Chapter 1

Introduction

What is a Medium?

- A means of *conveying and distributing information*. (e.g., print and broadcast media.)
- A material and form of artistic expression. (e.g, paper, stone, inks, musical instruments.)



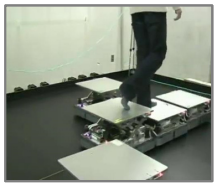

natural media

- A kind of natural medium

Natural and Digital Media

- Natural Media rely on physical elements -- paper, stone, inks and paints, and musical instruments.
- Digital Media rely on the computer and “bits”.
- In this course, “media” refer to the types of digital information carriers, such as textual data, images, audio, and video.
- Why digital?
  - Easy integration & sharing of resources (storage, transmission network).
  - Allows encryption, compression etc.
  - More reliable
  - Can be used and copied many times without losing quality.

What is Multimedia?



- Multimedia combines audio and visual material to provide computerized interaction of text, sound, graphics, images, animation & video to enhance communication and to enrich its presentation.
- Explores the use of various sensory channels and modes of expression.
- Five senses: sound, sight, touch, taste & smell. Current technology allows us to handle sound and sight and partially touch (Phantom in our SGI lab). Taste and smell are not tangible currently.

What is Multimedia? (2)

- Multimedia systems handle at least one type of "**continuous** media" as well as "**discrete** media".
- We used the following definition of **Multimedia**:

"*A multimedia system is characterized by computer-controlled, integrated production, manipulation, presentation, storage and communication of independent information, which is encoded at least through a continuous (time-dependent) and a discrete (time-independent) medium.*"

Why Multimedia?

- Picture, sound or even touch, taste and smell are sometimes **more effective** to convey information than boring text.
- One extreme example: Virtual Reality
- People are demanding such “TV features” as sound, image, and video from their computers.
- They also want such “computer features” such as interactivity and content-on-demand for the television.
- Merging of computer and TV industries
  e.g. Microsoft (computer company) partnered with MSNBC (TV broadcast company)
  e.g. Apple launched Apple TV

Why Multimedia? (2)

- Many MM applications are used to **emulate** human communications and to assist the human being in organizing & managing vast amounts of information in various media types (e.g., a photo database for fashion models).

The Implications

- TV: analog to digital, e.g. HDTV
- Digital TV: 2-way interaction + digital A / V => merging of 3 industries: computer, communication, and broadcasting
- To provide for creation, editing, transport, and distribution of digital A/V.
  (e.g., Netflex/NowTV video-on-demand service)
- Internet-based TV (youtube channel; TVB->myTV box) and radio stations
- Digital cinema with stereo vision (e.g. Avatar)

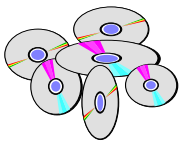
- Current 3D TV is only stereo TV
- Real 3D is like this

- Lytro camera
- Panoramic video (Walking New York | 360 VR Video)

Computer Multimedia Applications

- Some are used to embellish the human-computer interface by adding sounds, animation, and other forms of multimedia.
- (More importantly) some are used to manipulate multimedia data, and these data is central to the applications.
- Some common functions: to capture, generate, store, retrieve, process, transmit, and present multimedia information.

Multimedia Application



- Digital Photo Album
  - How to mange thousands of personal photos? Phototourism: https://photosynth.net/
- Home theater
  - A blu-ray disc can hold 25GB per layer. There can be up to 4 layers for BD-XL.
  - Provide high-quality video and audio
- Electronic Games
  - Each game is a highly interactive multimedia application that presents layered 2D/3D animation with synchronized sound effects and music.
  - New generation of computer games usually use DVD or even Blu-ray to store the game data.

- Hypermedia Browsers
  - A hyperdocument is a document composed of multiple media types. On the world-wide-web, hyperdocuments are organized by linkings. We can traverse and retrieve the hyperdocuments through these links.
- Multimedia Presentation Systems
  - An “engine” that displays, synchronizes, provides *interaction* with, and generally manipulates multimedia material. (e.g., Authorware, Macromedia Director, Powerpoint, Flash)
- Multimedia Mail / Instant Messengers / Blog / Album
  - Handles electronic messages containing audio, graphics, and other media, e.g. MIME mails, whatsapp, instagram, twitter

- Teleconferencing
  - Computers equipped with microphone, speakers, and video cameras, and placed on a multimedia network, can establish audio and video connections between each other.
  - Multi-user tools, such as group editors. A group editor allows conference participants to share documents, and to edit the documents simultaneously, e.g. Skype video chat, iphone FaceTime
- Multimedia Services
  - Interactive Shopping
  - Tele-banking
  - Education
  - Medical Services (telemedicine)
  - Video-on-demand, music-on-demand (iTunes)
  - Information-on-demand, wikipedia

Media Types

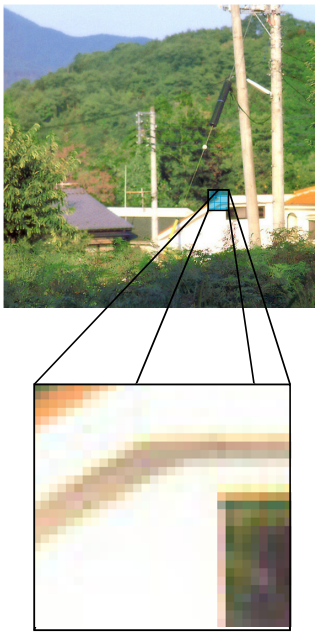
- Non-temporal (Discrete) -- do not have a time dimension, and their contents & meanings do not depend on the presentation time.
  - Text
  - Image (2D samples, e.g. GIF, JPEG)
  - Graphics (2D/3D geometry, e.g. CorelDraw, VRML)
- Temporal (Continuous, Isochronous) -- have a time dimension. They convey meanings only if “displayed” at a specific rate.
  - Video, Animation
  - Digital Audio (sampled sound, like image)
  - Music (e.g. MIDI, symbolic represented sound, like graphics)

Text

- Not visually exciting.
- Conveys essential and precise information.
- Text representation:
  e.g. ASCII - English (1 byte)
  e.g. BIG5, GB - Chinese (2 bytes)
  e.g. Shift-JIS - Japanese (2 bytes)
  e.g. UTF-8, UTF-16
  - Unicode multi-language (up to 4 bytes)
- Storage “friendly” (compact)
- Sometimes, certain information is too clumsy to be captured by words. e.g. try to write an essay to describe the right figure.

Digital Images

- Two-dimensional arrays of pixels of varying color.
- Color model: how to specify the color of a pixel?
  - RGB: Colors are represented by a numeric triple specifying red (R), green (G), and blue (B) intensities.
  - YUV, YIQ, HSV
  - CYMK for publishing
  - 24 bit really sufficient?

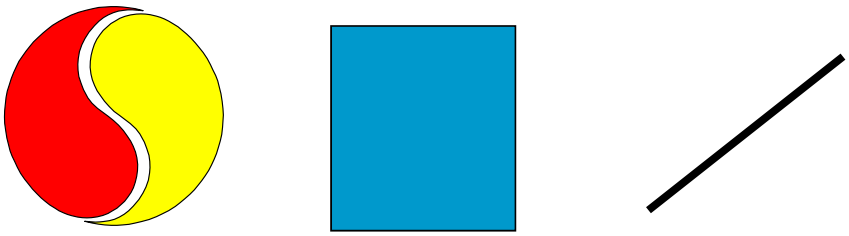


Digital Images (2)

- Compression
  - A page-sized 24-bit color image with 300 pixels per inch takes up about 20Mbytes.
  - Lossless and lossy.
  - Many different standards: JPEG, GIF, TIFF,....
- Image processing:
  blurring, sharpening, edge detection, filtering.
- Image transformation:
  e.g., morphing (transform one image into another).

Graphics

- Graphics data are represented by a geometric model and a rendering process.
- 2D graphics is sometimes called vector graphics.
  2D geometric primitives: circle, rectangle, line ...
  Rendering process: rasterization of these primitives
  e.g. CorelDraw, MacDraw.



Graphics (2)

- 3D graphics consists of 3D geometric primitives like triangles, circle, plane, surface patches, …
  The rendering process is actually a physical simulation of light propagation.
- Difference between image and graphics:
  - Image is actually a 2D array of color samples.
  - Once recorded, cannot be changed.
  - Usually large storage needed
  - e.g. JPEG, GIF
  - Graphics consists of 2D/3D geometric primitives and rendering process.
  - Can be easily modified.
  - Usually more compact, but not always.
  - e.g. VRML, OpenGL,WebGL, CorelDraw, Adobe Illustrator

Video

- A sequence of images called frames. "*persistence of vision*".
- Attributes: Frame rate, Resolution, Aspect Ratio, Interlacing.
- Formats. e.g. NTSC (National Television Systems Committee) PAL (Phase Alternation Line).
- Current video broadcast is in both analog and digital

format	frame rate	scan lines	aspect ratio
NTSC	30	525	4:3
PAL	25	625	4:3
HDTV(US)	30	1125	16:9
HDTV(EURO)	25	1250	16:9

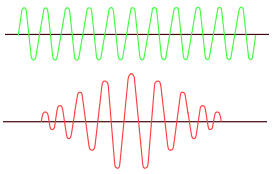
Video (2)

- Theoretically, almost any color can be produced by mixing 3 primary colors (red, green, blue). An analog video camera produces 3 distinct continuous signals, one for each color component.
- Luminance/chrominance principle: the three primary colors can be converted into 2 parts:
  - Luminance: information on the brightness of the image.
  - Chrominance: information on the color of the image.
- Because the human eye is not very sensitive to color information, the bandwidth of these 2 color components can be independently reduced before transmission.

Digital Video

- A video can also be represented by a sequence of digital images.
- broadcast quality video: 1 sec = 10MB.
- For lesser quality, and a good compression technique, it is possible to achieve: 1 sec = 1M bits => transfer rate of CD-ROM => VCD.
- Compression: Lossless and lossy.
- For lossy compression, can achieve 50:1 or higher.
- MPEG: The Moving Pictures Expert Group MPEG-1:1.2-1.5Mbps VHS quality video. MPEG-2: 4-10 Mbps. Used in DVD.

Digital Audio



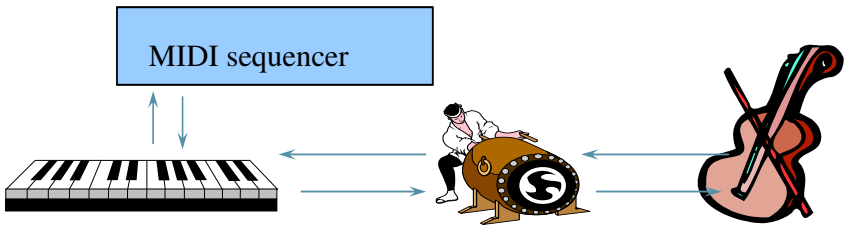
- Digital audio representation
  - Produced by sampling a continuous signal generated by a sound source.
  - A process called analog-to-digital conversion (A/D).
- Sampling frequency
  - Human ear is sensitive to frequencies of up to about 20kHz.
  - Sampling frequency > 40kHz (for CD => 44.1kHz, 16 bits per sample).
- Number of channels
  - 2 for stereo
  - Some audio editing equipment may have 16 or 32 channels.
  - 5 for AC3 (surround sound)

Digital Audio

- Storage
  - An hour of high quality stereo digital audio requires > 500MBytes of storage.
  - A CD-ROM can store about 650MBytes of data.
- Digital audio effects
  - Echoing
  - Equalization
  - Noise reduction
  - Pitch shifting
  - Acoustic environment

Music

- MIDI -- Musical Instrument Digital Interface
  - Digital musical instruments send MIDI messages to a sequencer.
    - Notes, temp, velocity, beat, bars, multiple instruments
  - The sequencer composes the music according to the messages received.
  - The sequencer/ synthesizer has a “palette” of sounds for each type of instrument



Music (2)

- Why distinguish digital audio and music?
- In fact any sound or music can be sampled, recorded and finally represented as digital audio.
- Analogous to image and graphics.
- What “music” mentioned here consists of musical primitives and synthesis process.
- Therefore it is more compact than digital audio.

Challenges

- Multimedia stresses all components of a computer system (volume & time constraints).
- CPU processing power
  - Fast speed for data capturing, codec, data enhancement. (Large amounts of data being processed in real-time).
- Storage and Memory
  - High capacity, fast access time, high transfer rates.
- System architecture
  - High bus bandwidth, efficient I/O.
- Software
  - Tools for retrieval and data management of continuous media data.

Challenges (2)

- Operating systems
  - Support for new data types, real-time scheduling, multimedia file systems, time-critical synchronization.
- Human Computer Interface
  - e.g. Multi-touch, kinect, HMD, 3D display, etc
- Networks
  - High bandwidth, low latency, low jitter.
- Application latency
  - e.g., video playback requires end-to-end jitter control within a couple of milliseconds.
- Synchronization
  - what is the tolerance?
  - how do we achieve?

Research Areas

1. fast processors
2. high-speed networks
3. large capacity storage devices
4. video & audio compression algorithms
5. graphics systems
6. human-computer interface
7. real-time operating systems
8. information storage and retrieval
9. hypertext & hypermedia
10. languages for scripting
11. parallel processing methods

Compression

- Throughput and storage
  - *"If a picture is worth a thousand words, then a video is worth 414 million (4-byte) words per minute,"* (or 25GBytes/hr !!!!)
  - What are the requirements on network data transfer rate? On disk I/O?
- Compression makes it possible.
  - 50:1 yields 0.5 GBytes/hr.
  - 200:1 for HDTV.
  - How expensive is the processing?
    Hardware solution?
    Software solution?

Compression (2)

- MM systems require data compression for 3 reasons:
  - the large storage requirements of MM data (30 MB/s).
  - relatively slow storage devices that cannot play MM data in real time.
  - Network bandwidth that does not allow real-time video data transmission.

Some Compression Standards

JPEG	Digital compression and coding of continuous-tone still images	Joint Photographic Experts Group	15:1 (full color still-frame applications)
H.261	Video coder/decoder for audio-visual services at p*64 Kbps	Specialist Group on Coding for Visual Telephony	100:1 to 2000:1 (video-based telecommunications)
MPEG	Coding of moving pictures and associated audio	Moving Pictures Experts Group	200:1 Motion-intensive applications

Multimedia Networking

- Many multimedia applications, such as video mail, video conferencing, and video-on-demand, require the support of a high performance network system.
- In these applications, the multimedia objects are stored at a server and played back at the clients' sites.
- Remote retrieval of multimedia objects has stringent time constraints.
- Delay: the amount of time it takes to transmit a data unit (e.g., a video frame) from a sender to a receiver.
- Jitter: delay variation.

Multimedia Networking (2)

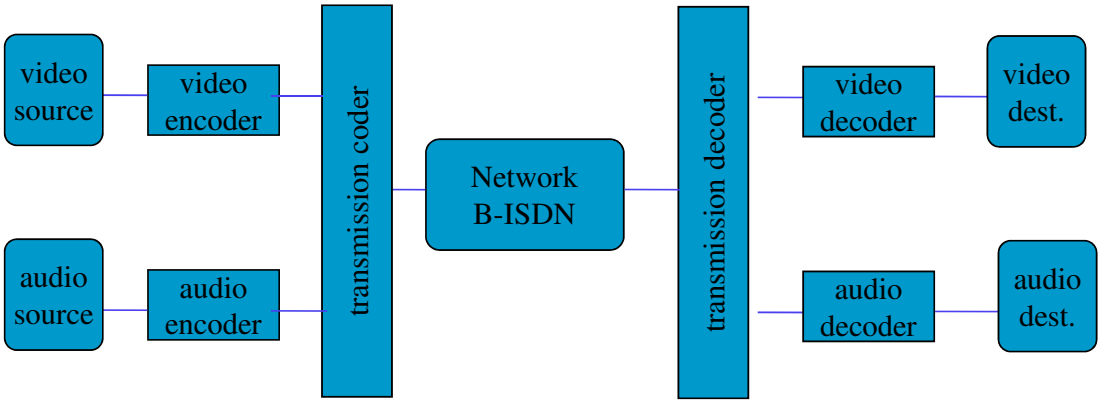
Characteristics	*Data Transfer*	*Multimedia transfer*
Data rate	Low	High
Traffic pattern	Bursty	Stream-oriented
Reliability requirements	No loss	Some loss
Latency requirements	None	Low, e.g., 20 ms
Mode of communication	Point-to-point	Multipoint, Point-to-point
Temporal relationship	None	Synchronized transmission

Synchronization

- Continuous synchronization
  - requires constant synchronization of lengthy events.
  - e.g., video-phone
- Point synchronization
  - a single point of one media block coincides with a single point of another media block.
  - e.g., slide show with blocks of audio allotted to each slide.

Synchronization Example

Asynchrony in a video telephone system



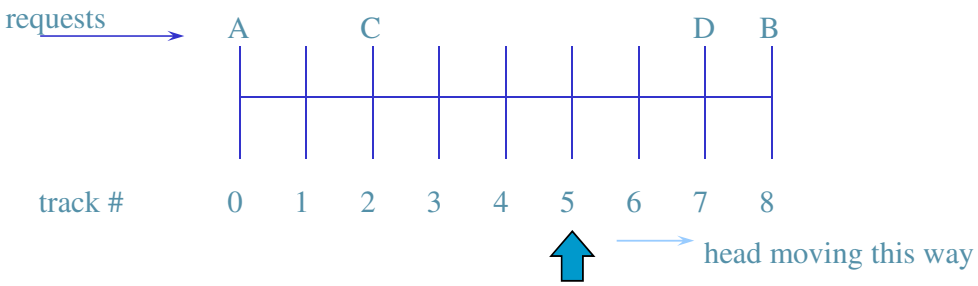
different encoding times for video and audio

variations in packet arrival times (jitter) introduced by:
-independent routing paths,
-packet loss,
-packet buffering

start-up delay due to decompression pipeline can cause video to trail audio where decompression delay is generally smaller.

I/O Scheduling

- In a disk-based database system, disk I/O occupies a major portion of transaction execution time.
- Disk seek time, which accounts for a very significant fraction of disk access latency, depends on the disk head movement. The order in which I/O requests are serviced, therefore, has an immense impact on the response time and throughput of the I/O subsystem.
- Example. HPF, Elevator, FD-Scan, HPGF.



- Highest Priority First (HPF) -- serve the request that has the highest priority (earliest deadline for example).
- Elevator -- scan the disk and pick up requests “on-the-way”.
- FD-SCAN -- always “target” the disk head towards the track with the highest priority request, but also service whatever requests are on its way.
- Highest Priority Group First (HPGF). Disk requests are grouped into a small number of priority levels. The disk is scheduled to service the highest priority group first. The Elevator Algorithm is used for the intra-group scheduling.

Multimedia Information Retrieval

- To retrieve a text document from the Web, we use keyword search via “Google”, for example.
- To retrieve a record from a relational database, such as Oracle, we use a SQL statement.
- To retrieve a picture, how shall we formulate a query?
- What about audio? How do we describe a sound?
- To be honest, this is still a hard research problem which is not yet been solved.

Summary

- Large data volume
  - How to store?
  - How to transfer?
  - Do we need careful scheduling?
- Multiple media channels
  - How to capture?
  - How to display/play/present?
  - How to synchronize them?

Text Alone is Never Sufficient