Image & Video Capture

- An image is captured when a camera scans a scene
  - Colour => Red (R), Green (G) and Blue (B) array of digital samples
  - Density of samples (pixels) gives resolution
- A video is captured when a camera scans a scene at multiple time instants
- Each sample is called a frame giving rise to a frame rate (frames/sec) measured in Hz
  - TV (full motion video) is 25Hz
  - Mobile video telephony is 8-15 Hz … jerky

Image Data (RGB)

- Colour still image:
  - 420 x 315 pixels, 8 bits/pixel = 387KB

Red Green Blue
8 bits: 0-255

(R,G,B) = (153,102,204)
(R,G,B) = (17,0,0)
(R,G,B) = (204,153,205)
**Video Technology: generating a colour**

- Frame buffer (2D array of 24 bit values)
- Colour guns
- Phosphor dots on display

<128, 128, 255>

**RGB value**

8 bits per colour

What you see

---

**Human Visual Perception**

- Mixing three primary colours in varying proportions, the perception of different colours can be created.

- Human eye build up of:
  - Cones to perceive colour
  - By exciting retina using different intensities of the three primary colours, the same colour may be perceived by the brain even if its unique wavelength is not present.

---

**Human Information processing**

- Identical colour combinations can cause different colour sensation under different conditions.
- Likewise two different colour can be perceived identical ...

- the human eye & brain
  - Interpolation
  - Pictures and events that can still be identified as separate
  - Colour interaction in the brain

- Adaptation
  - General-brightness adaptation
  - Lateral adaptation
  - Chromatic adaptation

---

**Colour**

- Colour is a visual feature which is immediately perceived.

- Salient chromatic properties are captured.

- Presence and distributions of colours induce sensations and conveys meanings in the observer according to specific rules.

- Representing colour on digital images and reproducing accurately on output devices are not at all straightforward.

- Distances in colour space should correspond to human perceptual distance.
Colour Space

- To deal with colour we need to quantify it in some way
  - gives us the notion of colour space or domain

- Hierarchy of colour sets
  - Perceivable by human beings
  - Displayed on a monitor screen
  - Calculated and stored in a frame memory

Representation of Colour Stimuli

- Points in three dimensional space
  - Perceivable by human beings
  - Displayed on a monitor screen
  - Calculated and stored in a frame memory

- Colorimetric models
  - CIE Chromaticity diagram

- Physiologically inspired models
  - CIE XYZ, RGB

- Psychological models
  - HSV, HW

- Hardware-oriented models
  - RGB, CMY, YIQ

- User-oriented models
  - HLS, HSV, HSB

Video Technology:
representing colour

- Colour models: RGB
  - RGB = Red Green Blue
  - directly modelled in device (i.e., corresponds to colour guns in display)
  - easy to implement

- Not based on visual (perceived) colours
- Not perceptually uniform
Video Technology: Colour Models: RGB Colour Space

- Cyan
- White
- Black
- Red
- Magenta
- Blue
- Green
- Yellow

Video Technology: Colour Models: RGB Colour Space

- Blue
  - (0,0,1)
- Cyan
  - (0,1,1)
- Red
  - (1,0,0)
- Yellow
  - (1,1,0)
- Green
  - (0,1,0)
- White
  - (1,1,1)
- Black
  - (0,0,0)
- Magenta
  - (1,0,1)

Video Technology: Colour Models: RGB

- Colour is labeled as a relative weights of three primary colours, in an additive system using the primaries Red, Green, Blue.
- It is perceptually non-linear space.
  - Equal distances in the space do not necessarily correspond to perceptually equal sensation.
- Non-linear relationship between RGB values & the intensity produced in each phosphor dot, low intensity values produce small changes in response to screen.
- It is not a good colour description system.

Video Technology: Colour Models: HSV

- HSV = hue, saturation, value (intensity)
- "painter's model"
- better model for representing colours as we see them ("I want a bright highly saturated apple green.")
- can be converted to/from RGB
- like RGB, axes not perceptually uniform
- variant: HLS (hue, lightness, saturation)
**Video Technology: Colour Models: HSV Colour Space**

- Non-linear transformation of RGB cube
- Hue: quality by which we distinguish one family from others
- Chroma: quality by which we distinguish a strong colour from weak ones
- Value: It is that quality by which we distinguish a light colour from a dark one
- H corresponds to selecting a colour; S corresponds to selecting the amount of white; selecting V corresponds to adding black
- Perceptually non-linear
  - Perceptual in the sense that we are using attributes that we normally think of
  - Attributes are not independent
- Variant: HLS (hue, lightness, saturation)

**Video Technology: Colour Models: YUV**

- Colour model used for TV signal transmission
- Y represents luminance (intensity of monochrome signal)
- U,V carry separate colour information (colour difference values)
- \[ Y = 0.2125R + 0.7154G + 0.0721B \]
- \[ U = B - Y, \ V = R - Y \]
- Typically, Y contributes most to signal bandwidth

**Image Data (YUV)**

- See:
Video Technology: CIE Colour Specification System

- Commission Internationale d’Éclairage
- colour labelling system
- “XYZ” space
- international standard (1931)
- based on colour matching functions determined by experiments with human subjects
- gives uniform colour spaces
- needs transformation into one of the other models

Video Technology: Colour Models: CMYK

- CMYK = cyan, magenta, yellow, black
- "printer’s model"
- a subtractive model
- set of practically available CMYK colours (“process colours”) are not equivalent to RGB set

Image & Video Capture

- Red, Green, Blue
- 8 bits: 0-255
- Y (luminance)
- U, V
- Time
- 0(black), ..., 255(white)

Video Sequence

- Consists of number of frames
  - Images produced by digitising time-varying signal generated by the sensors in a camera
  - Bit-mapped images
- Camera
  - Circuitry Inside a Camera
  - Purely digital signal (data stream) is fed into a computer via a high speed interface
  - IEEE 1394 (FireWire)
- Computer
  - Broadcast video is fed into a video capture card attached to the computer
  - Video capture card - analogue signal is converted into a digital form
Video Data

- **Desktop PC**
  - CIF (352 x 288), 8 bpp, 30hz = 8.7 MB/sec
  - 30 sec clip = 261 MB

- **Video to mobile device**
  - QCIF (176 x 144), 8 bpp, 30hz = 2.2 MB/sec
  - 30 sec clip = 65 MB

- **High Definition TV (HDTV)**
  - 1280 x 720, 24 bpp, 50hz = 0.4 GB/sec
  - 2.5 hour movie = 3.4 TB

Pushing the hardware

- **Consumers expectations are based on broadcast television**
- **Consumer equipment plays back at reduced frame rate resulting in jittery- dropped frames**
- **In order to accommodate low-end PCs considerable compromises over quality must be made**

Persistence of vision

- **If a sequence of still images is presented to our eyes at sufficiently high rate (frame rate~40 fps), we experience a continuous visual sensation rather than perceiving individual images**
  - A lag in the eye's response to visual stimuli which results in after images

- **If the consecutive images only differ by a small amount, any changes from one to next will be perceived as movement of elements within images**

- **Film projector displays an image twice (24 fps becomes 48 fps)**

Human Perception

- **What frame rate perceived as smooth?**
  - No identification of single frames if refresh frequency is high enough
  - Perception of 16 frames/s as continuous sequence
  - Depends on material

- **More sensitive to low frequencies**
- **More sensitive to changes in luminance and blue-orange axis**
- **Vision emphasizes edge detection**
Digitization: camera vs computer

- **Advantage**
  - Analogue signal transmitted on a cable get corrupted by noise
  - Noise will creep in if analogue data is stored on a magnetic tape
  - Camera is resistant to corruption by noise and interference

- **Disadvantage**
  - User has no control over digitization
  - Most conform to an appropriate standard

Image & Video Processing

- When processing image/video data we have two choices:
  - Raw data … termed uncompressed domain
    - Direct processing of the pixel values on either a global or local basis
    - Slow - more data, may require decode process
    - Possible to extract a wide range of expressive information from raw data
  - Encoded data … termed compressed domain
    - Fast - partial image reconstruction, real-time possible
    - Restricted to image/video data in bitstream
    - Compression is about throwing away information for efficient representation and transmission

Video Bit Rate Calculation

\[
\text{width} \times \text{height} \times \text{depth} \times \text{fps} \times \text{compression factor} = \text{bits/sec}
\]

- width ~ pixels (160, 320, 640, 720, 1280, 1920, …)
- height ~ pixels (120, 240, 480, 485, 720, 1080, …)
- depth ~ bits (1, 4, 8, 15, 16, 24, …)
- fps ~ frames per second (5, 15, 20, 24, 30, …)
- compression factor (1, 6, 24, …)

Examples

<table>
<thead>
<tr>
<th>Width</th>
<th>Height</th>
<th>Depth</th>
<th>fps</th>
<th>Comp</th>
<th>Kb/sec</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>120</td>
<td>5</td>
<td>15</td>
<td>25</td>
<td>92</td>
<td>Basic Rate ISDN</td>
</tr>
<tr>
<td>320</td>
<td>240</td>
<td>5</td>
<td>15</td>
<td>25</td>
<td>309</td>
<td></td>
</tr>
<tr>
<td>640</td>
<td>480</td>
<td>16</td>
<td>24</td>
<td>1232</td>
<td>MPEG1 (Primary Rate ISDN)</td>
<td></td>
</tr>
<tr>
<td>640</td>
<td>480</td>
<td>16</td>
<td>24</td>
<td>1232</td>
<td>MPEG1</td>
<td></td>
</tr>
<tr>
<td>640</td>
<td>480</td>
<td>24</td>
<td>30</td>
<td>36,864</td>
<td>MPEG2</td>
<td></td>
</tr>
<tr>
<td>640</td>
<td>480</td>
<td>24</td>
<td>30</td>
<td>221,184</td>
<td>Uncompressed</td>
<td></td>
</tr>
</tbody>
</table>
Video Data Size

<table>
<thead>
<tr>
<th>Image Size</th>
<th>1280x720</th>
<th>640x480</th>
<th>320x240</th>
<th>160x120</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sec</td>
<td>0.19</td>
<td>0.08</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>1 min</td>
<td>11.20</td>
<td>4.98</td>
<td>1.66</td>
<td>0.41</td>
</tr>
<tr>
<td>1 hour</td>
<td>671.85</td>
<td>298.60</td>
<td>99.53</td>
<td>24.88</td>
</tr>
<tr>
<td>1000 hours</td>
<td>871,846.40</td>
<td>298,598.40</td>
<td>99,532.80</td>
<td>24,883.20</td>
</tr>
</tbody>
</table>

Image size of video

- 1280x720 (1.77)
- 640x480 (1.33)
- 320x240
- 160x120

Compression

- When captured, audio/video data is referred to as "raw" or "uncompressed"
- In practice, undergo software/hardware process to compact data:
  - Termed "compression" or "encoding"
  - Results in an efficient bitstream that can be stored or transmitted
- Requires a (less) complex process to uncompress (decode) before it can be displayed
- A system for encoding & decoding is termed a "codec"

Image Compression

- Use frequency domain analysis
  - The discrete cosine transform (DCT)

- Use Pass Filter
  - Ignore low frequency info
  - Gain better compression

- Reuse inverse transform
  - Use inverse transform
  - Less perceptual losses

Image Compression

- Use more blocks
  - Less information to store or transmit over network
Effects of Compression

storage for 1 hour of compressed video in megabytes

<table>
<thead>
<tr>
<th>Resolution</th>
<th>1920x1080</th>
<th>1280x720</th>
<th>640x480</th>
<th>320x240</th>
<th>160x120</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1</td>
<td>671,846</td>
<td>298,598</td>
<td>99,533</td>
<td>24,883</td>
<td>6,221</td>
</tr>
<tr>
<td>3:1</td>
<td>223,949</td>
<td>99,533</td>
<td>33,178</td>
<td>8,294</td>
<td>2,074</td>
</tr>
<tr>
<td>6:1</td>
<td>111,974</td>
<td>49,766</td>
<td>16,589</td>
<td>4,147</td>
<td>1,037</td>
</tr>
<tr>
<td>25:1</td>
<td>26,874</td>
<td>11,944</td>
<td>3,981</td>
<td>995</td>
<td>249</td>
</tr>
<tr>
<td>100:1</td>
<td>6,718</td>
<td>2,986</td>
<td>995</td>
<td>249</td>
<td>62</td>
</tr>
</tbody>
</table>

3 bytes/pixel, 30 frames/sec

Compression

- Two types:
  - Lossless: doesn’t change data “simply” reorganizes
    - Used in medical applications (e.g. X-Rays) and document scanning (e.g. FAX)
  - Lossy: throws some data away during encoding
    - Used in most multimedia applications
- Popular image/video compression standards for multimedia applications:
  - JPEG (still images)
  - JPEG 2000 (enhanced functionality/quality)
  - MPEG-1 (video from CD-ROM)
  - MPEG-2 (Digital TV, DVD)
  - MPEG-4 (mobile and content-based functionality)
  - Also: ITU-T real-time telecommunications standards e.g. H.261, H.263, H.264/MPEG-4 AVC

Video codecs

- Video capture boards
  - Digitization and compression
  - Decompression and digital to analogue transformation
  - Devices compressor/decompressor (codecs)
- Hardware codecs
  - Store them on a computer
  - Then play them back to an external video monitor (TV set) attached to the VCC
  - Most hardware codecs can not provide full motion video to monitor
  - We can not know our audience will have any hardware codec available
- Software codec
  - Program that performs the same operation

What is MPEG

- MPEG: Moving Picture Experts Group (Created in 1988)
- ISO (Int. Standards Organization) / IEC (Int. Electro-technical Commission)
  - ISO/IEC JTC 1 / SC 29 / WG 11
- Develop standards for the coded representation of moving pictures and associated audio
**Video Technology:**

**MJPEG**

- motion JPEG
- just applies JPEG to each frame
  - YCbCr
  - apply to each channel
- used for compression during video capture
- compression ratios of 7:1
- no temporal compression
- allows users to set quality parameters
- not a standard
  - MJPEG-A

**Vector Quantization**

- Iterative algorithm
  - Pick set of reference blocks (code book)
  - Code picture blocks by code book entries
  - Entropy/RLE code the code symbols
- How to select code book
  - Step 1: pick reference blocks
  - Step 2: compare reconstructed image to original
  - Step 3: add additional reference blocks
  - REPEAT UNTIL ERROR IS SMALL
- Slow encode, fast decode

**MPEG Standards**

- **MPEG-1:** Storage of moving picture and audio on storage media (CD-ROM) 11 / 1992
  - aimed at low bit-rates of 1.5 Mb/s
  - typical of CD-ROM
- **MPEG-2:** Digital television 11 / 1994
  - aimed at bit rates of 8-15 Mb/s
  - DVD
- **MPEG-4:** Coding of natural and synthetic media objects for multimedia applications v1: 09 / 1998  
  v2: 11 / 1999
  - introduction of objects into the specification
  - wide range of data rates
  - important for multimedia
- **MPEG-7:** Multimedia content description for AV material

**Video Technology:**

**MPEG-1 compression approach**

- Spatial compression for individual frames
- based on JPEG-like technique
- temporal compression of sequences of frames
  - looks for areas of change
  - creates difference frames
- based on 16X16 macroblocks
**Temporal Compression**

- Make use of similarities of frames
  - Only difference between frames is encoded
  - Process often termed motion compensation

  ![Image of S1 and S2 frames](Image)

  Second one (S2) can be approximated by pieces of the first one (S1)
  - S1 acts as a reference frame

**Motion Vectors**

- Algorithm searches for Best matching Block
- Needs to calculate error term (Matching block)
- Needs to capture/convey spatial translation
  - Motion vector

  ![Image of Motion Vectors](Image)

**Predicted Frames**

- Consider S3
  - Has macroblocks in common with S1
  - Could be reconstructed from S1
  - S3 would be then a Predicted (P) frame

  ![Image of Predicted Frames](Image)

**Bidirectional frames**

- Consider S2
  - Has macroblocks in common with S1 and S3
  - Could be constructed using pieces of S1 and S3
  - S2 would be then a Bidirectional (B) frame
  - Both S1 and S3 acting as reference frames

  ![Image of Bidirectional Frames](Image)
Question?

- How can we know at the time S2 is coded that there will be a matching block in S3?
- Answer:
  - S3 needs to be available for reference at the time of F2 is coded
  - i.e., S1, S2, S3 would need to be buffered
  - S2 only sent (transmission order) once it has been interpolated from S1 and S3

Summary (from example)

- S1 is an I frame – it is encoded without reference to any other frame
- S3 is a P frame – it is predicted from a reference frame: in this case S1
- S2 is a B frame – it is interpolated from S1 and S3

Display Order

Bitstream order

- What about decoder ...
- How to handle B frames
  - Needs info from later I or P frames in order to construct B frame
- Display Order

- Solution: reorder the sequence

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 sequence (display Order)
    - IBBPBBBI or IBBPBPPBB
    - N=9
  - Bitstream order
    - IPBBBIIBBB or IPBBPBBBIIB
Video Sequence

- Commence with a sequence header
- Followed by n GOPS where n> 0
- End with a sequence_end_code
- GOP
  - Each GOP must contain at least I frame
  - Assist random access into the sequence
    - Therefore greater apps need for RA the shorter should be the size of GOP

Role of I frames

IPBBPBBIBBB

You want to resume from a given frame …

What if frame is I frame
P frame
B frame

I frames act as synchronisation points
Delay between occurrence of successive I frames should not exceed 400ms

Video Technology: MPEG
Frame Types: I Frames

- Intra-coded images
  - similar to a JPEG still of the frame
- Expensive but required
  - I-frames expensive as they have to compress the entire scene
  - needed as start frame for differences
  - needed for scene changes

Video Technology: MPEG
Frame Types: P Frames

- Predictive coded frames
  - based on predicting the movement of blocks from their position in the previous frame (I or P)
Video Technology: MPEG

Frame Types: B Frames

- Bi-directional frames
  - based on pair of I/P frames, before and after

Profiles and levels

- MPEG-2 supports greater choice of bit rate
  - Up to HDTV picture size and resolution
  - Allows greater chrominance resolution
    - 4:2:2; 4:4:4
  - Support for wider range of apps
    - Family of compression schemes
    - Schemes defined by a profile and level
      - No single encoder/decoder has to implement all functionality
      - Comparability between newer and older equipment
  - 5 Profiles
    - High, Main, Simple, Spatially scalable, SNR scalable, 4:2:2, multiview etc.

MPEG 2

- Motivation …
  - Provide different qualities if image for different domains (with differing target bit rates)
    - E.g., studio quality motion video
  - MPEG-2 took on the mantle of MPEG-3
    - Encoding and compression for HDTV
  - Standard for digital broadband TV
  - Interlaced video
  - DVD quality

MPEG-4

- Motivation …
  - Original objective: develop a low bit rate video compression method
  - Now a set of tools for interactive multimedia scene composition, multiplexing and synchronisation
    - Digital television
    - Interactive graphics application
    - Interactive multimedia
  - MPEG-4 provides
    - The standardised technological elements enabling the integration of production, distribution and content access paradigm of the fields of interactive multimedia, mobile multimedia,…