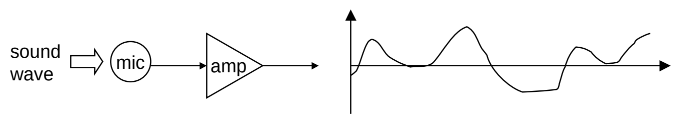
Distributed Video Systems
Chapter 2
Video Coding Technologies

Jack Yiu-bun Lee
Department of Information Engineering
The Chinese University of Hong Kong

Contents

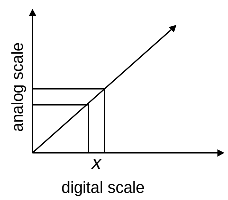
- 2.1 AV Signal Processing
- 2.2 Digitizing Audio
- 2.3 Digitizing Video
- 2.4 Digital Video Standards
- 2.5 Video Compression
- 2.6 MPEG Compression Standards

- Analog and Digital Signals
  - From the physical world, a *sensor* transforms the time-dependent or space-dependent physical variables into electrical signals.
  - For example: recording audio



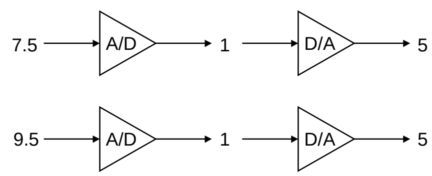
However, digital computer/systems cannot handle analog signals directly.

- Analog and Digital Signals
  - We need Analog-to-Digital (A/D) Conversion:



  - Use a *number* to represent a *range of values* in the analog scale.
  - For example, represent 5~10mV as 1, 10~15mV as 2, etc.

- Analog-to-Digital (A/D) Conversion
  - Sampling Accuracy
    - The A/D conversion process is also referred to *quantization*.
    - The problem is digital number covers a range of analog values, hence the mapping is not one-to-one.
    - For example, a 7.5mV input converted to a digital number of 1 is only an *approximation*.
    - Because another input of say 9.5mV will also be represented by a digital number of 1.



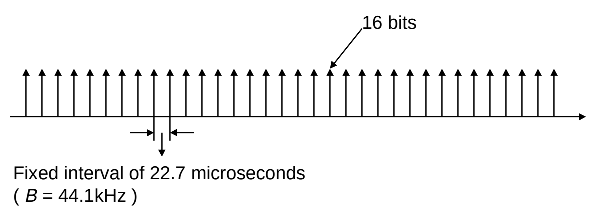
- Analog-to-Digital (A/D) Conversion
  - Sampling Accuracy
    - The amount of digital numbers used is called quantization level, and is usually measured in bits.
    - If *n* bits are used, then there are 2^n numbers or levels to represent *distinct* signal values.
    - For example:
      – CD-audio uses 16 bits for audio, hence there are a total of 2^{16} or 65536 levels.
    - A digital signal is usually represented as a binary *codeword*:
      – e.g. 01101001
      $= (0\times2^7)+(1\times2^6)+(1\times2^5)+(0\times2^4)+(1\times2^3)+(0\times2^2)+(0\times2^1)+(1\times2^0)$
      $= 0 + 64 + 32 + 0 + 8 + 0 + 0 + 1$
      $= 105$

2.1 AV Signal Processing

- Analog-to-Digital (A/D) Conversion
  - Sampling Rate
    - How often do we take samples of the analog signal and convert it to digital form?
    - For example:
      - If we take one sample every second, then the sampling rate is 1Hz.
      - CD audio uses a sampling rate of 44.1kHz.
    - How fast should I sample?
      - Nyquist in 1924 showed that if the sampling rate is *twice* the max. frequency of the signal, than no information will be lost.
      - Hence CD audio's 44.1kHz covers the entire range of human-audible frequencies (20~20kHz).

2.2 Digitizing Audio

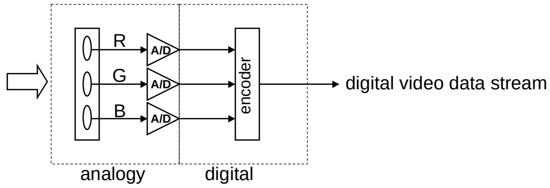
- Data Volume
  - Sampling frequency = *B* Hz
  - A/D precision = *L* bits
  - Data rate = *L* x B bits per second (or bps or b/s)
  - E.g. CD Audio



- Data rate R = 44.1 x 16 = 705.6 kbps (mono)
- Data rate R = 2 x 705.6 = 1411.2kbps (stereo)

Periodic digital signals are also called continuous media or isochronous media.

- Operation Model



- Primary Colors
  - Red-Green-Blue (RGB)
    - Red = 700 nanometers light wave
    - Green = 546 nanometers light wave
    - Blue = 436 nanometers light wave

- Analog Video Standards
  - NTSC (National Television Systems Committee)
    - In use in American, Canada, Japan, & Latin America, etc;
    - Signal Composition:

$Y = 0.30R + 0.59G + 0.14B$

$I = 0.74(R-Y)-0.27(B-Y) = 0.60R + 0.28G + 0.32B$

$Q = 0.48(R-Y)+0.41(B-Y) = 0.21R + 0.52G + 0.31B$

    - Interlaced scanning w/ 4:3 aspect ratio;
    - Resolution is 525 lines per frame at 29.97 frames per second (fps).

2.3 Digitizing Video

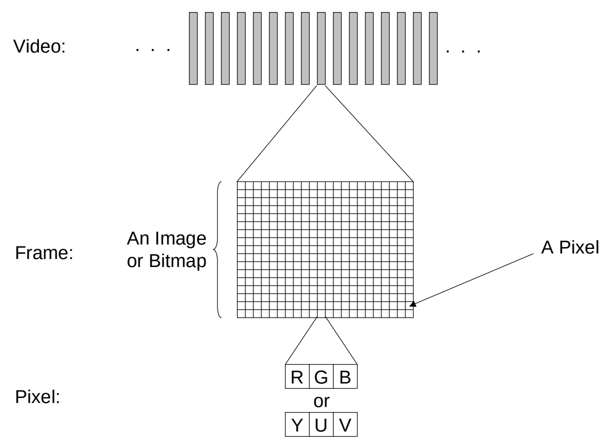
- Analog Video Standards
  - PAL (Phase Alteration Line)
    - In use in Hong Kong, Europe, Australia, etc;
    - Signal Composition:

      Y = 0.30R + 0.59G + 0.11B
      U = 0.493(B-Y) = -0.15R - 0.29G + 0.44B
      V = 0.877(R-Y) = 0.62R + 0.52G + 0.10B

    - Interlaced scanning w/ 4:3 aspect ratio;
    - Resolution is 625 lines per frame at 25 fps.
  - Others like SECAM, etc.

2.3 Digitizing Video

- Digitization Model



2.4 Digital Video Standards

- Studio-Quality TV (ITU-R 601)
  - Sampling Rate
    - Y(13.5Mhz), U (6.75Mhz), V (6.75Mhz);
  - Digitizing NTSC Video Signal
    - Raw data rate = (13.5+6.75+6,75) x 8 = **216 Mbps**.
    - Raw Pixel Resolution = 864 x 525 pixels (removing retrace ,etc.).
    - Active Video Area = 720 x 486 pixels.
    - Sub-sampling (4:2:2) (reduce bit-rate by 33%) Y (720x486), U (360x486), V (360x486)
    - 8-bits per sample per signal channel.
    - Net data rate after sub-sampling = **168 Mbps**.

2.4 Digital Video Standards

- Studio-Quality TV (ITU-R 601)
  - HDTV (US)
    - 720,000 pixels per frame
    - 24 bits per pixel
    - 60 fps
    - Data rate = **1.0368 Gbps**.(!)

2.4 Digital Video Standards

- Videoconferencing Quality - CIF
  - Common Interchange Format (CIF), (ITU-TS H.261)
  - Frame size (4:1:1 sub-sampling):
    - 352 x 288 for luminance (Y)
    - 176 x 144 for chrominances (U, V)
    - Data rate = **36 Mbps**.
- Videoconferencing Quality - QCIF
  - Quarter-Common Interchange Format (QCIF)
  - Frame size:
    - 176 x 144 for luminance (Y)
    - 176 x 144 for chrominances (U, V)
    - Data rate = **18 Mbps**.

2.4 Digital Video Standards

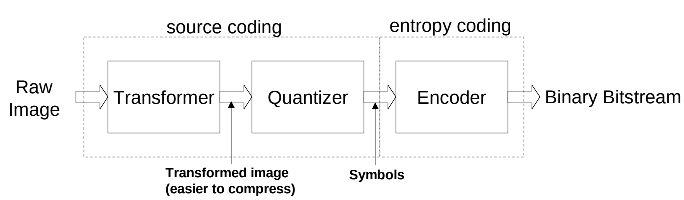
- Videoconferencing Quality - Super-CIF
  - Super-Common Interchange Format (Super-CIF)
  - Frame size (4:1:1 sub-sampling):
    - 704 x 576 for luminance (Y)
    - 352 x 288 for chrominances (U, V)
    - Data rate = **146 Mbps**.
- VCR Quality - SIF
  - Standard Interchange Format (Defined in MPEG-1)
  - Frame size (4:1:1 sub-sampling):
    - 352 x 240 (NTSC) or 352 x 288 (PAL/SECAM) for luminance (Y)
    - 176 x 120 or 144 for chrominances (U, V)

2.5 Video Compression

- Motivation
  - Digital audio and video generates vast amount of data that are difficult to process and deliver quickly.
- What is compression?
  - Reduce the number of bits used to encode the same information by exploiting:
    - Spatial redundancy
      - Correlation between neighboring pixels
    - Spectral redundancy
      - Correlation between color components
    - Psycho-visual redundancy
      - Perceptual properties of the human visual system

2.5 Video Compression

- Types of compression
  - Lossless compression
    - No information is loss in the encode/decode process.
  - Lossy compression
    - Some information is loss in the encode/decode process.
- A Generic Model for Compression:



2.5 Video Compression

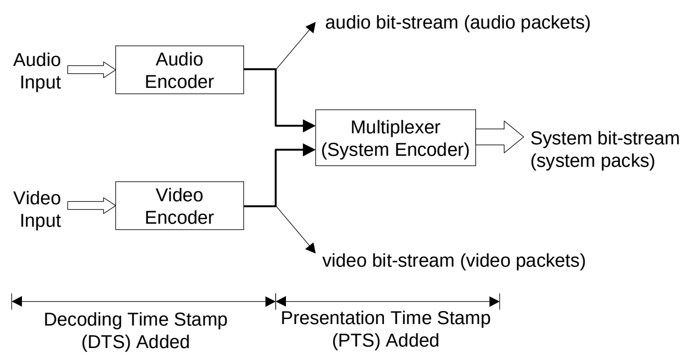
- A Generic Model for Compression
  - Transformer
    - A one-to-one mapping to transform the signal from the spatial domain to other domains, which are easier to compress.
    - Common transformers
      - Discrete Cosine Transform (DCT)
      - Wavelet Transform
  - Quantizer
    - A many-to-one mapping to reduce the data rate.
    - Loss in information is introduced in this stage.
  - Encoder
    - Maps symbols generated by Quantizer to bit-strings.
    - Exploits statistical knowledge to reduce bit rate.

2.5 Video Compression

- Two Types Compression
  - Constant Bit-Rate (CBR)
    - The bit-rate of the compressed video stream over a short time interval is constant.
    - The video quality is not constant. Loosely speaking, more motions degrade video quality.
    - CBR videos are good for system design but bad for the user.
  - Variable Bit-Rate (VBR)
    - The video quality is constant for the entire video stream.
    - The bit-rate is adjusted to maintain a constant video quality.
    - VBR videos are good for the user but bad for system design.

- MPEG Compression Standards
  - MPEG standards for Motion Picture Expert Group
    - It is a standard for video compression.
  - Composition
    - MPEG-1
      - VCR-quality video up to 8 Mbps;
      - Used in Video-CD, CD-I and Video-on-Demand systems.
    - MPEG-2
      - Broadcast quality video from 3 to >10 Mbps;
      - Used in DVD, HDTV, and Video-on-Demand systems.
    - MPEG-3
      - Originally slated for HDTV but later dropped due to the incorporation of HDTV into MPEG-2.
    - MPEG-4
      - Low-bit rate video for video telephony systems.

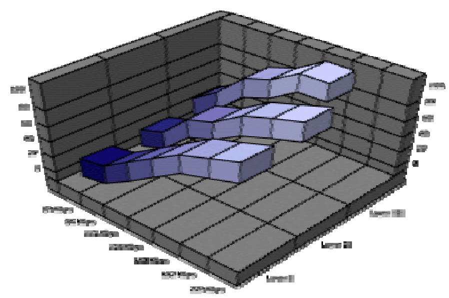
- MPEG System Structure
  - Encoding Process:



- MPEG System Structure
  - Bit-stream Structure
    - Audio and video are compressed and encoded individually into audio packets and video packets.
    - Decoding Time Stamps (DTS) are added to the packets to guide the decoder controller in the decoding process.
    - The audio and video packets are then multiplexed into a system stream by a system encoder (or multiplexer).
    - Presentation Time Stamps (PTS) are then added to *synchronize* the audio and video streams.

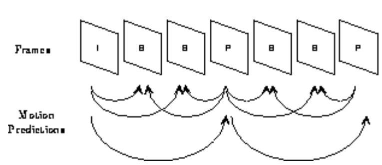
- MPEG System Structure
  - Audio Compression
    - How does it work?
      - MPEG Audio strips information in the audio signal that is less sensitive to the human perception system (ear).
      - This is called "perceptual coding".
    - MPEG Audio Layers
      - The Layer I psychoacoustic model only uses frequency masking.
        - This means that it strips frequencies that are hidden behind others. You shouldn't encode at higher compression than 384 Kbps.
      - Layer II does more filtering.
        - In layman's terms, it decides better what information can be stripped. Encoding at 160 Kbps sounds good, at 192 Kbps it becomes difficult to hear the difference, and at 256 Kbps and above produce very good quality audio.

- MPEG System Structure
  - Audio Compression
    - MPEG Audio Layers
      - Layer III is the most complex MPEG Audio model.
        - It does even more filtering than Layer II and uses a Huffman coder. While encoding at 112 Kbps sounds good, 128 Kbps is even closer to the original; at 160 Kbps and 192 Kbps you won't hear a difference to the original.



2.6 MPEG Compression Standards
Jack Y.B. Lee

- MPEG System Structure
  - Video Compression
    - Two Basic Compression Techniques:
      - Block-based motion compensation for the reduction of the **temporal redundancy**, and
      - Transform domain (DCT) coding for the reduction of **spatial redundancy**.
    - Temporal Redundancy Reduction



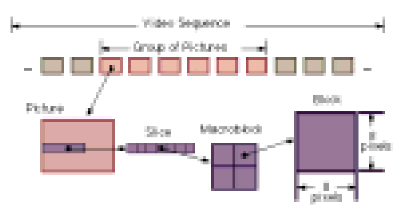
Three types of frames: intra pictures (I frames), predicted pictures (P frames), and bidirectionally interpolated pictures (B frames).

- MPEG System Structure
  - Video Compression
    - Temporal Redundancy Reduction
      - I frames
        - Compressed *independently* and provide access points for random access, but only with a *moderate* compression.
      - P frames
        - Coded with reference to a *previous* frame, which can be either an I or P frame, with higher compression.
      - B frames
        - Intended to be compressed with a low bit rate, using both the previous and future references (I, P).

```
IBBPBBPBBPBBIBBPBBPBBPBBIBBPBBPBBPBBI . . .
```

A group of picture (GOP)

- MPEG System Structure
  - Video Compression
    - Motion Compensation (Estimation)
      - Each picture is divided into blocks of 16 x 16 pixels, called a *macroblock*.
      - Each macroblock is predicted from the previous or future frame, by estimating the amount of the *motion* in the macroblock during the frame time interval.
      - This process is very computationally intensive.



- MPEG System Structure
  - Video Compression
    - Spatial Redundancy Reduction
      - For the reduction of spatial redundancy in each I picture or the prediction error in P and B pictures, the MPEG standard uses
        - Discrete cosine transform (DCT)
        - Quantization
        - Run-length encoding

source coding

entropy coding