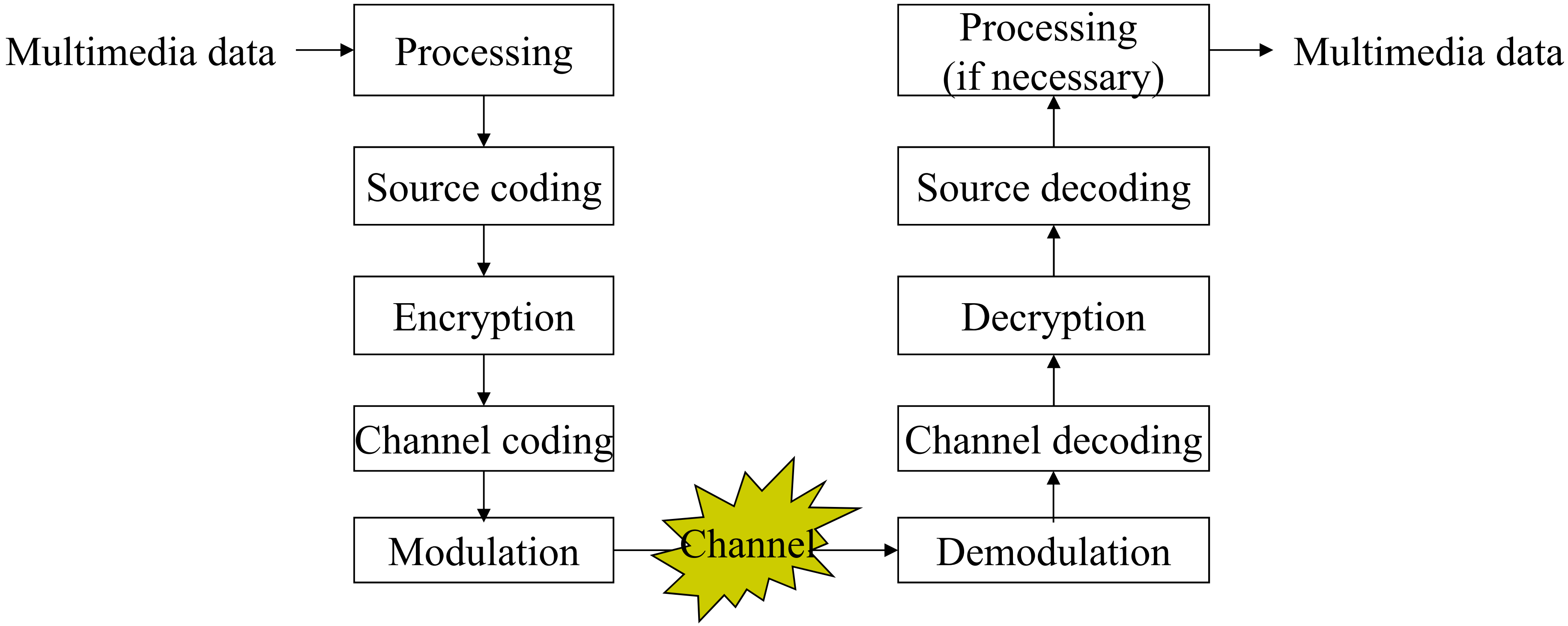
Data Compression

Introduction

中央大學資工系
蘇柏齊

Multimedia Communication System

- Multimedia processing/transmission/security

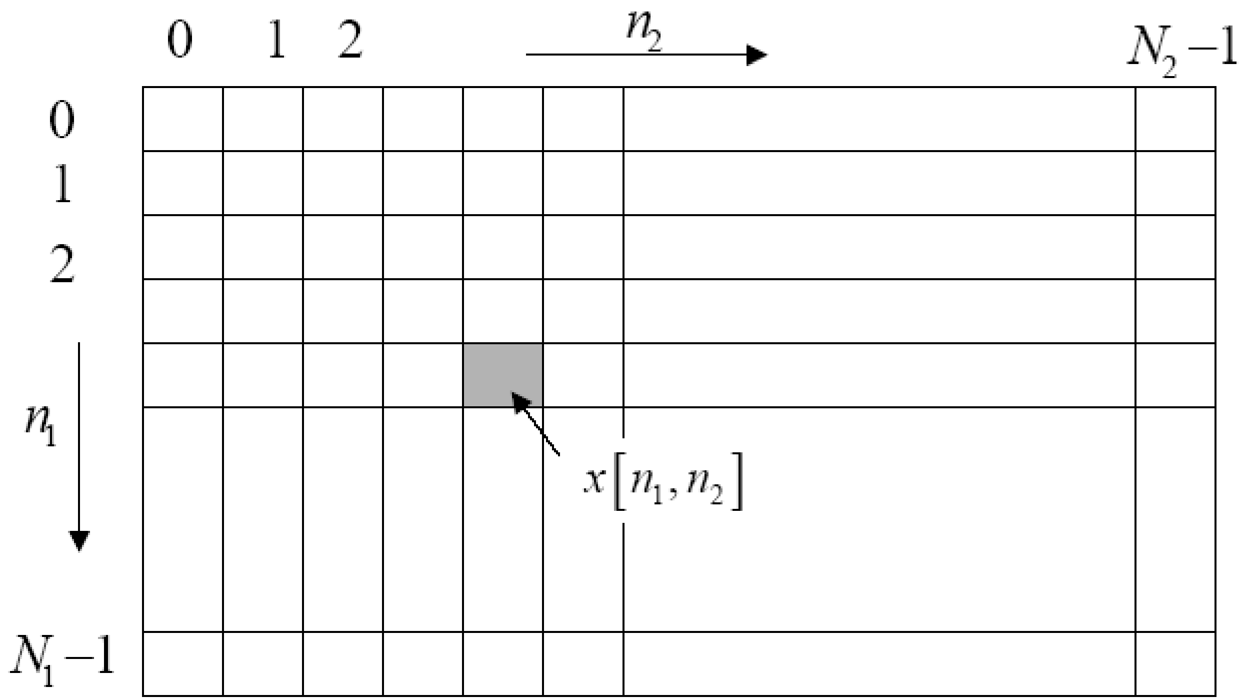


A Multimedia Communication System (cont.)

- The source encoder/decoder are used to reduce the redundancy →data compression
- The channel encoder / decoder are used to combat channel noises by adding some redundancy
  - CRC-4, Reed-Solomon Code
  - If the channel is noise-free, we can omit the channel coder
- The modulator takes in the channel encoder/source encoder output and outputs waveforms that suit the physical nature of the channel
- Encryption may be applied after the source coder
- We will focus only on source coder

Multimedia Data: Digital Images

- A digital image is a two-dimensional sequence of samples $x[n_1, n_2], \quad 0 \le n_1 < N_1, \quad 0 \le n_2 < N_2$



Discrete Image Intensities

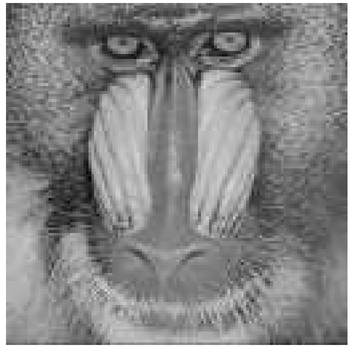
- Unsigned B-bit imagery

$$x[n_1, n_2] \in \{0, 1, \ldots, 2^B - 1\}$$

- Signed B-bit imagery

$$x[n_1, n_2] \in \{-2^{B-1}, -2^{B-1} + 1, \ldots, 2^{B-1} - 1\}$$

- Most common: *B=8*, but larger *B* are used in medical, military, or scientific applications.



Multiple Image Components

- Color images typically represented by three values per sample location, for red, green and blue primary components

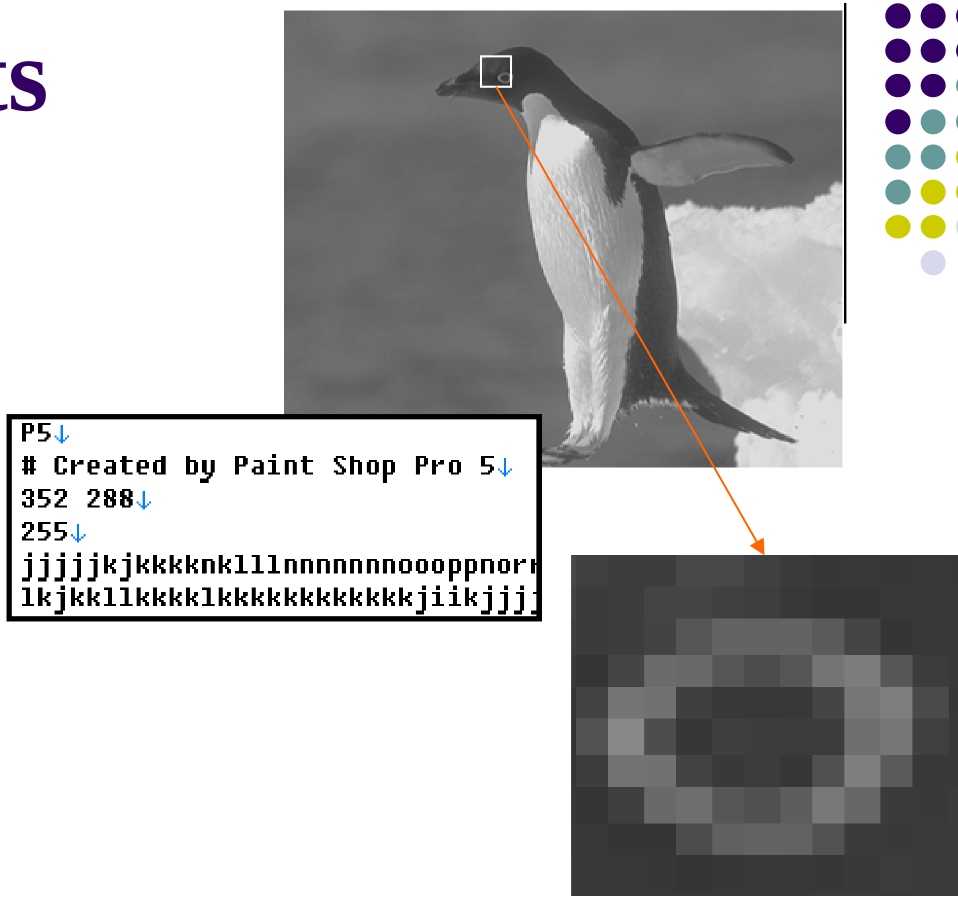
$$x_R[n_1,n_2], \quad x_G[n_1,n_2], \quad x_B[n_1,n_2]$$

- General multi-component image

$$x_C[n_1,n_2], \quad c=1,2,\ldots,C$$

Popular Formats for Images

- Raw data
  - .raw, .ppm, .pgm
- Formats (usually) with compression
  - JPEG
  - GIF
  - PNG
  - BMP/TIFF
  - JPEG2000
- Try Irfanview (www.irfanview.com)



Multimedia Data: Audio/Video

- The data volume of audio and video file is larger!
  - Digital audio: 80 Mins, 700 Mb
    - $44100 \times 2 \times 2 \times 60 \times 80 / 2^{20} = 747$ MB (x 2048/2352=703MB)
  - Digital video: 1 hour CIF video
    - $352 \times 288 \times 3 \times 30 \times 60 \times 60 / 2^{30} = 30.6$ GB
- Audio
  - MP3, AAC…
- Video
  - MPEG4, H.264, RMVB…

Multimedia Compression

- Multimedia compression looks for efficient representations of digital image/video/audio
  - Reduce the data volume to meet a bit-rate requirement
    - Minimize the bits required to represent a signal source for efficient transmission/ storage
  - Maintain the quality of the reconstructed data for target applications
    - Digital photography
    - VCD, DVD, Digital TV
    - Digital Audio
    - FAX /teleconferencing/video streaming
    - Medical image archiving
    - Fingerprint database
    - Remote sensing images
  - Complexity of computation involved is affordable
    - Hardware implementation
    - Cost
    - Real-time requirement

How to Achieve Compression? Making Good Use of Redundancy

- Utilizing (removing) the redundancy in the signal
  - Quick examples:
    - Text data:
      - Alphabetic redundancy:
        - Assigning variable size codes to the letters, with “E” getting the shortest code and “Z” getting the longest one.
      - Contextual redundancy:
        - The letter “Q” is usually followed by “U”.
    - Image data: Adjacent pixels tend to have similar colors
- Classification
  - Statistical redundancy
    - Inter-sample redundancy
    - Coding redundancy
  - Psychovisual redundancy

Inter-Sample Redundancy

- Spatial redundancy
  - For a sampled TV signal, normalized correlation between a row (column) and a one-pixel shift row (column) is very close to 1.
  - Intensity value of a pixel can be guessed from its neighbors.
    - Predictive coding, differential coding
- Temporal redundancy
  - 176x144=25344 pixels, only 3.4% have large value changes
  - For video-phone like signal, correlation between adjacent frames is around 0.8
  - Inter-frame predictive coding

Coding Redundancy

- Representation of information
- Example
  - Symbol / Occurrence probability / Code1 / Code2

Symbol	Occurrence probability	Code1	Code2
a1	0.1	000	0000
a2	0.2	001	01
a3	0.5	010	1
a4	0.05	011	0001
a5	0.15	100	001

  - Code 2: 1.95 bits per symbol
- Prefix code = instantaneous code
- Huffman coding, arithmetic coding…

Psychovisual Redundancy

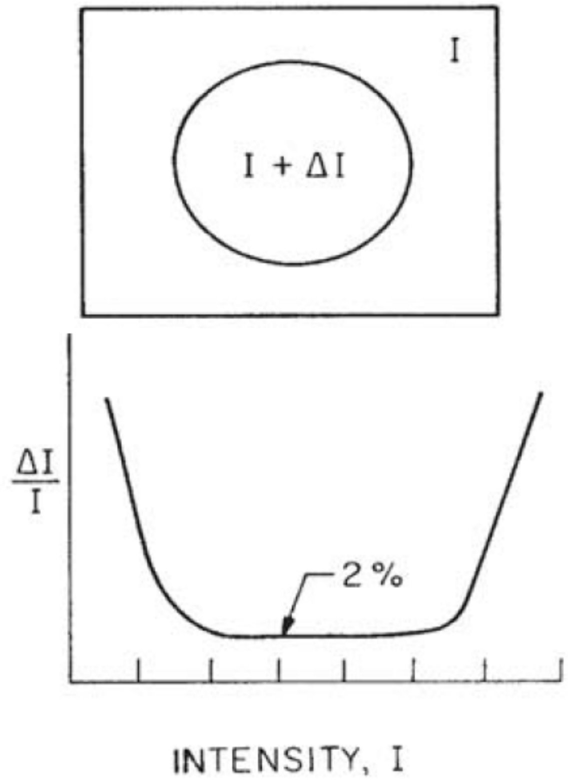
- Characteristics of human perceptual system
  - Visual information is not perceived equally; some information may be more important than other information (which is psychovisually redundant).
  - If we apply few data to represent less important visual information, perception will not be affected.
- Masking
  - Destructive interaction or interference among stimuli that are closely coupled in time or space, which may result in a failure in detection or errors in recognition.
- Masking in human visual system (HVS)
  - Contrast masking
  - Texture masking
  - Frequency masking
  - Temporal masking
  - Color masking

Contrast Masking

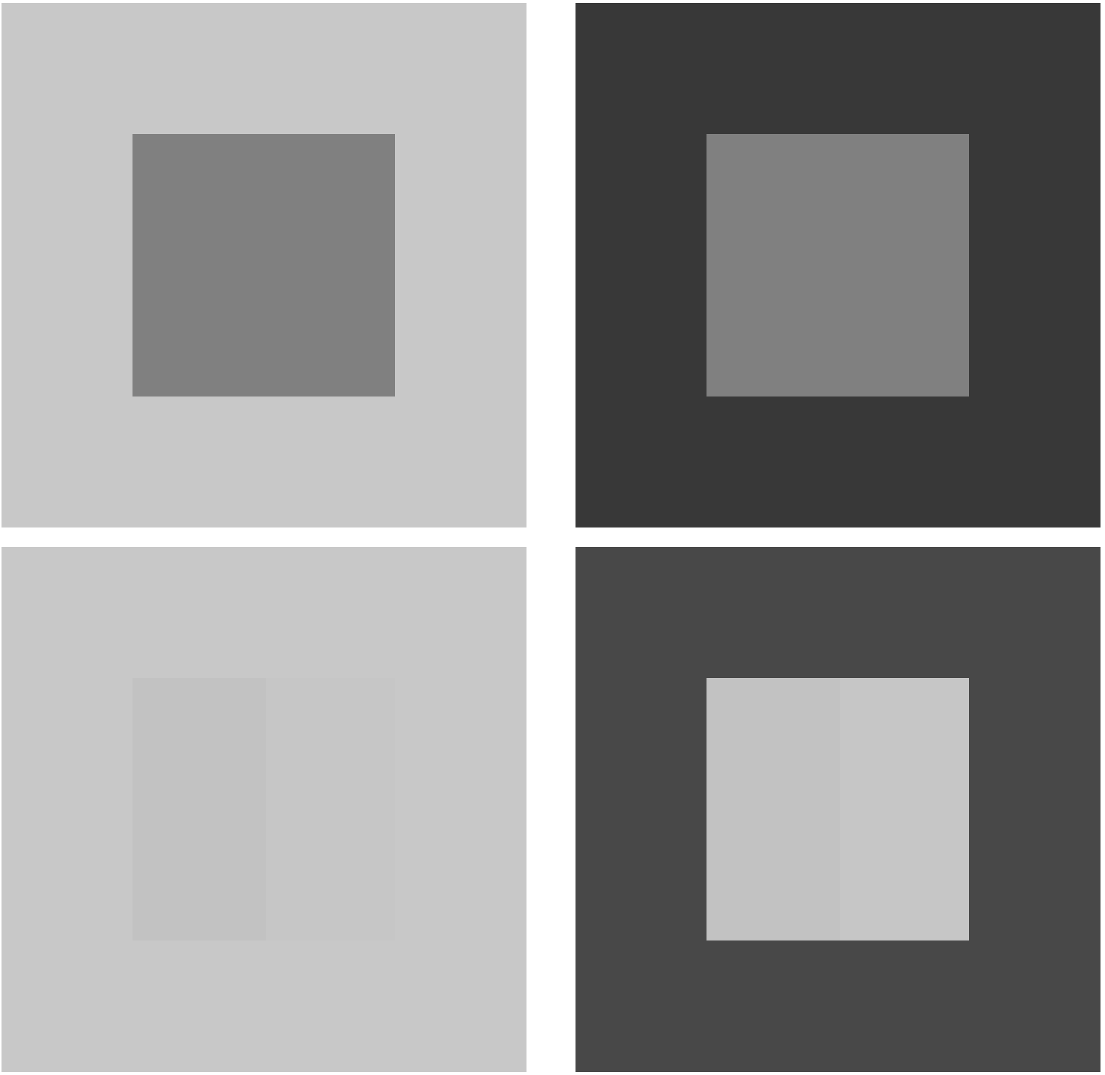
- Effect of one stimulus (masker, background)
  on the detectability of another stimulus (circle)
- Just noticeable difference (JND)
- Weber's law:
  - $$\frac{\Delta I}{I} = const \sim 0.02$$
  - For a relatively wide range of I, the threshold for discrimination is directly proportional to the intensity I.
    - When the background is bright, a larger difference is needed.
    - The intensity difference required could be smaller if the background is relatively dark.
  - d(log I)=dI/I
    - In some image processing, operations are performed on the logarithm of the intensity of an image point.
- Watson's model:

$$\Delta I = I_0 \times \max\left\{1, \left(\frac{I}{I_0}\right)^a\right\}$$

where I_0 is the luminance detection threshold when the gray level of the background is equal to zero and a is a constant, approximately equal to 0.7.



Contrast Masking (cont.) Local Contrast Adaptation

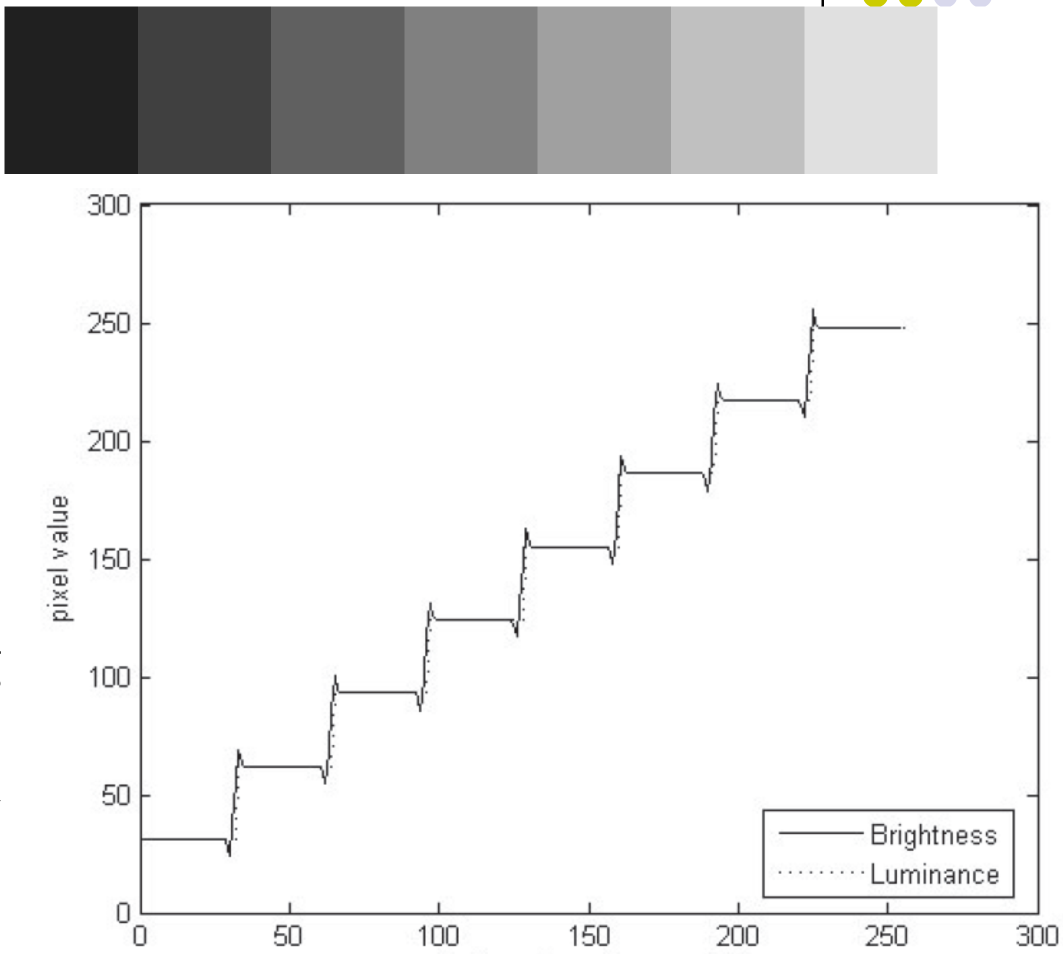


HVS adapts to surrounding brightness levels when it interprets the brightness of an object.

JND moves upward as the background brightness moves away from the average contrast of the object.

Contrast Masking (cont.)

- Mach Band
  - Each strip is darker at its right side than at its left
  - The Mach band overshoot in brightness is a consequence of the spatial frequency response of the eye
- Effects of contrast masking
  - The eye possess a lower sensitivity to high/low spatial frequencies than mid-frequencies
  - Perfect fidelity of edge contours can be sacrificed



Texture Masking

- Detail dependence, spatial masking, activity masking.
- The discrimination threshold increases with increasing picture detail.
- The stronger the texture, the larger the discrimination threshold.
- Example:
  - colors reduced to 16

Frequency Masking

- Frequency dependence
- Picture independent
- Example:
  - Adding noises and then reduce the number of colors to 16.
  - Low frequency error is converted to the high-frequency noise and the HVS is less sensitive to the high-frequency content.
- Human eyes function like a low-pass filter.
- Drop some high-frequency coefficients in the DCT domain.

Temporal Redundancy

- It takes a while for the HVS to adapt itself to the scene when the scene changes abruptly.
- The HVS is not sensitive to details during the transition. The masking takes place both before and after the abrupt change.
  - forward temporal masking
  - backward temporal masking
- This implies that we should take temporal masking into consideration when allocating data in video and audio coding.

Color Masking

- A color is an energy with an intensity as well as a set of wavelengths associated with the electromagnetic spectrum, hue & saturation.
  - Hue: the dominant wavelength
  - Saturation: purity of a color
    - A pure color has a saturation of 100
    - White light has a saturation of 0
- RGB model
  - The color sensitive area in the HVS consists of three different sets of cones and each set is sensitive to the light of one of the 3 primary colors: red, green, blue.
  - Color sensed by HVS can be considered as a linear combination of the 3 primary colors.
  - Acquisition and display
  - Luminance-chrominance color used in signal processing.

Luminance-Chrominance Model

- HSI model

$$H = \begin{cases} \theta & \text{if } B \le G \\ 360 - \theta & \text{if } B > G \end{cases}$$

$$\theta = \cos^{-1}\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\}.$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)].$$

$$I = \frac{1}{3}(R+G+B).$$

- YUV model (PAL)

$$Y = 0.299R + 0.587G + 0.114B$$

$$U = 0.492(B-Y) \qquad V = 0.877(R-Y)$$

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

- YIQ model (NTSC)

$$I = -0.545U + 0.839V \qquad Q = 0.839U + 0.545V$$

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

- YDbDr model (SECAM)

$$Db = 3.059U \qquad Dr = -2.169V$$

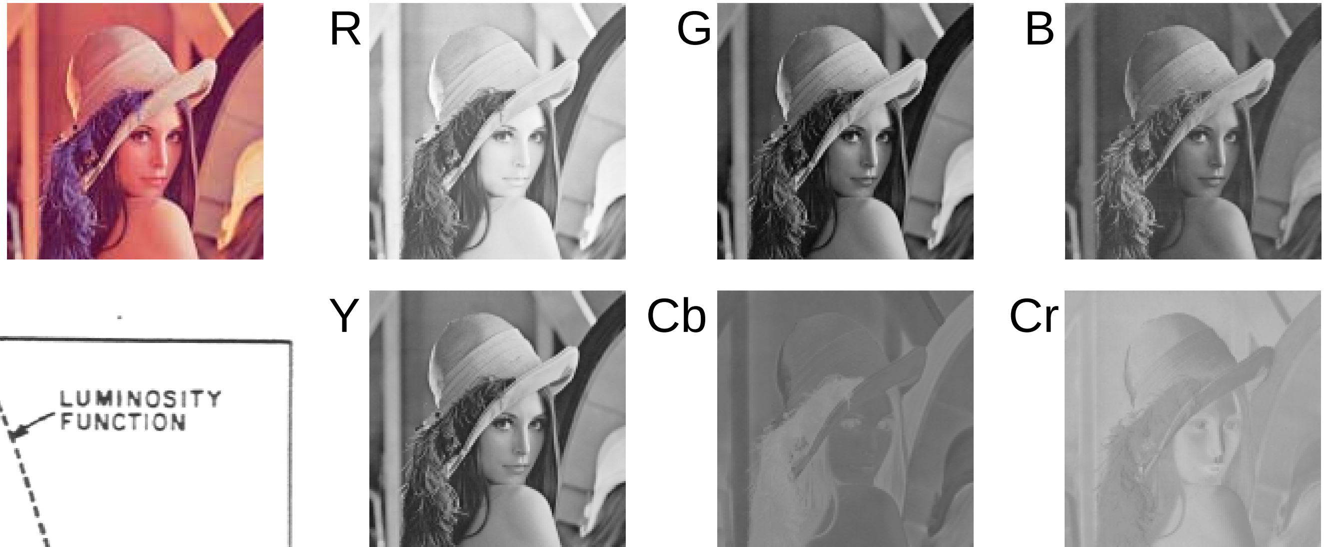
$$\begin{pmatrix} Y \\ Db \\ Dr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.450 & -0.883 & 1.333 \\ -1.333 & 1.116 & -0.217 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

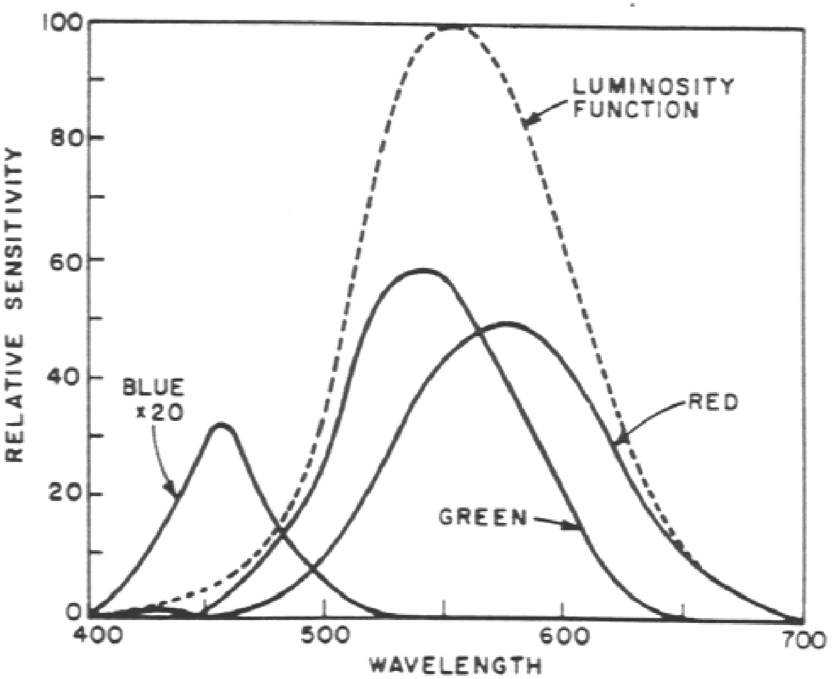
- YCbCr model

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix}$$

Y’=219(0.299R’+0.587G’+0.114B’)+16
Cb’=224(-0.169R’-0.331G’+0.500B’)+128
Cr’=224(0.500R’-0.419G’-0.081B’)+128
RGB [0,1] Y' [16,235] Cb' and Cr' [16, 240] with zero difference at 128, Other levels are reserved for synchronization and signal processing head- foot-rooms

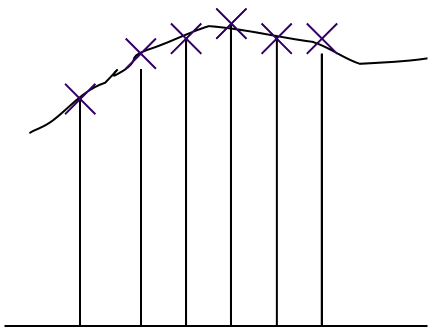
YCbCr Example





Utilizing “Signal Structure” for Data Compression

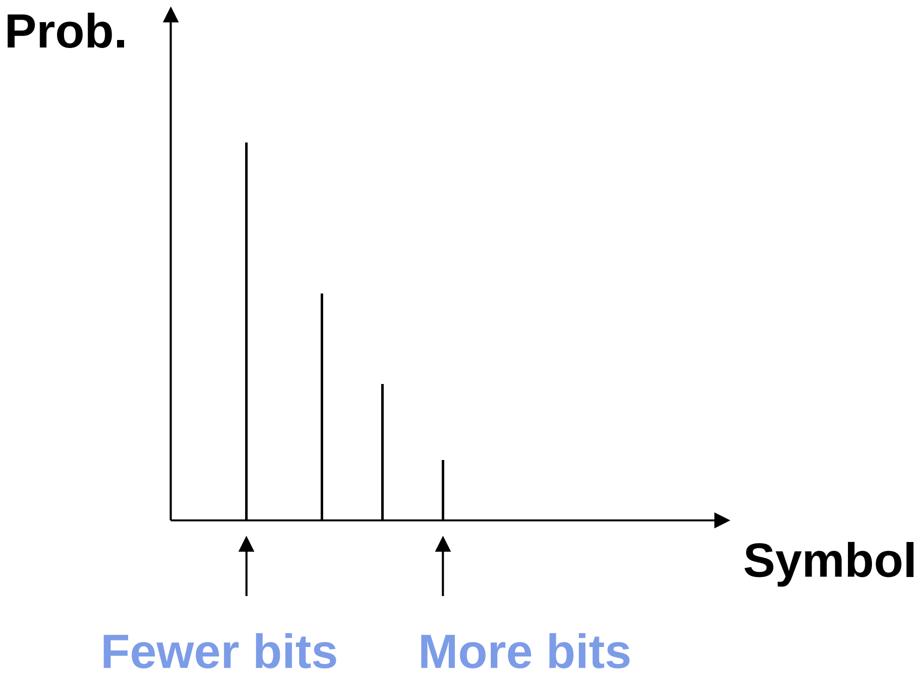
- Exploiting the structure in the data
  - Predicted coding
    - DPCM (Differential Pulse Code Modulation)



    - Motion predication for video compression

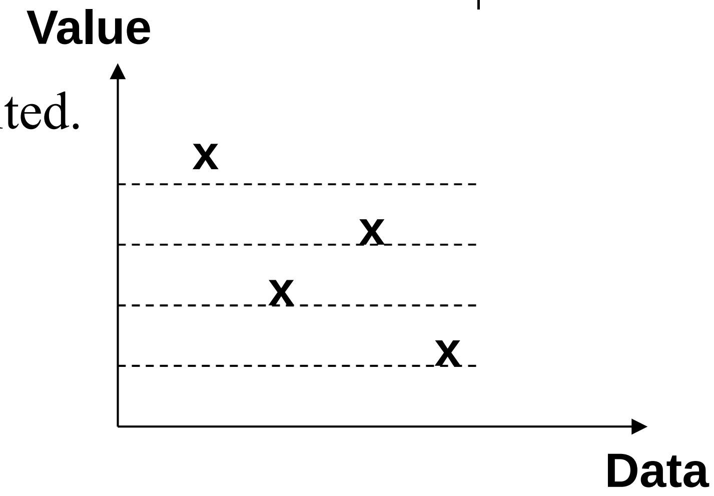
Utilizing “Statistics” for Data Compression

- Exploit the statistical properties of signals
  - Probability density function of the signal



Utilizing “Approximation” for Data Compression

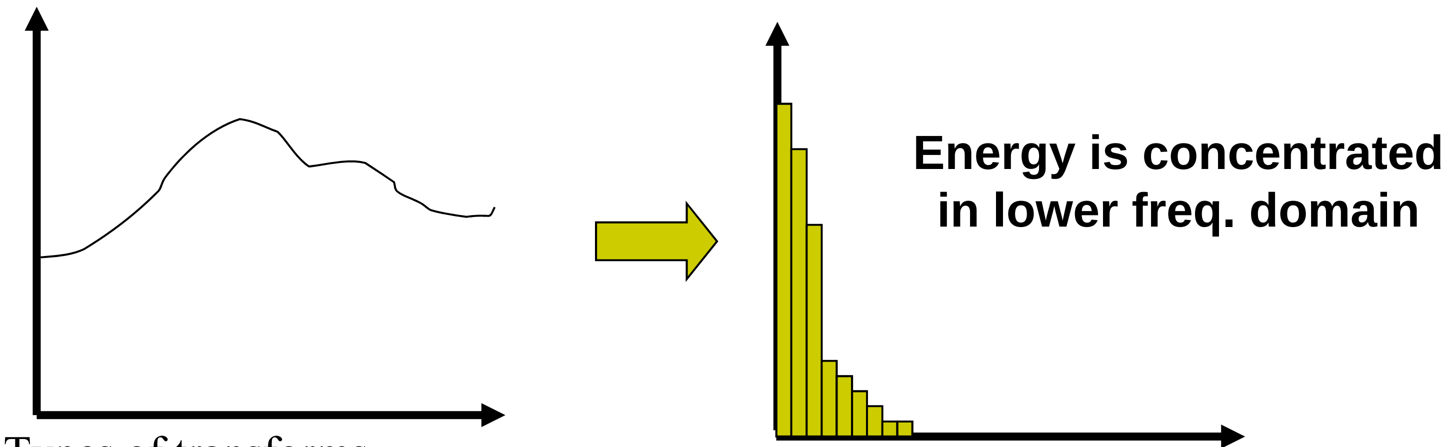
- Classification of compression: lossless and lossy compression
- Lossless compression:
  - All bits must be reconstructed.
  - Achievable compression usually rather limited.
  - Applications
    - Binary images (facsimile)
    - Medical images
    - Master copy before editing
    - Palletized color images



- Lossy compression is usually preferred
  - Make use of “approximation” acceptable to humans’ perception
  - Some deviation of decompressed data from the original is acceptable:
    - Human visual system might not perceive loss, or tolerate it
    - Digital input is imperfect representation of the real scene
  - Quantization
    - Goal: reduce the number of possible amplitude values for coding
    - Quantization error
  - Much higher compression than with lossless
  - Lossy compression is widely used for natural images & video

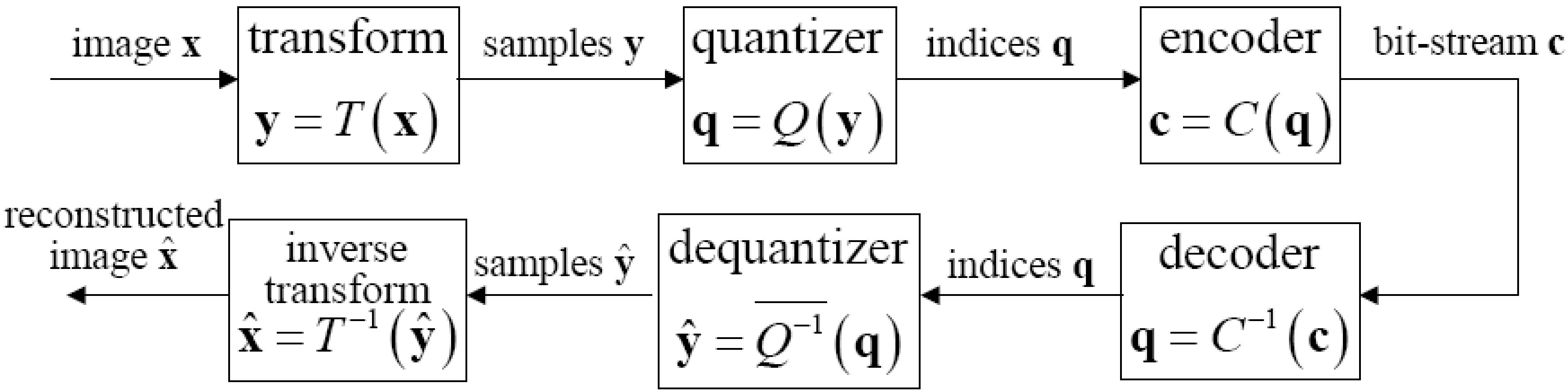
Utilizing "Transform" for Data Compression

- Goal: represent the image samples in a different form, such that statistical dependencies are greatly reduced
- Energy compaction via transform



- Types of transforms
  - Discrete cosine transform (widely used in standards: JPEG/MPEG)
  - Wavelet transform (JPEG2000, MPEG4 texture coding, EZW)
    - EZW: Embedded Zerotree Wavelet Compression (Shapiro)
    - SPIHT: Set Partitioning in hierarchical Trees (Pearlman)
    - EBCOT: Embedded Block Coding with Optimized Truncation (Taubman)

Example: Typical Still Image Compression System



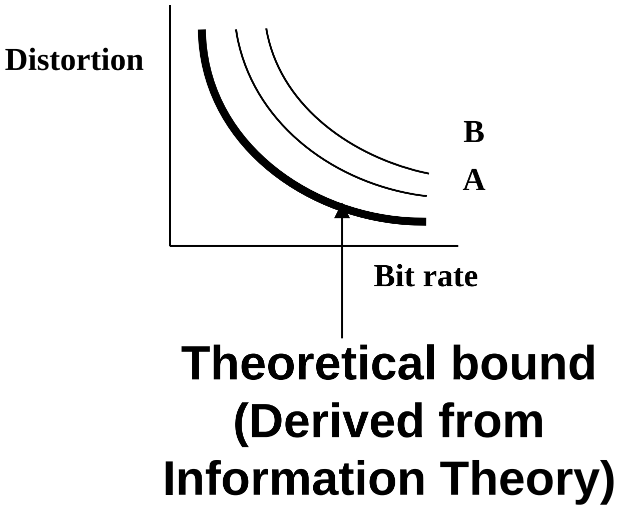
- Transform $T(\boldsymbol{x})$ invertible
- Quantization Q(**y**) not invertible, introduces distortion
- Combination of encoder C(**q**) and decoder C^{-1}(**c**) lossless

Utilizing "Scalability"

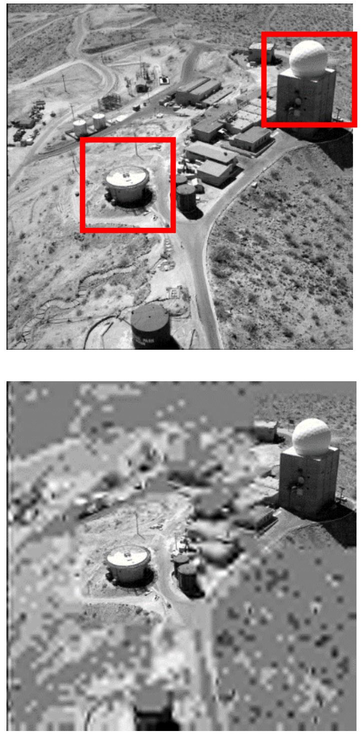
- Multi-resolution/Embedded/ Progressive property/Scalability
  - Spatial resolution
  - Quality resolution
  - Temporal resolution

Other Functionalities

- Better codec



- Interactivity with contents
  - ROI (Region of Interest)
- Error resiliency
- Error concealment



Measures of Compression

- Image:
  - Image represented by a "bit-stream" **c** with length $\|\mathbf{c}\|$.
  - Compare number of bits w/ and w/o compression

$$\boxed{\text{compression ratio} = \frac{N_1 N_2 B}{\|\mathbf{c}\|}} \quad \text{or} \quad \boxed{\text{bit-rate} = \frac{\|\mathbf{c}\|}{N_1 N_2} \text{ bits/pixel}}$$

  - Bit-rates substantially dependent on image content
    - For typical natural images
      - Lossless compression: *(B-3)*bpp (bits per pixel)
      - Lossy compression,
        - High quality: 1 bpp
        - Moderate quality: 0.5 bpp
        - Usable quality: 0.25 bpp
  - Perceived distortion depends on sampling density and contrast
- Audio/Video: bits per second (bps)

Lossy Compression: Measuring Distortion

- Most commonly employed:
  - Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left(x[n_1, n_2] - \hat{x}[n_1, n_2] \right)^2$$

  - Peak Signal to Noise Ratio (PSNR)

$$\text{PSNR} = 10 \log_{10} \frac{\left(2^B - 1\right)^2}{\text{MSE}} \text{ dB}$$

- Advantages
  - Easy calculation
  - Mathematical tractability in optimization problems
- Disadvantage
  - Neglects properties of human vision