LBSC 690: Week 9

Multimedia

Jimmy Lin
College of Information Studies
University of Maryland

Monday, April 2, 2007

The Gullibility of Human Senses

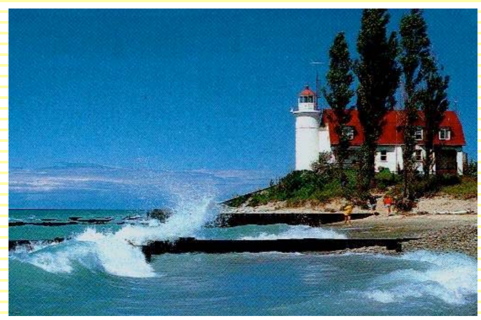
- Three simple tricks for producing
  - Images
  - Video
  - Audio
- But how do you move the bits around fast enough?
  - Remove redundancy
  - Throw away stuff that doesn't matter
- Synchronizing different media to create multimedia

Nothing new…

Seurat, Georges, A Sunday Afternoon on the Island of La Grande Jatte

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

A Picture…



UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

… close up

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Visual Perception

- Closely spaced dots appear solid
  - But irregularities in diagonal lines can stand out
- Any color can be produced from just three
  - Red, Blue and Green: "additive" primary colors
- High frame rates produce apparent motion
  - Smooth motion requires about 24 frames/sec
- Visual acuity varies markedly across features
  - Discontinuities easily seen, absolutes less crucial

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Basic Image Coding

- An image = a collection of picture elements (pixels)
  - Each pixel has a "color"
- Different types of pixels
  - Binary (1 bit): black/white
  - Grayscale (8 bits)
  - Color (3 colors, 8 bits each): red, green, blue
- A 1024x768 image requires 2.4 MB
  - So a picture is worth 400,000 words!

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Monitor Characteristics

- Technology (CRT, flat panel)
- Size
  - 15, 17, 19, 21 inch, or even bigger!
  - Measured diagonally
- Resolution
  - Measured in pixels: 640x480, 800x600, 1024x768, 1280x1024, etc.
- Dot pitch (0.26, 0.28)
- Refresh rate (60, 72, 80 Hz)

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Some Questions

- How many images can a 64 MB flash card store?
  - But mine holds 120. How?
- How long will it take to send an image at 64kb/s?
  - But my Web page loads faster than that. How?
- But in reality images don't have these problems
  - How do we get around these problems?

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Compression

- Goal: send the same information using fewer bits
- Originally developed for fax transmission
  - Send high quality documents in short calls
- Two basic types of data compression:
  - Lossless: can reconstruct exactly
  - Lossy: can't reconstruct, but looks the same
- Two basic strategies:
  - Reduce redundancy
  - Throw away stuff that doesn't matter

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Run-Length Encoding

- Opportunity:
  - Large regions of a single color are common
- Approach:
  - Record # of consecutive pixels for each color
- An example with text:

  Sheep go baaaaaaaaaa and cows go mooooooooooo
  → Sheep go ba<10> and cows go mo<10>

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Using Dictionaries

- Opportunity:
  - Data often has shared substructure, e.g., patterns
- Approach:
  - Create a dictionary of commonly seen patterns
  - Replace patterns with shorthand code
- An example with text:

  The rain in Spain falls mainly in the plain
  → The r* ^ Sp* falls m*ly ^ the pl* (*=ain,^=in)

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Using “Back Pointers”

- Opportunity:
  - Data often has shared substructure, e.g., patterns
- Approach:
  - Replace commonly seen patterns with “pointers” to previous location in data stream
- An example with text:

The rain in Spain falls mainly in the plain
→ The rain <3,3>Sp<9,4>falls m<11,3>ly <16,3><34,4>pl<15,3>

<*x*, *y*> = go back *x* characters and take *y* characters

Huffman Coding

- Opportunity:
  - Common colors are sent more often
- Approach:
  - Use fewer bits to represent common colors

1	Blue	75%	75x1 = 75	75x2 =150
01	White	20%	20x2 = 40	20x2 = 40
001	Red	5%	5x3 = 15	5x2 = 10
			130	200

Palette Selection

- Opportunity:
  - No picture uses all 16 million colors
  - Human eye does not see small differences
- Approach:
  - Select a palette of 256 colors
  - Indicate which palette entry to use for each pixel
  - Look up each color in the palette
- What happens if there are more than 256 colors?

This is GIF!

Tricking Human Senses

- Opportunity:
  - Eye sees sharp lines better than subtle shading
  - Eye more sensitive to small changes in brightness than in color
- Approach:
  - Retain detail only for the most important parts
  - Approximate changes in image with mathematical curves: Discrete Cosine Transform
    - Efficiently captures smooth transitions and shading
    - Not as good at capturing sharp edges

This is JPEG!

Mixing and Matching

- Image encoding formats mix and match techniques described above
  - GIF: palette selection, LZW compression
  - JPEG: takes advantage of human visual system using DCT
  - PNG: replacement for GIF (no patent restrictions)
  - etc.

Variable JPEG Compression

37 kB (20%)

4 kB (95%)

Hands-on Exercise

- Comparing JPEG and GIF formats
- Load and save two images
  - http://www.umiacs.umd.edu/~jimmylin/LBSC690-2005-Fall/Week9-images/image1.jpg
  - http://www.umiacs.umd.edu/~jimmylin/LBSC690-2005-Fall/Week9-images/image2.gif
- Download the two images:
  - Use MS photo editor to convert each to the other format, and compare the quality and the size
  - Vary the compression rate for the JPEG image, and compare the quality

Discussion Point

- JPEG vs. GIF
- Which format should I use for images in my Web pages?
  - photos
  - text images
  - drawings

Discussion Point

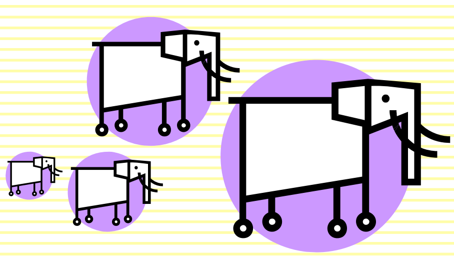
- When is Lossless Compression Important?
  - For images?
  - For text?
  - For sound?
  - For video?

Raster vs. Vector Graphics

- Raster images = bitmaps
  - Actually describe the contents of the image
- Vector images = composed of mathematical curves
  - Describe *how* to draw the image

What happens when you scale vector images?
What happens when you scale raster images?

Example of Vector Images



Basic Video Coding

- Display a sequence of images
  - Fast enough for smooth motion and no flicker
- NTSC Video
  - 60 “interlaced” half-frames/sec, 720x486
- HDTV
  - 30 “progressive” full-frames/sec, 1280x720

Complexity Example

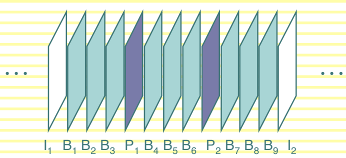
- "NTSC" Quality Computer Display
  - 640 X 480 pixel image
  - 3 bytes per pixel (red, green, blue)
  - 30 Frames per Second
- Bandwidth
  - 26.4 MB/second: corresponds to 180X CD-ROM
  - Exceeds bandwidth of almost all disk drives
- Storage
  - CD-ROM would hold 25 seconds worth
  - 30 minutes would require 46.3 GB
- Some form of compression required!

Video Compression

- Opportunity:
  - One frame looks very much like the next
- Approach:
  - Record only the pixels that change
- Standards:
  - MPEG-1: Web video (file download)
  - MPEG-2: HDTV and DVD
  - MPEG-4: Web video (streaming)

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

MPEG Encoding

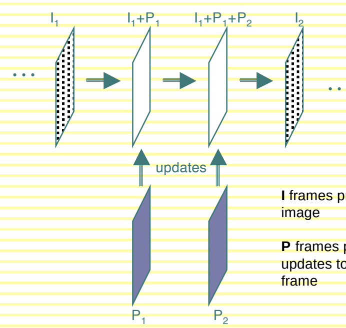


Frame Types

I	Intra	Encode complete image, similar to JPEG
P	Forward Predicted	Motion relative to previous I and P's
B	Backward Predicted	Motion relative to previous & future I's & P's

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Frame Reconstruction

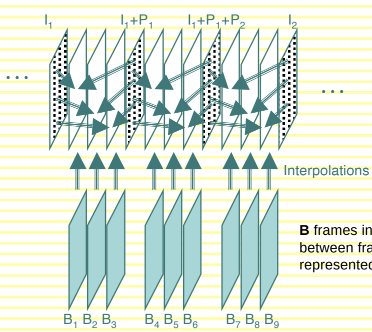


I frames provide complete image

P frames provide series of updates to most recent I frame

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Frame Reconstruction



B frames interpolate between frames represented by I's & P's

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Basic Audio Coding

- Sample at twice the highest frequency
  - 8 bits or 16 bits per sample

- Speech (0-4 kHz) requires 8 kB/s
  - Standard telephone channel (8-bit samples)
- Music (0-22kHz) requires 172 kB/s
  - Standard for CD-quality audio (16 bit samples)

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Speech Compression

- Opportunity:
  - Human voices vary in predictable ways
- Approach:
  - Predict what's next, then send only any corrections
- Standards:
  - Real audio can code speech in 6.5 kb/sec

Demo at http://www.data-compression.com/speech.html

Compact Disk Recording

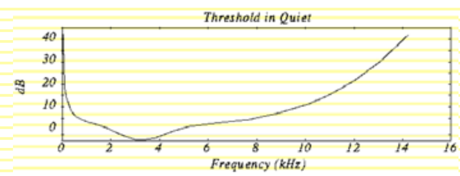
- Parameters
  - 44,100 samples per second
    - Sufficient for frequency response of 22KHz
  - Each sample 16 bits
    - 48 dB range
  - Two independent channels: stereo sound
    - Dolby surround-sound uses tricks to pack 5 sound channels + subwoofer effects
- Bit Rate
  - 44.1K samples/sec x 2 channels x 2 bytes = 172 KB/sec
- Capacity
  - 74 Minutes maximum playing time
  - 747 MB total

How do MP3s work?

- Opportunity:
  - The human ear cannot hear all frequencies at once, all the time
- Approach:
  - Don't represent things that the human ear cannot hear

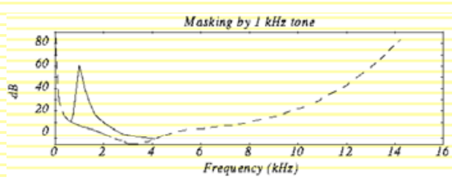
Human Hearing Response

Experiment: Put a person in a quiet room. Raise level of 1kHz tone until just barely audible. Vary the frequency and plot the results.



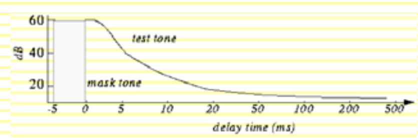
Frequency Masking

Experiment: Play 1kHz tone (masking tone) at fixed level (60db). Play test tone at a different level and raise level until just distinguishable. Vary the frequency of the test tone and plot the threshold when it becomes audible.



Temporal Masking

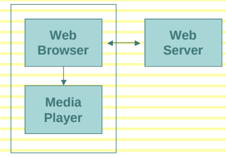
If we hear a loud sound, then it stops, it takes a while until we can hear a soft tone at about the same frequency.



Putting it all together…

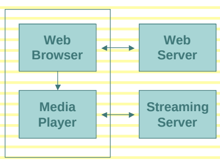
- Psychoacoustic compression:
  - Eliminate sounds below threshold of hearing
  - Eliminate sounds that are frequency masked
  - Eliminate sounds that are temporally masked
  - Eliminate stereo information for low frequencies

Multimedia: Take One



- Object stored in a file
- File transferred as an HTTP object:
  - Received entirely at the client
  - Passed to media player

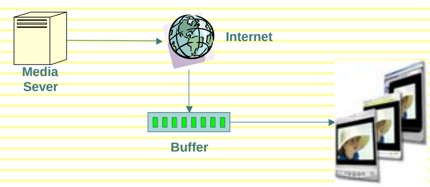
Multimedia: Take Two



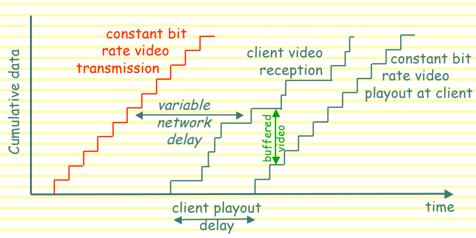
- Browser gets metafile over HTTP
  - Launches media player to interpret the metafile
- Media player contacts streaming server

Streaming Audio and Video

- Buffering a portion of audio/video
- Playing along with receiving
- Analogy: filling and draining a basin concurrently

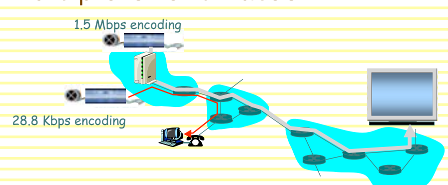


Client Buffering



- Client-side buffering: playout delay compensate for network-added delay

Multiple Client Rates



Q: how to handle different client receive rate capabilities?

- 28.8 Kbps dialup
- 100Mbps Ethernet

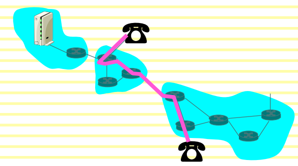
A: server stores, transmits multiple copies of video, encoded at different rates

Discussion Point

- What are the effects of buffering on…
  - Internet radio?
  - Digital TV “on demand”?
  - Skype?
  - AIM?

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Example: Internet Telephony



- Characteristics:
  - “Live”
  - Alternating talk spurts

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

IP Phones: Network Issues

- **Network loss:** packets lost due to network congestion
- **Delay loss:** packets arrives too late for playout at receiver
  - Delays: processing, queueing in network; end-system (sender, receiver) delays
  - Typical maximum tolerable delay: 400 ms
- **Loss tolerance:** depending on voice encoding packet loss rates between 1% and 10% can be tolerated

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

IP Phones: Playout Delay

- Receiver attempts to playout each chunk exactly *q* ms after chunk was generated
  - Chunk has time stamp *t*: play out chunk at *t*+*q*
  - Chunk arrives after *t*+*q*: data arrives too late for playout, data “lost”
- Tradeoff for *q*:
  - Large *q*: less packet loss
  - Small *q*: better interactive experience
- Fancier algorithms: variable playout delay

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Hands On: RealPlayer

- View streaming real video
  - http://www.c-span.org
- Select “Tools/Playback statistics”
- Pay attention to bandwidth and lost packets

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Simple RealAudio Demo

- .ram file specifies
  - URL for the RealAudio stream
  - Dimensions of the picture
  - URL for the picture
- See demo at
  - http://www.umiacs.umd.edu/~jimmylin/LBSC690-2005-Fall/SMIL-demo/

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Synchronizing Multiple Media

- Scripting Languages
  - Synchronized Multimedia Integration Language (SMIL)
- Custom applications
  - Macromedia Flash

SMILe

- W3C standard
  - Player-specific extensions are common
- XML, with a structure similar to HTML

```
<smil>
<head> ... </head>
<body> ... </body>
</smil>
```

- See demo at
  - http://www.umiacs.umd.edu/~jimmylin/LBSC690-2005-Fall/SMIL-demo/

Elements in SMIL

- Window controls (in <head>)
  - Controlling layout: <region>, <root-layout>
- Timeline controls (in <body>)
  - Sequence control: <seq>, <excl>, <par>
  - Timing control: <begin>, <end>, <dur>
- Content types (in <body>)
  - <audio>, <video>, <img>, <ref>

From Media to Multimedia…

- Tricking the human senses:
  - Blending pixels into a seamless image
  - Rapidly cycling through images to create motion
  - Sampling analog waveforms to create digital recordings
- Lots of information required to encode images, movies, and sounds
  - Key is compression!
- Synchronization of different media sources leads to multimedia applications