Handling Pictures and Sound

Outline:

- Information content of pictures (still and video) and audio
- Perception limits our requirements
- Does Moore's law apply?
    - CCD and CMOS imaging costs can be extrapolated
        - 10x decrease in cost over 7 years
        - same takedown rate for prof CCDs, consumer Mpel CCDs, and tethered computer cams
    - Coding effectiveness in video (see below)
- Coding and compression reduce the implementation costs
- Applications drive standards for these
    - Proprietary standards (early) displaced by open ones.
        - But there is still a license fee threshold
    - CD-ROM interactive multimedia established MPEG-1
    - Satellite TV established MPEG-2
        - 40M homes in US (40%), greater impact in Asia
    - Streaming audio and video over the Web still in flux
    - HDTV (as opposed to digital SDTV) – not yet, but sports and feature films are captured and archived in HDTV.
- Almost all handling of multimedia now digital
    - G.722, AC-2, MP3, AC-3 for audio
    - TV still has a strong analog heritage (NTSC,PAL,SECAM)
    - H.261, MPEG-1, MPEG-2, MPEG-4 (only DivX so far)

How much information in a picture?

Consider the earliest photographers:
1850-1870 in the Middle East and American Civil War, then the far west – mules carried tents, giant cameras, wet collodion on glass plate technology, and pictures were 18”x22” (2400 cm^2).
1 cm^2 = 10^6 possible silver grains, so a picture was worth 2.4Gb or 300 MB (very rough estimate)!

Color films, finer grain increase the information density, but todays pictures are made on 35-70 mm film at most.

A quality laser printer (1200 dpi) gives 6 MB in a 5”x7” print.

Broadcast TV (NTSC, 1950s): 30 Hz (actually 60 Hz, but interlaced half-frames) x 525 lines vertically, about 700 points resolution horizontally => 5.5 MHz bandwidth
(PAL, SECAM) have 25Hz x 625 lines x 800 points and use 6.5 MHz bandwidth

TV separates colors into luminance (Y) and chrominance (U,V)
Where Y = .3R + .6G + .1B, U ~ B-Y, V ~ R-Y, and devotes 2x to 4x as much bandwidth to Y as to U,V.

TV and printers are clearly perception-limited, yet look good.

What’s in audio?

Humans sensitive to 20-20K Hz, and a pressure range of about 100 db. This requires >~ 40K samples per second (by Nyquist/Shannon), and 14-16 bits of data per sample. CD-audio, using PCM (pulse-coded-modulation – that is, just list the digital values of the data sampled at each time) pretty much does that.

Stereo doubles the bit rate to support two ear perception. 4, 5, and 7 channel variants contain the same information, but delivered redundantly to spread the experience over a larger spatial area.

If the only criteria is intelligible voice transmission, as in G.722:
Sample at 16 KHz, take 8 bits/sample, Fourier transform and allocate 6 bits/sample to the spectrum between 50-4KHz, 2 bits for 4K to 7KHz. Result can be transmitted in 32Kbps.

Higher quality audio codecs: MP3 (actually the audio of MPEG-1, level 3), AC-2 (for theaters), AC-3 (for HDTV)…

Experimental observations of data density (stereo voice + music):
CD-audio – 650MD (one CD) = 60 min ==> 10-12 MB/minute
MP3 compression ==> 1 MB/minute

Coding Techniques
(simplistic view)

If you can't see it or hear it, don't send it
Examples: YUV → (4,2,0), G.722 for audio

Runlength encoding for adjacent degeneracies – seen best in facsimile transmission of documents.
"Motion compensation" in the MPEG codes exploits space degeneracy and constancy of the image in time.

Move to appropriate space or representation (usually Fourier, increasingly wavelets) and use the shortest codewords for the most common symbols
Morse
Huffman

Subband encoding
Permits progressive transmission, as well as sequential
Framework in which you can throw away the least perceivable bits (usually high frequencies).

Important note: the fancier the code, especially codes involving transforms, filtering, and inverse transforms in decoding, the longer the delay incurred. Telephony codes usally < 10ms, but MP3 and MPEGs can be close to 100 ms round trip – a problem if used in interactive applications (conversations).

Video Standards

Proprietary:

Indeo and earlier RCA/SRI efforts for CD-interactive

RealNetworks audio and video – for streaming on the Web.

MS .asf formats for the Web

Open:

JPEG for images (based on 8x8 pixel blocks), use of DCT, and discarding high spatial frequencies to meet a bit rate budget.

MPEG 1,2,(4 or DivX) are founded on JPEG plus temporal reduncancy.

H.261 – a predecessor, used for videoconferencing

MPEG-1 (1992) 360 x 240 (NTSC@30Hz)
360 x 288 (PAL,SECAM@25Hz)
coded for CD-ROM, at max of 1.5Mbps (0.2 MB)
360 x 240 for luminance, but only 180 x 120 for chroma
motion comp – 16x16 blocks, search over 31x31 square

MPEG-2 (1996-8) for digital TV (esp. satellite)
Coded for max of 5 Mbps, 640 x 480 (and other levels)

MPEG-4 (just starting with the DivX and MS codecs, which are "naïve video," simple level, and non-conforming): objective is quality video at streamable bandwidths. Syntax permits rendering objects, but at present only rectangles supported. And DivX renders 320 x 240 pixels at present.

Video Standards in Practice

Even “open” isn’t free -- $6/copy for MPEG license
    (But cf. $90/copy for the Windows license on every PC)

What’s a GOP? MPEGs use
    I frames (full key frames, no prediction)
    P frames, differences from preceding frame after motion comp.
    B frames, after motion comp, interpolate between two frames.
    (the difference signals are encoded using JPEG)
        P frames are ~30% of I frames in data volume
        B frames are only ~12% of I frames
    And the sequence consists of Groups Of Pictures, such as:
        B (I B B P B B P B B P B B P B B P) I B …. (used in DVD)

Note that frames are not sent in presentation order, because B frames can’t be decoded until the following P or I frame is available. This means a lot more buffer is needed (and adds latency).

Video Practical Issues, ctd.

Observed data rates (from movies on my laptop):

MPEG-1	0.22-.24 MB/sec
MPEG-2	0.3 – 0.5 MB/sec
Div-X ('01)	0.04-0.06 MB/sec

(factor out the 4X increase in picture size for MPEG-2, and this is just 2X better encoding in each generation – another Moore's Law!)

Compute requirements:

JPEG, MPEG-1 encode can be done on a DSP (and is, in many still and video cameras) in real time

MPEG-2 decode is possible on a Pentium-I (no MMX) 233 MHz

DivX decode takes P-III (with MMX), 450 MHz

(note the value of SIMD in MMX and Mac's G4 PowerPC, for dealing with all the 8-bit data in the image or video stream)

In general, encode takes 3-5x as long as encode, and data transport times may be more important than compute delays. Thus to date, real time MPEG-2 encode with any quality requires special hardware, and software DivX encode is 4-8x real time, or typically overnight for a single movie.

Update -- newer software, e.g. PowerDirector, can now handle MPEG encoding. I have seen the following cpu requirements quoted:

MPEG-1 encode	PII (450 MHz)
MPEG-2 encode	PIII(650 MHz)
DV to MPEG-2 transcode	PIII(1GHz)
DivX encode in near real-time	PIV (1.5GHz)