

# Multimedia Computing

Keith Cheverst

## Additional (Background) Reading...

- "Understanding Networked Multimedia: applications and technology" by Francois Fluckiger, Prentice Hall, 1995.
  - ISBN 0131909924
- "Digital Multimedia" by Nigel Chapman and Jenny Chapman, Wiley 2000.
  - ISBN 0471983861
- "Multimedia Communications" by Fred Halsall, Addison-Wesley, 2001.
  - ISBN 0201398184

## Aims of the lecture...

- Provide an awareness of the context of the emergence of multimedia computing and the factors that have driven its development
- Cover the nature and representation of the various media types employed in multimedia computing
  - classification of media types
  - basic encoding types
  - benefits and drawbacks of compression
  - resource requirements

## A Definition of digital MM

- Digital multimedia is the field concerned with the computer-controlled integration of text, graphics, still and moving images, animation, sounds, and any other medium where every type of information can be represented, stored, transmitted, and processed digitally.  
*[Fluckiger, 95]*

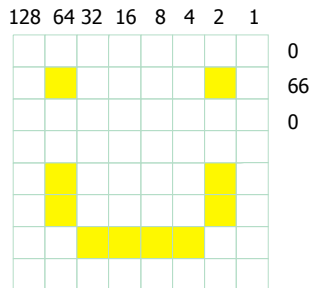
## Characteristics of MM systems

- Four main characteristics:
  - Info handled must be represented *digitally*
  - *Computer controlled*
  - *Integrated*
  - User interface may permit *interactivity*

## Discrete Media: Bitmapped Images

- Encoded as a matrix of dots (pixels)
- Explicit resolution and aspect ratio
- Used for range of purposes
  - Screens, printers, digital cameras...
- Tolerances range from
  - 50dpi – screen resolution
  - 600dpi – publishable quality

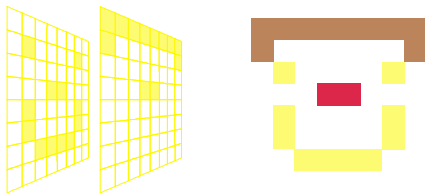
## Bitmap Example



## Colour bitmaps...

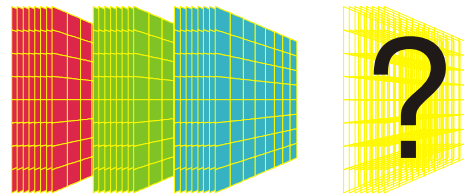
- Colour encoding
  - Need multiple bits per pixel, or bit-planes
- Either paletted
  - Combinations of bits differentiate colours
  - Colour lookup table or palette
- Or true-colour
  - 24 bitplanes upwards
  - Encoded as RGB components

## Paletted bitmaps



- n-bit depth
- $2^n$  colour combinations (e.g.  $2^2 = 4$ )

## True colour bitmaps



- Red/green/blue components
- Special effects bits

## Storage Requirements

- Volume = Width x Height x Depth
- Volume in bits, so Bytes = Volume / 8...
- Monochrome 160x120 = 2.4Kbyte
- VGA 640x480x8 (paletted) = 300Kbyte
- XGA 1024x768x24 = 2.25Mbyte
- A4 Print = 4800x6600x24 = 90.6Mbyte

## Graphics

- Computer generated images
- Two or Three dimensional
- Encoded using **vectors**...
- Differ from images (bitmaps)
  - Scalable – render at any resolution
  - revisable

## Graphics ctd.

- Storage overheads vary
  - Coarse images small (KBytes)
  - Detailed images huge (GBytes)
- Rendering overhead
  - Xmen window scene
  - 26 hrs / frame – approx 1 month / second
  - Real-time – e.g. MTV/Computer games

## Continuous Media

- All media so far have been **discrete**
- **Continuous media** has temporal properties
  - Constantly changing with time
  - Look at audio for this – more intuitive
- Continuous media **ANALOGUE**
  - need digital representation
  - how do we encode?

## Digitizing Continuous Media

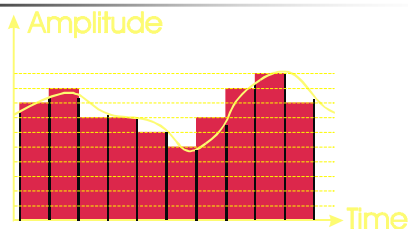
- Four steps
  - Sampling
  - Quantization
  - Codeword Generation
  - Compression
- Parameters and techniques chosen to avoid distortion

## Sampling



- **Periodic** reading of amplitude.
- Splits time into **discrete intervals**

## Quantization



- Splits amplitude into discrete levels
- More levels => better reproduction

## Distortion

- Clipping
  - Too few quantization levels
- Band-limiting
  - Sampling too infrequent

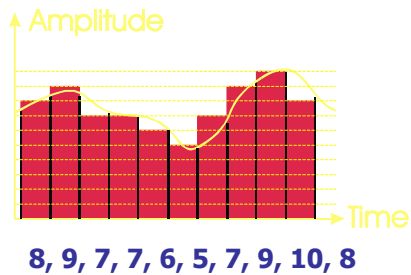
## Tolerances

- Range of frequencies important
  - Give quality of sound
  - Constrained by either sensor, media or channel
- E.g. human ear 50Hz – 22KHz
  - Music uses whole range
  - Speech only requires 11KHz
  - CD audio sampled at 44KHz ... why?

## Code word Generation

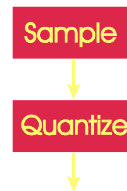
- Linear encoding
  - Microsoft WAV / CD audio (PCM)
- Differential encoding
  - DPCM – as used by telephones
  - Records change in amplitude between samples
  - Better for voice than music... why?

## Example – Audio Clip



## Without compression steps are...

- Linear Encoding



- Examples: CD-audio (PCM), WAV, BMP

## Compression

- As we've seen continuous media is BIG
  - Take advantage of redundant information
- Two basic categories
  - Lossless compression
    - Run-length encoding
  - Lossy compression
    - Examples later...

## Run-length encoding

- Optimises out consecutive patterns of symbols in the data stream
- For example:  
**0,7,3,4,4,4,4,4,4,5,6,2,2,2,2,0**  
Could be encoded as:  
**0,7,3,4!6,5,6,2!4,0**

Saving 4 bytes.

## Video Compression & Standards...

## Background Reading...

- "Understanding Networked Multimedia: applications and technology" by Francois Fluckiger, Prentice Hall, 1995. ISBN 0131909924
  - \*\*\*\* Pages 536-552 + Pages 498-505 \*\*\*\*
- "Digital Multimedia" by Nigel Chapman and Jenny Chapman, Wiley 2000. ISBN 0471983861
  - Pages 300-325
- "Multimedia Communications" by Fred Halsall, Addison-Wesley, 2001. ISBN 0201398184
  - Pages 193-223

## Intro: Why we need video compression...

- Motivation for video compression...
  - Making video a manageable size for both:
    - Transmission
    - Storage

## An example...

- Medium quality video (no audio)
  - 640x480 pixels, 8 bit colour (RGB), 25 fps, 90 mins
- Consider transmission...
  - Bit rate =  $\text{frame size} * \text{bits per sample} * \text{no. of components} * \text{frame rate}$ 

$$= (640 \times 480) * 8 * 3 * 25$$

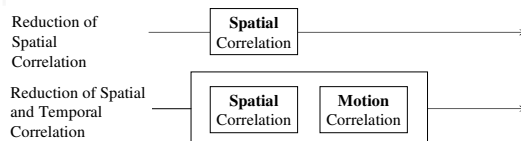
$$= 184 \text{ Mbit/s}$$
- Consider storage...
  - Storage =  $\text{bytes per second} * \text{no. seconds}$ 

$$= (184 \text{ Mbit/s} / 8) * 60 * 90$$

$$= 124,200 \text{ Mbytes (124.2 Gbytes)}$$

## Compressing moving images...

- Two possible approaches...
  - Spatial correlation
  - Temporal correlation
    - Only difference between frames is encoded
    - *Motion compensation*



## Spatial correlation...

- Concerned with individual images
- Two approaches (complementary)
  - cf JPEG compression for still images
    - Sample
    - Quantize
    - Frequency transform etc.
  - Also have sub-sampling based on Luminance and Chrominance signals...

## Luminance and Chrominance signals...

- A colour pixel is made up of:
  - 3 colour components., RGB
  - Or, luminance (Y) and 2 chrominance signals (UV)
  - Chrominance signals calculated from colour difference...
  - In practice only two colour difference signal are required, i.e.  $C_{d1}$  and  $C_{d2}$ .
- However...
  - the human eye more is sensitive to brightness than to colour...

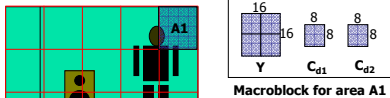
[http://www.pcmag.com/encyclopedia\\_term/0,2542,t=YCbCr+sampling&i=55148,00.asp](http://www.pcmag.com/encyclopedia_term/0,2542,t=YCbCr+sampling&i=55148,00.asp)

## Spatial correlation: Sub-sampling...

- Motivation...
  - Economise the bandwidth and storage capacity.
  - Approach...
- Subsample the colour components...
  - Give greater resolution/accuracy to luminance (Y) over colour differences  $C_{d1}$  and  $C_{d2}$
  - Fewer samples per line, even fewer lines per frame
  - Standard notation for sampling ratios
    - Y sampling frq:  $C_{d1}$  sampling frq:  $C_{d2}$  sampling frq

## Example...

- Consider MacroBlocks of 16x16 pixel area
  - Using 4:1:1 subsampling a MB would comprise:
    - 4 8x8 pixel blocks for luminance (Y)
    - 1 8x8 pixel block for (Colour difference 1)  $C_{d1}$
    - 1 8x8 pixel block for (Colour difference 2)  $C_{d2}$



- Resulting bitstream is 50% of original

## Confusion...

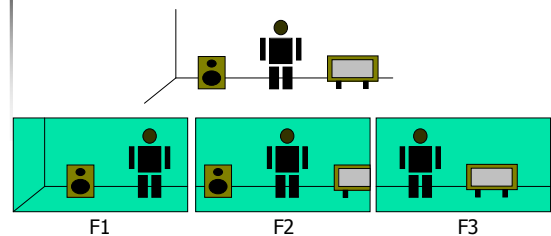


- This is wrongly labelled!
- These are analog outputs so should be labelled  $P_b$  and not  $C_b$  etc.

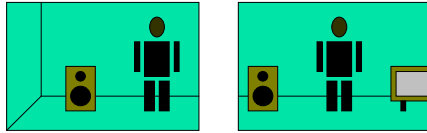
## Temporal Correlation...

- Make use of similarities of frames...
- Employs DPCM techniques...
  - Only the *difference* between frames is encoded
  - Process often termed *motion compensation*...

## A Simplified Example...



## Example...

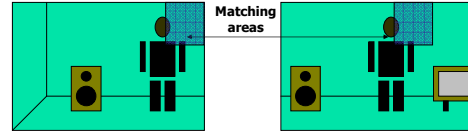


F1

F2

- Reference and *Intracoded* (I) frames...
- F2 can be approximated by pieces of the F1 frame
- F1 acts as Reference frame

## Motion Vectors...

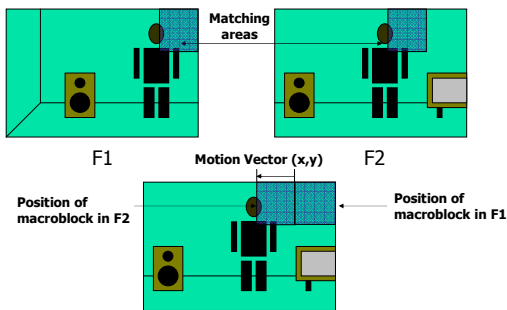


F1

F2

- Algorithm searches for *Best Matching MacroBlock*
- Needs to calculate *error term* (also MacroBlock)
- Needs to capture/convey *spatial translation*
  - *Motion Vector*

## Translation of Macro Block...



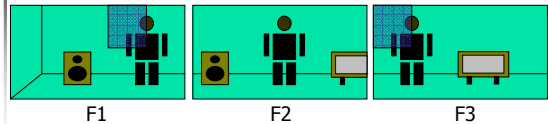
F1

F2

Position of macroblock in F2

Position of macroblock in F1

## Predicted and Bidirectional frames (1)...



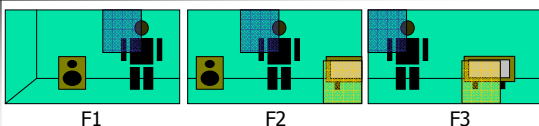
F1

F2

F3

- Consider F3 (ignoring F2 for now...)
  - Has macroblocks in common with F1
  - Could be constructed from F1
  - F3 would then be a *Predicted* (P) frame

## Predicted and Bidirectional frames (2)...



F1

F2

F3

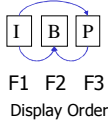
- Consider F2
  - Has macroblocks in common with F1 and F3
  - Could be constructed using pieces of F1 and F3
  - F2 is called a *Bidirectional* (B) frame
    - Note: Both F1 and F3 acting as Reference frames!

## Question?

- How can we know at the time F2 is coded that there will be a matching macroblock in F3?
- Answer:
  - F3 needs to be available for reference at the time F2 is encoded
  - i.e. F1, F2 and F3 would need to be buffered
  - F2 only sent (transmission order) once it has been interpolated from F1 and F3

## Summary (from example)...

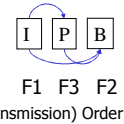
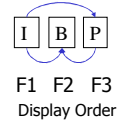
- F1 is an I frame – it is encoded without reference to any other frame
- F3 is a P frame – it is predicted from a reference frame: in this case F1
- F2 is a B frame – it is interpolated from F1 and F3



## Bitstream order...

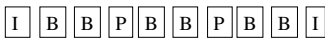
- What about the decoder...

- How to handle B frames
- Needs info from later (I or P) frames in order to reconstruct the B frame
- Solution: reorder the sequence
- Display order -> Bitstream order

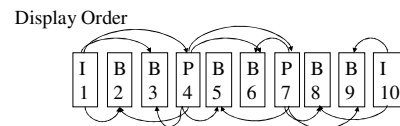


## GOPS...

- Encoders typically use a repeating sequence of I,P and B frames
- This is known as a GOP (Group Of Pictures)
  - Always begin with an I frame
  - Common sequences
    - IBBBBBBI, or IBBPBBPBI
    - N=9

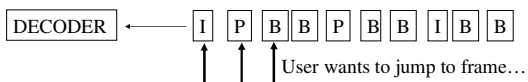


## Sequences of I,P and B frames



Bit stream (transmission) Order

## Role of I frames...



- You want to resume from a given frame...
  - What if frame is a I frame?
  - What if frame is a P frame?
  - What if frame is a B frame?
- I frames act as synchronisation points
  - Delay between occurrence of successive I frames should not exceed 400 ms

## Implications on the network...

- Original example showed a transmission over the network of 184 Mbit/s
  - Constant bit rate stream
- However, different parts of video have greater redundancy than others
- Implication...
  - Variable bit rate stream

## MPEG standards...

## MPEG standards...

- ISO
  - International Standards Organisation
  - Motion Pictures Expert Group
- Each standard composed of 3 parts:
  - MPEG-Audio
  - MPEG-Video
  - MPEG-System
    - How audio and video bitstreams should be multiplexed
    - Achieving synchronisation between the two streams

## ISO MPEG-1...

## Intro to MPEG-1

- Intended for video playback from CD-ROM
  - Assumes transmission rate of 1.5 Mbps (1993/4)
  - Audio channel requires 200-250 Kbps for CD quality
  - Therefore bit rate of video shouldn't exceed 1.2 Mbps
  - Targets VCR quality
- Defines a data-stream syntax and a decompressor
  - Manufacturers are free to develop different compressors

## MPEG-1: Image Format

- Optimised for Standard Interchange Format (SIF)
  - 352\*240 pixels NTSC (30 fps)
  - 352\*288 pixels PAL/SECAM (25 fps)
    - Note same resulting bit rate per second
- Storage:
  - 650 Mbyte CD-ROM can store approx 40 mins video at 352x288 at 25 fps.

## MPEG-1 Video Coding

- DPCM technique based on Macroblocks
  - Computationally expensive...
  - Only decoding process required to be real/time
- 4:1:1 subsampling
- Utilises I,P and B frames
- *GOPs*
  - Each GOP must contain at least 1 I-frame

## Contrast with ITU H.261

- Intended application domains
  - videoconferencing, video telephony
- Transmission requirements
  - video conferencing over ISDN (multiples of 64 kbit/s)
- Coding/decoding requirements
  - Both encoding/decoding must be carried out in real-time
  - So uses only I and P frames
- Supported picture formats
  - CIF (352x288)
  - QCIF (176x144)

## ISO MPEG 2...

## MPEG-2 Intro...

- Motivation...
  - Provide different qualities of image for range of different app domains (with differing target bit rates)
    - e.g. Studio-quality motion video
  - MPEG-2 took on the mantle of MPEG-3
    - Encoding and compression of HDTV

## Profiles and Levels...

- MPEG-2 supports greater choice of bit rate
  - up to HDTV picture size and resolution
- Allows greater chrominance resolution
  - E.g. 4:2:2, 4:4:4
- Support for wider range of apps
  - Family of compression schemes
  - Schemes defined by a profile and a level
- 5 Profiles
  - Specifies details of compression algorithm
  - High, Main, Simple, Spatial Resolution etc.

## Levels...

Level	Frame Size	Target BitRate
High	1920*1152 samples/frame (HDTV)	80/100 Mbps *
High-1440	1440*1152 samples/frame (consumer-HDTV)	60/80 Mbps *
Main	720*576 samples/frame (studio quality TV)	15/20 Mbps *
Low	352*288 samples/frame (VCR) – compatible MPEG-1	4 Mbps

\* Dependent on subsampling used

## Further info...

- ISO MPEG standards for video compression
  - [www.mpeg.org](http://www.mpeg.org)
  - [www.chiariglione.org/mpeg/](http://www.chiariglione.org/mpeg/)
- Also...
  - [MPEG-4](#),
    - the standard for multimedia for the fixed and mobile web,
  - [MPEG-7](#),
    - the standard for description and search of audio and visual content
  - [MPEG-21](#),
    - the Multimedia Framework.
- [http://www.pixeltools.com/h264\\_paper.html](http://www.pixeltools.com/h264_paper.html)
- Try out...
  - [www.flaskmpeg.net](http://www.flaskmpeg.net)
  - [www.avs4you.com](http://www.avs4you.com)

## What does the iPod Use?

- Uses H.264 standard..
- Ratified as part of MPEG-4 standard (MPEG-4 Part 10)
- Gives good results across a broad range of bandwidths,
  - Mobile Content 176x144, 10-15 fps 50-60 Kbps
  - Std Def 640x480, 24 fps 1-2 Mbps
  - High Def 1280x720, 24p 5-6 Mbps
  - Full High Def 1920x1080, 24p 7-8 Mbps
- Delivers at a third to half the data rate for equivalent quality coded using MPEG-2
- [http://images.apple.com/quicktime/pdf/H264\\_Technology\\_Brief.pdf](http://images.apple.com/quicktime/pdf/H264_Technology_Brief.pdf)

## Multimedia Applications...

## Types of App...

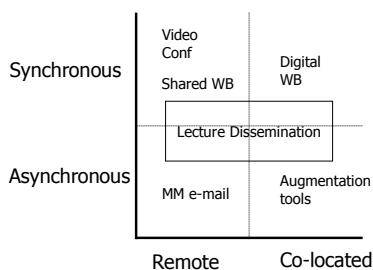
- Taxonomy of distributed MM Apps...
  - People-to-systems
    - Access, receive or interact with MM
    - Two broad categories
      - Distribution apps, Interactive apps
  - People-to-people
    - Improve communication
    - Interpersonal vs. group oriented
    - Synchronous vs. asynchronous (e.g. MM e-mail)

## CSCW

- Computer Supported Cooperative Work
  - Group communications (multimedia & remote)
  - Groupware...

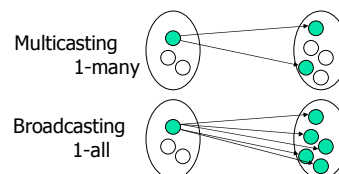
"groupware be viewed as the class of applications, for small groups and for organisations, arising from the merging of computers and large information bases and communications technology. These applications may or may not specifically support cooperation." - Ellis,91

## Time-space Matrix...



## (P to S) Distribution apps...

- Server initiated Audio-video distribution...
- Disseminate audio/video to...
  - Multiple individuals, Multiple locations
  - Open or Closed groups



## (P-to-S) Interactive Apps...

- User initiated
  - Retrieval
    - Access, locate and display MM info
- Two classes
  - Single request apps
    - cf downloading a MM doc
  - Multiple interaction apps
    - Interactivity continues during the session
    - Customisation of presentation
    - cf play, pause, ff etc.

## Multimodal Applications...

- Often retrieval based
- Emphasis on usability – providing a natural interface
- Input
  - gesture recognition,
  - voice recognition etc.
- Output
  - speech output
  - avatars (e.g. MPEG-4)

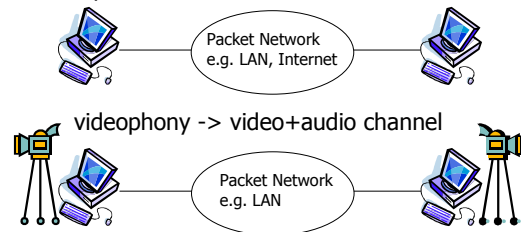


## People-to-People MM Apps

- Synchronous...
  - Interpersonal apps
  - Distribution apps
    - Person-to-group (1-N)
    - One-way
  - Collaborative Virtual Environments
  - Group 'Teleconferencing' apps
    - Bidirectional
    - (N-N)
      - Audio, audio-video, shared workspace,

## Audio-video Interpersonal Apps... Packet telephony & videophony

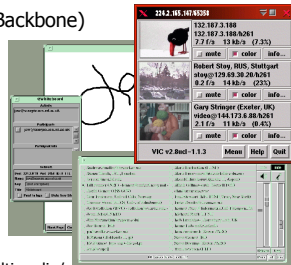
- The underlying network is not the telephone...
- If IP packet network -> VoIP



## Long-haul packet tele/videophony

### MBONE tools (IP Multicast Backbone)

- VIC
  - Video conferencing tool
  - Adaptive operation
  - H.261/H.263/mjpeg.
  - Voice activated switching
- VAT
  - Audio conferencing tool
- Also
  - Wb, Sdr



<http://www-mice.cs.ucl.ac.uk/multimedia/>  
<http://research.microsoft.com/barc/mbone/mbonetools.htm>