excellent compression technique which provide lossy compression.

It is a compression technique for gray-scale or color image and uses a combination of discrete cosine transform, quantization, run-length and Huffman coding.

JPEG operates differently in that it stores changes in colour. As the eye is very sensitive to brightness changes and less on colour changes, then if these changes are similar to the original then the eye perceives the recovered image as very similar to the original.

File Format – Image file

JPEG files format:

The JPEG file normally complies with JFIF (JPEG file interchange format) which is a defined standard file format for storing a gray scale or YC_bC_r colour image.

Data within the JFIF contains segments separated by a 2-byte marker.

This marker has a binary value of 1111 1111 (FF_h) followed by a defined marked field.

If a 1111 1111 (FF_h) bit field occurs anywhere within the file (and it isn't a marker), the stuffed 0000 0000 (00_h) byte is inserted after it so that it cannot be read as a false markers.

File Format - Multimedia

- **MPEG**

ISO established the Moving Picture Expert Group with the mission to develop standards for the coded representation of moving picture and associated audio information on digital storage media.

MPEG standardization are best known for their video compression, they also support high-quality audio compression.

There are currently **Five MPEG Standards** in various stages of completion.

Each consists of numerous parts addressing aspects of video compression, storage, and delivery.

File Format - Multimedia

MPEG-1:

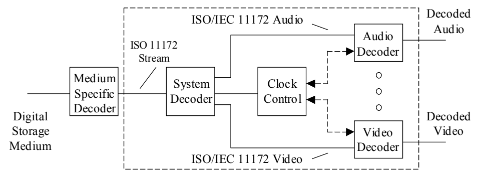
ISO 11172, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbits/s. (such as CD-ROM)

A file contain a video, an audio stream, or a system stream stream which interleaves some combination of audio and video stream.

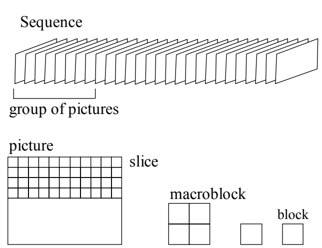
The MPEG-1 bit-stream syntax allows for picture sizes of up to 4095 x 4095 pixels. Many of the applications using MPEG-1 compressed video have been optimized the SIF (source intermediate format) video format.

As in MPEG-1, input pictures are coded in the YC_bC_r colour space. This is referred to as the 4:2:0 subsampling format.

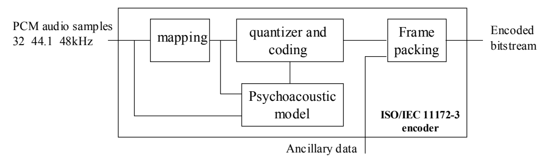
MPEG1 (ISO/IEC 11172) Decoder



Temporal Picture Structure (MPEG-1)

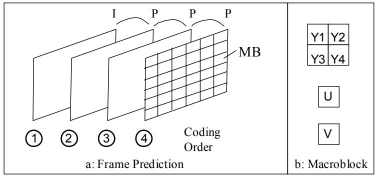


MPEG1 (ISO/IEC 11172) Audio Encoder



MPEG-1 Coding Algorithm

Illustration of I-picture (I) and P-picture (P) in a video sequence



File Format - Multimedia

MPEG-2:

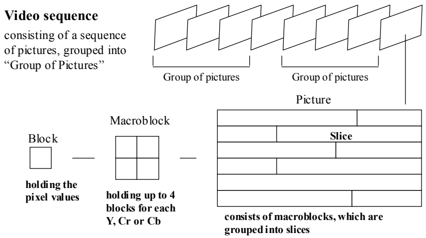
ISO 13818, "Generic Coding of Moving Pictures and Associated Audio"

The maximum coded bit rate to 10Mbits/s.

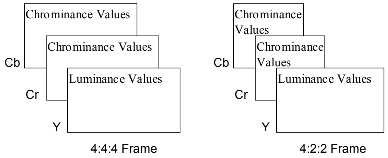
It could successfully support higher bit rates, e.g 80-100Mbits/s for HDTV applications.

MPEG-2 supports the 4:2:2 colour subsampling format.

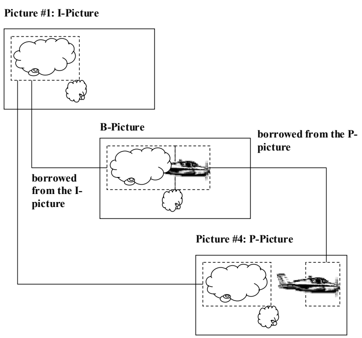
Basic Objects in MPEG-2



Matrixes for 4:4:4 and 4:2:2 Sampled Frame



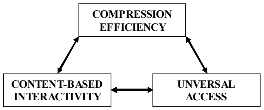
Encoding of the B Frame



File Format - Multimedia

MPEG-4:

MPEG-4 is designed for very low bit rate system. This would allow for low-quality video over internet connection.



Functionalities in MPEG-4 Visual standard

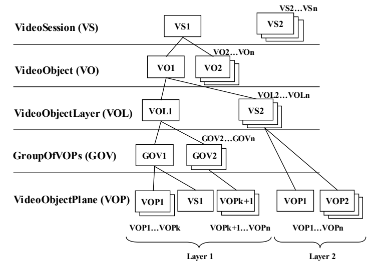
MPEG-4

Bit rate targeted for the MPEG-4 video standard:

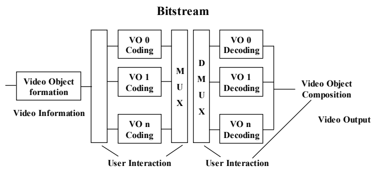
5 - 64 Kbit/s	**for mobile or PSTN video**
64 Kbit/s – 2 Mbit/s	**for TV/film applications**

The object of MPEG-4 was to standardize algorithms for audiovisual coding in multimedia application, allowing for interactivity, high compression, scalability of video and audio content, and support for natural and synthetic audio and video content.

MPEG-4 Video Bit Stream Logical Structure.



MPEG-4 Video Coding And Decoding



File Format - Multimedia

MPEG-7:

Multimedia Content Description Interface and referenced to as MPEG-7, that will extend the limited search capabilities of today to include more information types.

MPEG-7 Initiatives

- **to take care of the situation that more and more audiovisual information is available in digital form.**

  (Before using them, to locate them is important, or even more important.)

- **Current search engines, mainly for searching textual information**

File Format - Multimedia

- **For Multimedia Databases –**
  - **to allow searching for pictures using characteristics**

    such as colour
    texture
    shapes of objects
  - **to allow searching for audio (or say a song) characteristics of personal features using certain specification of a program**

There are many applications that includes digital libraries (image catalog, musical dictionary), multimedia directory services (yellow pages), broadcast media selection (radio and TV channel) and multimedia edition (personalized electronic news services).

File Format - Multimedia

AVI stands for Audio Video Interleave. It is a special case of the Resource Interchange File Format (RIFF) which create by Microsoft for audio/video data on PC and it becomes the de facto standard for uncompressed data. Other RIFF file includes WAV for audio and others.

AVI files contains the FOURCC (Four Character Codes) codec ID for the video compressor in the video stream header.

http://www.jmcgwan.com/avi.html

Compression Data

Compression data is becoming an important subject as increasingly digital information is required to be stored and transmitted.

Data Compression techniques exploit inherent redundancy and irrelevancy by transforming a data file into a smaller file from which the original image file can later be reconstructed, exactly or approximately.

Some data compression algorithms can be classified as being either lossless or lossy.

Video and sound images are normally compressed with a lossy compression whereas computer-type data has a lossless compression.

Bandwidth Requirement

- E.g. bandwidth of a typical video file with 24bit true colour and a resolution of 320 (H) x240 (V) pixels
  - Data per frame = 320 x 240 x 24 / 8 = 230kB
- For desirable motion, we need 15frame per second.
- Therefore uncompressed data per second

  = 230k x 15=3.45MB
  - If the video programme is 60 minutes in length, a storage capacity for the video programme requires

    = 3.45M x 60 x 60 = 12.42GB
  - If we send it over the network, the bandwidth requirement is

    = 3.45M x 8 = 27.6Mbps

Lossless Compression and Lossy Compression

There are two broad categories of data compression:

Lossless Compression:

- Such as planar RGB, tiff and gzip

Lossy Compression:

- Such as mp3, gif and jpeg

Lossless	Lossy
An exact copy of the original data is obtain after decompression	Original information content is lost
Structure data can be compression to 40 – 60% of original size	Any data can be compressed. Sometimes by 90% or more

Lossless Compression

Lossless compression – where the information, once uncompressed, will be identical to the original uncompressed data.

This will obviously be the case with computer-type data, such as data files, computer programs.

Any loss of data will cause the file to be corrupted.

A **lossless compression algorithm** the aim is to reduce the amount of source information to be transmitted.

When the compressed information is decompressed, there is no loss of information.

Lossless compression is **reversible.**

Lossless Compression Methods

An example application of lossless compression is for the transfer over a network of a text file since, in such applications, it is normally imperative that no part of the source information is lost during either the compression or decompression operations.

Lossless compression consists of those techniques guaranteed to generate an exact duplicate of the input data stream after a compress/expand cycle.

The most popular lossless coding techniques are **Entropy**, **Huffman Coding** and **Arithmetic Methods**.

Lossy Compression

Lossy Compression – where the information, once uncompressed, cannot be fully recovered.

Lossy compression normally involves analyzing the data and determining which information has little effect on the resulting compressed data.

Compression of an image might be used to reduce the resolution of the image.

Lossy compression allows much lower data rates and file sizes than lossless compression, so lossy codecs are commonly used for final production of video delivered using CD-ROM or the Internet.

Lossy Compression

The aim of **lossy compression algorithms,** is normally not to reproduce an exact copy of the source information after decompression but rather a version of it which is perceived by the recipient as a true copy.

In general, with such algorithms the higher the level of compression being applied to the source information the more approximate the received version becomes.

Example applications of lossy compression are for the transfer of digitized images and audio and video streams.

In cases, the sensitivity of the human eye or ear is such that any fine details that may be missing from the original source signal after decompression are not detectable.

Perceptual Coding

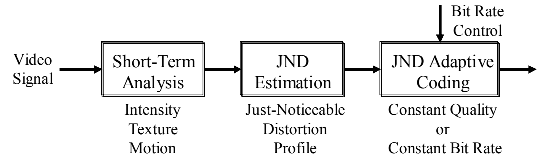
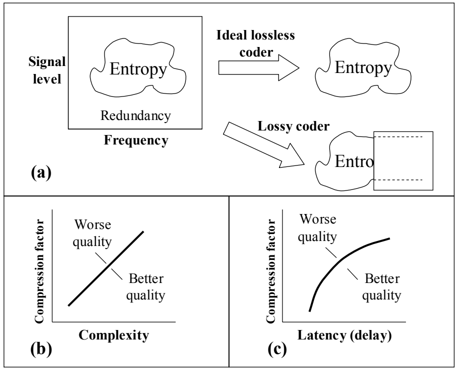


Figure: Block diagram of generic image-coding algorithm

Operation of a Coder



Compression Methods Used by Various Image File Formats

	BMP	GIF	PNG	JPEG
RLE	X			X
LZ		X	X	
Huffman			X	X
DCT				X

Entropy Encoding

Entropy encoding is lossless and independent of the type of information that is being compressed. It is concerned solely with how the information is represented.

Run-length Encoding

Run-length encoding are when the source information comprises long substrings of the same character or binary digit.

Instead of transmitting the source string in the form of independent codewords or bits, it is transmitted in the form of a different set of codewords but also an indication of the number of characters/bits in the substring.

Entropy Encoding

In an application that involves the transmission of long strings of binary bits that comprise a limited number of substrings, each sub-string can be assigned a separate codeword.

For example, if the output of the data was:

000000011111111110000011...

this could be represented as: **0,7 1,10 0,5 1,2** . . . .

To send this in a digital form, the individual decimal digits would be sent in their binary form and, assuming a fixed number of bits per codeword, the number of bits per codeword would be determined by the largest possible substring.

Entropy Encoding

Statistical Encoding

A set of ASCII code-words are often used for the transmission of strings of characters.

Normally, all the codewords in the set comprise a fixed number of binary bits, for example 7 bits in the case of ASCII.

Fro ***L=8*** message, elements are $\{a_1,a_2,\ldots,a_8\}$

Length of codeword required = $\log_2 \boldsymbol{L} = \log_2 8 = 3$

Message	Codeword	Message	Codeword	Message	Codeword	Message	Codeword
a_1	000	a_2	001	a_3	010	a_4	011
a_5	100	a_6	101	a_7	110	a_8	111

Entropy Encoding

Variable-Length Codeword Assignment

Statistical encoding exploits this property by using a set of variable-length codewords with the shortest codewords used to represent the most frequently occurring symbols.

It is always the case that some message possibilities are more likely to be available than others.The major idea is to assign less bits to frequent message (e.g a,e,i,o,u), while to assign long word to messages that appear rarely (e.g j,k,x,z).

Entropy Encoding

<delete>

$$\text{Average number of bits per codeword} = \sum_{i=1}^{n} N_i P_i$$

where n is the number of different symbols in the source stream and P_i is the probability of occurrence of symbol i.

The theoretical minimum average number of bits that are required to transmit a particular source stream is known as the **entropy** of the source and can be computed using a formula attributed to Shannon:

$$\text{Entropy}, H = -\sum_{i=1}^{n} P_i \log_2 P_i$$

Entropy Encoding

Example:

A statistical encoding algorithm is being considered for the transmission of a large number of long text files over a public network. Analysis of the file contents has shown that each file comprises only the six different characters **M, F, Y, N, O**, and **1** each of which occurs with a relative frequency of occurrence of 0.25, 0.25 0.125, 0125, 0,125, and 0.125 Respectively. If the encoding algorithm under consideration uses the following set of codewords:

M = 10, F = 11, Y = 010, N =011, O = 000, 1 = 001

Entropy Encoding

(A) Average number of bits per codeword

$$\sum_{i=1}^{6} N_i P_i = 2(2 \times 0.25) + 4(3 \times 0.125)$$
$$= 2 \times 0.5 + 4 \times 0.375 = 2.5$$

(B) Entropy of source

$$H = -\sum_{i=1}^{6} P_i \log_2 P_i = -(2(0.25\log_2 0.25) + 4(0.125\log_2 0.125))$$
$$= 1 + 1.5 = 2.5$$

(C) The minimum number fixed-length codewords would required a minimum of 3bits (8 combination) to code 6 different characters.

Huffman Coding

Huffman coding the character string to be transmitted is first analyzed and the character types and their relative frequency determined.

- A simple approach to obtain a coding near to its entropy.

Example:

A series of messages is to be transferred between two computers over a internet. The messages comprise just the characters A through H. Analysis has shown that the probability of each character is as follows:

A and B = 0.25, C and D = 0.14 E,F,G and H=0.055

Huffman Coding

A) Use Shannon's formula to derive the minimum average number of bits per character.

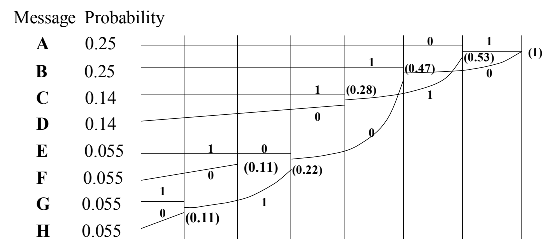
Entropy, $H = -\sum_{i=1}^{n} P_i \log_2 P_i$ bits per coderword

$$\mathbf{H} = -(2(0.25\log_2 0.25) + 2(0.14\log_2 0.14) + 4(0.55\log_2 0.55))$$
$$= 1 + 0.794 + 0.921 = 2.175 \text{ bits per coderword}$$

B) Use Huffman coding to derive a **codeword set**

The characters at the bottom of the list in weight order and the two characters at the bottom of the list are assigned to the (1) and (0) branches.

Huffman Coding



A — 10 (2 bits)		E — 0001 (4 bits)	
B — 01 (2 bits)		F — 0000 (4 bits)	
C — 111 (3 bits)		G — 0011 (4 bits)	
D — 110 (3 bits)		H — 0010 (4 bits)	

Huffman Coding

C) An average number of bit per codeword using Huffman coding is:

2 (2 x 0.25) + 2(3 x 0.14) + 4(4 x 0.55))

= 2.72 bits per coderword

D) Using fixed-length binary codewords:

There are 8 character and hence 3 bits per codeword

Arithmetic Coding

The codewords produced using arithmetic coding always achieve the Shannon value.

Arithmetic coding, however, is more complicated than Huffman coding.

To illustrate how the coding operation takes place, consider the transmission of a message comprising a string of characters with probabilities of:

$$a_1 = 0.3, \quad a_2 = 0.3, \quad a_3 = 0.2, \quad a_4 = 0.1, \quad a_5 = 0.1$$

Arithmetic Coding

Arithmetic coding yields a single codeword for each encoded string of characters.

The first step is to divide the numeric range from **0 to 1** into a number of different characters present in the message to be sent and the size of each segment by the probability of the related character.

Hence the example for our set of five characters may be as shown in following Figure.

O = 0.3, L = 0.3, Y = 0.2, P = 0.1, U = 0.1

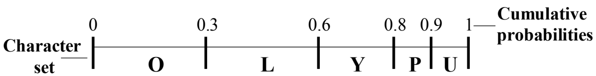


Figure: Example character set and their probabilities

Arithmetic Coding

There are five segments, the width of each segment being determined by the probability of the related character.

Once this has been done, we are ready to start the encoding process. An example is shown in figure.

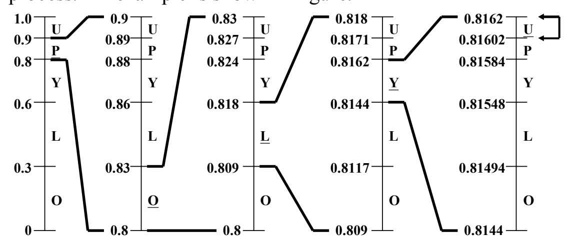


Figure: Arithmetic coding principles: encoding of the string POLYU

Arithmetic Coding

In the about figure, at that point, the segment range of "U" is from 0.81602 to 0.8162 and hence the codeword for the complete string is any number within the range:

0.81602 < codeword > 0.8162

In the static mode, the decoder knows the set of characters that are present in the encoded messages it receives as well as the segment to which each character has been assigned and its related range.

Hence with this as a start point, the decoder can follow the same procedure as that followed by the encoder to determine the character string relating to each received codeword.

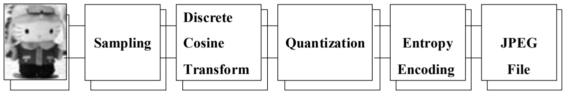
JPEG Standard

- The most widely-adopted standard relating to the compression of digitized pictures has been developed by an international standards body known as the Joint **Photographic Experts Group (JPEG).**
- JPEG also forms the basis of most video compression algorithms.
- JPEG is defined in the international standard IS 10918.
- The standard defines a range of different compression modes, each of which is intended for use in a particular application domain.

Lossy Sequential Mode

- **Lossy Sequential Mode** - also known as the **baseline mode**
- Since it is this which is intended for the compression of both monochromatic and color digitized pictures/images as used in multimedia communication applications.
- There are five main stages associated with this mode:
  - Image/Block Preparation
  - Forward DCT (Discrete Cosine Transform)
  - Quantization
  - Entropy Encoding and
  - Frame Building

JPEG Encoding Overview



JPEG Encoder

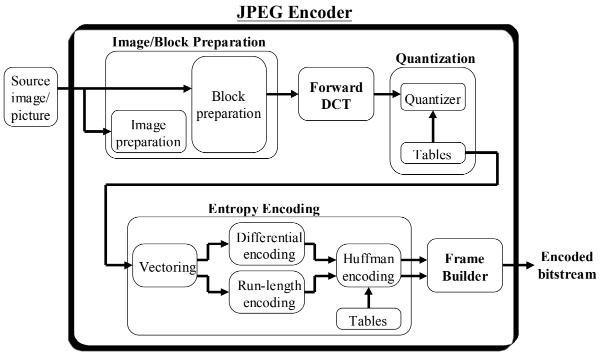


Figure: JPEG encoder schematic

JPEG Decoder

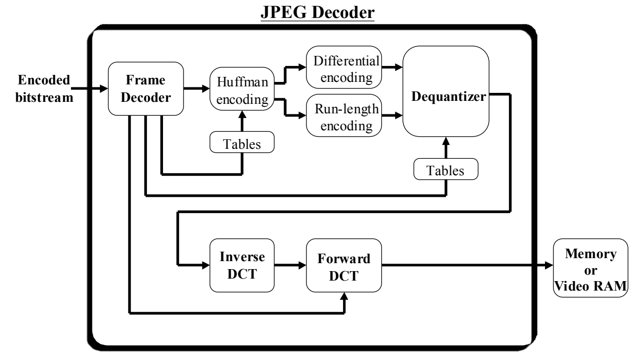


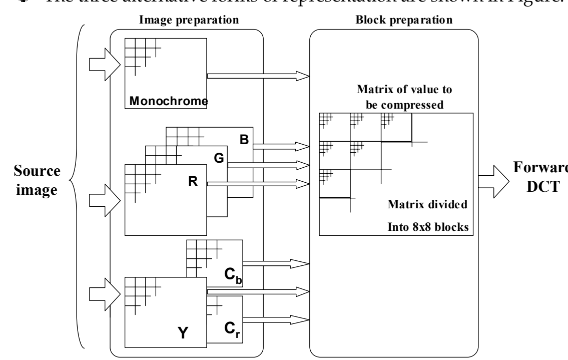
Figure: JPEG Decoder Schematic

Source Image/ Picture

- The source image/picture is made up of one or more 2-D matrices of values.
- **Monochrome Image:**
  - A single 2-D matrix is required to store the set of bit gray-level values that represent the image.
- **Color Image:**
  - The image is represented in an R, G, B format three matrices are required.
  - Y, C_b, C_r, can optionally be used. The two chrominance signals, C_b, and C_r, require half the bandwidth of the luminance signal, Y.

Source Image/ Picture

- The three alternative forms of representation are shown in Figure.



Spatial Compression

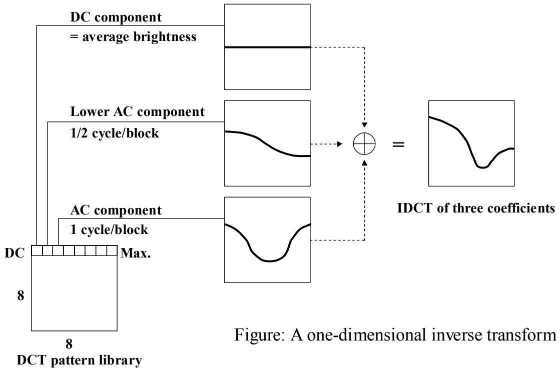


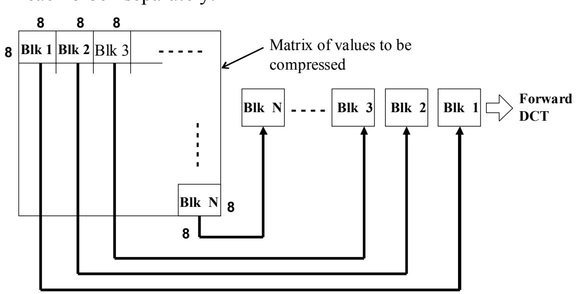
Figure: A one-dimensional inverse transform

Image/Block Preparation

- Once the source image format has been selected and prepared, the set of values in each matrix are compressed separately using the **DCT**.
- However, **block preparation** is carried out. To compute the transformed value for each position in a matrix requires the values in all the locations of the matrix to be processed.
- It would be too time consuming to compute the DCT of the total matrix in a single step so each matrix is first divided into a set of smaller 8 **x** 8 sub-matrices.
- Each is known as a block and these are then fed sequentially to the DCT which transforms each block separately.

Image/Block Preparation

- Each is known as a block and, as we can see in the figure, these are then fed sequentially to the DCT which transforms each block separately.



Forward DCT

- In order to compute the forward DCT, all the values are first centered around zero by subtracting 128 from each intensity/luminance value.
- The input 2-D matrix is represented by: P[***x***, ***y***] and the transformed matrix by F[***i***, ***j***], the DCT of each 8 **x** 8 block of values is computed using the expression:

$$F(i,j) = \frac{1}{4} C(i) C(j) \sum_{x=0}^{7} \sum_{y=0}^{7} P[x,y] \cos\frac{(2x+1)i\pi}{16} \cos\frac{(2y+1)j\pi}{16}$$

where ***C(i)*** and ***C(j)*** = $1/\sqrt{2}$ for ***i,j***=0
= 1 for all other values of ***i*** and ***j***

And ***x***, ***y***, ***z*** and ***j*** all vary from 0 through 7

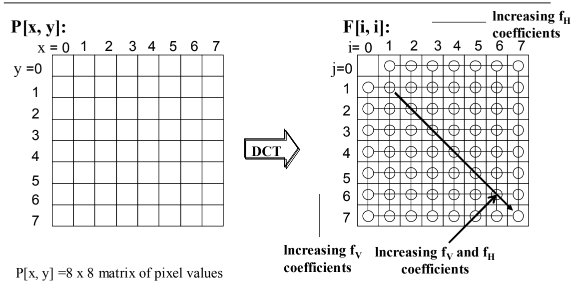
Forward DCT

- Each pixel value is quantized using 8 bits which produces a value in the range 0 to 255 for the intensity/luminance values - *R, G,* B, or Y - and a value in the range -128 to +127 for the two chrominance values - C_b, and C_r.
- The follow points are summarized in Figure:
  1. All 64 values in the input matrix, P[***x, y***] contribute to each entry in the transformed matrix, F[***i, j***] .
  2. For ***i*** = ***j*** = 0, the two cosine terms (and hence horizontal and vertical frequency coefficients) are both 0. Also, since cos(0)= 1, the value in location *F* [0,0] of the transformed matrix is simply a function of the summation of all the values in the input matrix. Essentially, it is the mean of all 64 values in the matrix and is known as the **DC coefficient.**

Forward DCT

3. Since the values in all the other location s of the transformed matrix have a frequency coefficient associated with them – either horizontal (***x*** = 1 – 7 for ***y*** = 0), vertical (***x*** = 0 for ***y*** = 1 – 7) or both (***x*** = 1 – 7 for ***y*** =1 – 7) – they are known as the **AC coefficient.**
4. For ***j*** = 0, only horizontal frequency coefficients are present which increase in frequency for ***i*** = 1 – 7.
5. For ***i*** = 0, only vertical frequency coefficients are present which increase in frequency for ***j*** = 1 – 7.
6. In all other locations in the transformed matrix, both horizontal and vertical frequency coefficients are present to varying degrees.

Forward DCT



P[x, y] =8 x 8 matrix of pixel values

F[i, j] = 8 x 8 matrix of transformed values/spatial frequency coefficients

In F[i, j]: ☐ = DC coefficient ☐ = AC coefficients

f_H = horizontal spatial frequecy coefficient

f_V =vertical spatial frequency coefficent

Figure: DCT computation features

Forward DCT

- Each regions of picture that contain a single color will generate a set of transformed blocks all of which will have firstly, the same **DC coefficients** and secondly, only a few **AC coefficients** within them.
- The whole areas of a picture which contain color transitions that will generate a set of transformed blocks with different DC coefficients and a larger number of AC coefficients within them.
- It is these features that are exploited in the quantization and entropy encoding phases of the compression algorithm.

Quantization

- The quantization process aims to reduce the size of the DC and AC coefficients so that less bandwidth is required for their transmission.
- This property is exploited in the quantization phase by dropping - setting to zero - those **spatial frequency coefficients** in the transformed matrix whose amplitudes are less than a defined threshold value.
- Comparing each coefficient with the corresponding threshold value, a division operation is performed using the defined threshold value as the divisor.
- These are held in the **quantization table** with the threshold value to be used with a particular DCT coefficient in the corresponding position in the matrix.

Quantization

- The threshold values used vary for each of the 64 DCT coefficients. These are held in a two-dimensional matrix known as the **quantization table** with the threshold value to be used with a particular DCT coefficient in the corresponding position in the matrix.
- The JPEG standard includes two default quantization table values
  - One for use with the luminance coefficients and
  - the other for use with the two sets of chrominance coefficients.
- It also allows for customized tables to be used and sent with the compressed image.

Quantization

- An example set of threshold values is given in the quantization table shown in Figure together with a set of DCT coefficients and their corresponding quantized values. We can conclude a number of points from the values shown in the tables:
  - The computation of the quantized coefficients involves rounding the quotients to the nearest integer value.
  - The threshold values used, in general, increase in magnitude with increasing spatial frequency.
  - The DC coefficient in the transformed matrix is largest.
  - Many of the higher-frequency coefficients are zero.

Quantization

DCT coefficients

120	60	40	30	4	3	0	0
70	48	32	3	4	1	0	0
50	36	4	4	2	0	0	0
40	4	5	1	1	0	0	0
5	4	0	0	0	0	0	0
3	2	0	0	0	0	0	0
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Quantizer

Quantized coefficients (rounded value)

12	6	3	2	0	0	0	0
7	3	2	0	0	0	0	0
3	2	0	0	0	0	0	0
2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

10	10	15	20	25	30	35	40
10	15	20	25	30	35	40	50
15	20	25	30	35	40	50	60
20	25	30	35	40	50	60	70
25	30	35	40	50	60	70	80
30	35	40	50	60	70	80	90
35	40	50	60	70	80	90	100
40	50	60	70	80	90	100	110

Quantization table

Figure: Example computation of a set of quantized DCT coefficients

Entropy Encoding

- The entropy encoding stage comprises four steps: ***vectoring***, ***differential encoding***, ***run-length encoding***, and ***Huffman encoding***.

Vectoring:

- The various entropy encoding algorithms operate on a one-dimensional string of values, that is, ***a vector***.
- The output of the quantization stage is a 2-D matrix of values and apply entropy encoding to the set of values in the form of a single-dimension vector. This operation is known as **vectoring.**
- In order to exploit the presence of the large number of zeros in the quantized matrix, a **zig-zag** scan of the matrix is used as shown in figure

Entropy Encoding

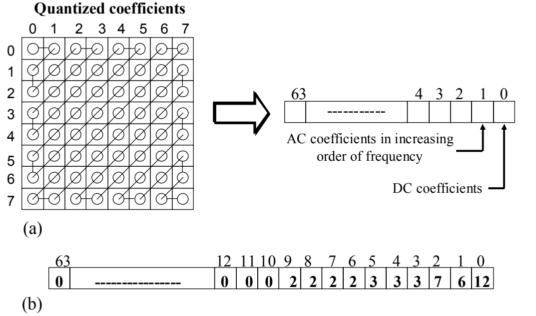


Figure: Vectoring using a zig-zag scan: (a) principle; (b) vector for example

Entropy Encoding

- From the figure, with this type of scan, the DC coefficient and lower-frequency AC coefficients - both horizontal and vertical - are scanned first.
- All the higher-frequency coefficients are in a sequential order so making this form of representation more suitable for compression.
- Differential encoding, which is applied to the DC coefficient only, and the second is run-length encoding, which is applied to the remaining values in the vector containing the AC coefficients.

Entropy Encoding

Differential Encoding:

- The first element in each transformed block is the DC coefficient which is a measure of the average color/luminance/chrominance associated with the corresponding 8 x 8 block of pixel values.
- DC coefficient is the largest coefficient and, because of its importance, its resolution is kept as high as possible during the quantization phase.
- Because of the small physical area covered by each block, the DC coefficient varies only slowly from one block to the next.

Entropy Encoding

- The first difference value always being encoded relative to zero. The difference values are then encoded in the form ***(SSS, value)*** where the ***SSS*** field indicates the number of bits needed to encode the value and the ***value*** field the actual bits that represent the value.
- The rules used to encode each value are summarized in next page figure.
- The number of bits required to encode each value is determined by its magnitude.
- A positive value is then encoded using the unsigned binary form
- A negative value by the complement of this.

Entropy Encoding

(a)

Difference value	Number of bits needed (SSS)	Encoded value
0	0	
-1,1	1	1= 1 , -1 = 0
-3, -2, 2, 3	2	2 = 10 , -2 = 01 3 = 11 , -3 = 00
-7..-4, 4..7	3	4 = 100 , -4 = 011 5 = 101 , -5 = 010 6 = 110 , -6 = 001 7 = 111 , -7 = 000
-15…-8, 8…15	4	8 = 1000 , -8 = 0111 :

Figure: Variable-length coding – coding categories

Entropy Encoding

- **Example:**

  Determine the encoded version of the following difference values which relate to the encoded coefficients from consecutive DCT blocks:

  12, 1, -2, 0, -1

Answer:

Value	SSS	Value
12	4	1100
1	1	1
-2	2	01
0	0	
-1	1	0

Entropy Encoding

Run-length Encoding:

- The remaining 63 values in the vector are the AC coefficients and, because of the zig-zag scan, the vector contains long strings of zeros within it.
- The **AC coefficients** are encoded in the form of a string of pairs of values.
- Each pair is made up of **(skip, *value)*** where ***skip*** is the number of zeros in the run and ***value*** the next non-zero coefficient.
- Hence the 63 values in the vector shown earlier in figure would be encoded as:

(0,6)(0,7)(0,3)(0,3)(0,3)(0,2)(0,2)(0,2)(0,2)(0,0)

Entropy Encoding

- Note that the final pair (0,0) indicates the end of the string for this block and that all the remaining coefficients in the block are zero.
- The value field is encoded in the form SSS/value.
- **Example:** Derive the binary form of the following run-length encoded AC coefficients:

(0,6)(0,7)(3,3)(0,-1)(0,0)

Answer:

AC coefficient	Skip	SSS/Value
0,6	0	3 110
0,7	0	3 111
3,3	3	2 11
0,-1	0	1 0
0,0	0	0

Entropy Encoding

Huffman Encoding:

- A table of codewords is used with the set of codewords precomputed using the Huffman coding algorithm.
- For the differential-encoded **DC coefficients** in the block, the bits in the ***SSS*** field are not sent in their unsigned binary form as shown in Differential Encoding Example.
- This is done so that the bits in the ***SSS*** field have the prefix property and this enables the decoder to determine unambiguously the first SSS field from the received encoded bitstream.

Entropy Encoding

- **Example:** Determine the Huffman-encoded version of the following difference values which relate to the encoded DCT coefficients from consecutive DCT blocks:

12, 1, -2, 0, -1

Default Huffman codewords defined in the following table.

Number of bits needed (SSS)	Huffman codeword
0	010
1	011
2	100
3	00
4	101
5	110
6	1110
7	11110
⋮	⋮
11	111111110

Entropy Encoding

Answer:

Value	SSS	Huffman-encoded SSS	Encoded value	Encoded bitstream
12	4	101	1100	**1011100**
1	1	011	1	**0111**
-2	2	100	01	**10001**
0	0	010		**010**
-1	1	011	0	**0110**

The Encoded bitstream that contain the default-encoded SSS value and the related encoded value.

Entropy Encoding

- For each of the run-length encoded **AC coefficients** in the block, the bits that make up the ***skip*** and ***SSS*** fields are treated as a single composite symbol and this is then encoded using either the default table of Huffman codewords.
- In the next example, to decode the received bitstream the receiver first searches the bitstream for a valid codeword and determines the corresponding SSS fields from the Huffman table.
- The use of variable-length codewords in the various parts of the entropy encoding stage, this is also known as the variable-length coding (VLC) stage.

Entropy Encoding

- **Example:** Derive the composite binary symbols for the following set of run-length encoded AC coefficients:

  (0,6)(0,7)(3,3)(0,-1)(0,0)

  Assuming the ***skip*** and ***SSS*** fields are both encoded as a composite symbol, use the Huffman codewords shown in the following table to derive the Huffman-encoded bitstream for this set of symbols.

Skip /SSS	Codeword	Skip /SSS	Codeword
0 / 0	1010 (=EOB)	3 / 1	11101
0 / 1	00	3 / 3	11111110111
0 / 6	111000	3 / 9	1111111110010101
0 / 7	1111000	3 / 10	1111111110010110
0 / 9	1111111110000010	14 / 1	111111110110
0 / 10	1111111110000011	14 / 10	1111111111110100
1 / 1	1100	15 / 0	111111110111
1 / 10	1111111110001000	15 / 10	1111111111111110

Entropy Encoding

Answer:

AC coefficient	Composite skip	Symbol SSS	Huffman codeword	Run-length value
0,6	0	3	100	110 = 6
0,7	0	3	100	111 = 7
3,3	3	2	111110111	10 = 2
0,-1	0	1	00	1 = 0
0,0	0	0	1010	

The Huffman-encoded bitstream is then derived by adding the run-length encoded value to each of the Huffman codewords:

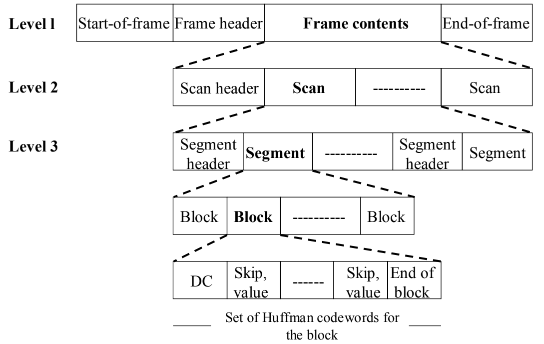
100110 100111 11111011110 000 1010

Entropy Encoding

Frame Building:

- the bitstream output by a JPEG encoder, which is stored in the memory of a computer ready for either integrating with other media if necessary or accessing from a remote computer.
- The JPEG standard, also includes a definition of the structure of the total bitstream relating to a particular image/picture. This is known as a ***frame*** and its outline structure is shown in the next page figure.
- The role of the **frame builder** is to encapsulate all the information relating to an encoded image/picture in this format and the structure of a frame is hierarchical.

Entropy Encoding



JPEG encoder output bitstream format

Entropy Encoding

- **At the top level**, the complete frame-plus-header is encapsulated between a ***start-of-frame*** and an ***end-of-frame*** delimiter which allows the receiver to determine the start and, end of all the information relating to a complete image/picture.
- The ***frame*** header-contains a number of fields that include:
  - the overall width and height of the image in pixels;
  - the number and type of components that are used to represent the image (CLUT, R/G/B, Y/ C_b/C_r) ;
  - the digitization format used (4:2:2, 4:2:0 etc.).

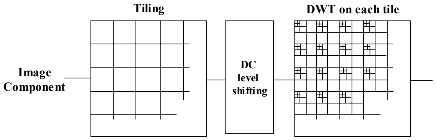
Entropy Encoding

- At the second level, a frame consists of a number of components each of which is known as a ***scan***. These are also preceded by a header which contains fields that include:
  - the identity of the components (R/G/B etc.);
  - the number of bits used to digitize each component;
  - the quantization table of values that have been used to encode each component.

Entropy Encoding

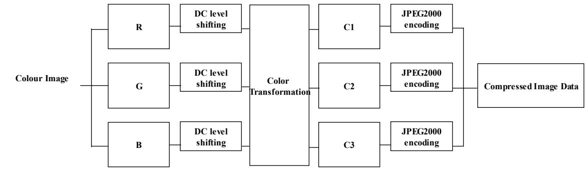
- Typically, each scan/component comprises one or more ***segments*** each of which can contain a group of (8 x 8) blocks preceded by a header.
- This contains the Huffman table of values that have been used to encode each block in the segment should the default tables not be used.
- Each segment can be decoded independently of the others which overcomes the possibility of bit errors propagating and affecting other segments.
- Each complete frame contains all the information necessary to enable the JPEG decoder to identify each field in a received frame and then perform the corresponding decoding operation.

JPEG-2000 Image Tiles

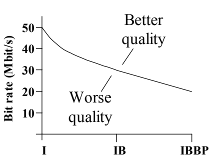


Tiling, DC level shifting and DWT on each image tile component

JPEG 2000 Multiple-component Encoder



MPEG Bidirectional Coding



GOP-Group of picture

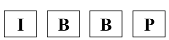


Figure: Constant Quality Curve

I=Intra or spatially coded ‘anchor’ frame
P= Forward predicted frame. Coder sends difference between I and P. Decoder adds difference to create P
B=Bidirectionally coded picture. Can be coded from a previous I or P picture or a later I or P picture. B frames are not coded from each other.

An MPEG-2 Coder

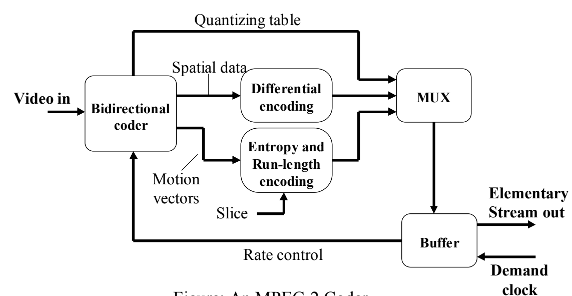


Figure: An MPEG-2 Coder

Forward Adaptive Bit Allocation (MPEG):

The **psychoacoustic models** associated with the various ***MPEG coders*** control the quantization accuracy of each subband sample by computing and allocating the number of bits to be used to quantize each sample.

Since the quantization accuracy that is used for each sample in a subband may vary from one set of subband samples to the next, the ***bit allocation information*** that is used to quantize the samples in each subband is sent with the actual quantized samples.

This information is then used by the decoder to dequantize the set of subband samples in the frame.

Forward Adaptive Bit Allocation (MPEG):

This mode of operation of a perceptual coder is known, therefore, as the ***forward adaptive bit allocation*** mode and, for comparison purposes, a simplified schematic diagram showing this operational mode is given in figure.

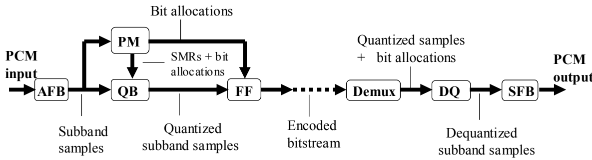


Figure: Perceptual coder schematics: forward adaptive bit allocation (MPEG)

Forward Adaptive Bit Allocation (MPEG):

It has the advantage that the psychoacoustic model is required only in the encoder.

It has the disadvantage, that a significant portion of each encoded frame contains bit allocation information which, in turn, leads to a relatively inefficient use of the available bit rate.

Reference

- Halsall, Fred, Multimedia Communications – Applications, Networks, Protocols and Standards, Addison-Wesley, 2001.
- Richardson, Iain, Video Codec Design – Developing Image and Video Compression Systems, John Wiley, 2002.
- Rao, K.R., Bojkovic, Z.S., Milovanovic, D.A., Multimedia Communication Systems, Techniques, Standards and Networks, Ed., Prentice Hall PTR, 2002.
- Special Issue on Multimedia Signal Processing, Part One, Proceedings of The IEEE, Vol. 86, No. 5, pp. 749-1024, May 1998.