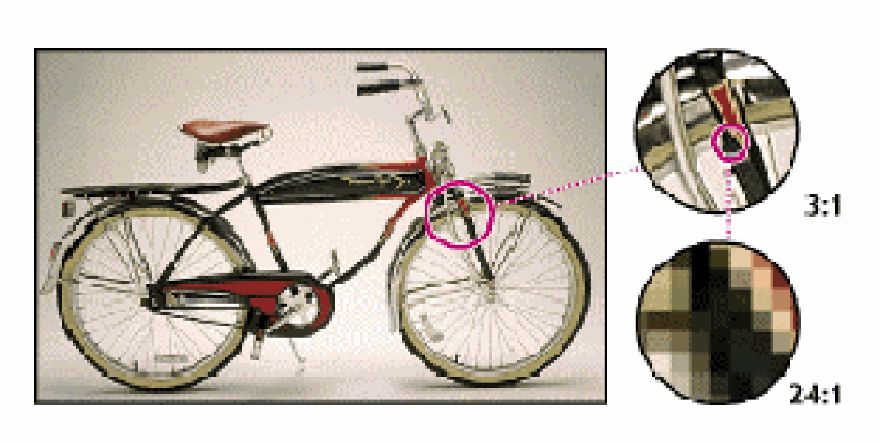
Chapter 1 Introduction

Outline

- Basic concept
- Why do we need compression ?
- Why compressible ?
- Visual quality measurement
- Information theory results

Digital Image

- Representation of raw digital image
  - Pixels (picture element)
  - RGB (e.g. 24 bits)
  - 2 D (e.g. 512x512)



Lenna

Digital Video

- A sequence of video frames (moving images)
- How to determine the frame rate?
  - Persistence of vision
    - The brain retains an image for a split second
    - Instead of seeing the flashing of a series of images, the brain sees the illusion of motion
    - It was determined that any rate less than 16 fps caused the brain to see flashing images
      - Today's theatrical film runs at 24 fps
      - Digital video records at an equivalent to 29.97 frames per second
      - Television in the US displays a complete new image at just under 30 fps and at a comparable rate in other countries.

Why Compression?

- Huge amount of data
  - Modem: max. bit rate of 56.6 kbps
  - 352x288 (CIF) digital color images, each pixel represented in RGB (1 byte/color or 24 bits/pixel)
  - Frame rate: 30 fps
  - Bit rate would be 352x288x24x30=72 mbps
  - A compression ratio of about **1289** is needed to view such video through the phone modem
  - What about DVD (720x480), HDTV(1960x1024)?

Why Compression?

- Storage requirement huge
  - 2-hour movie (720x480)--> 208 Giga bytes
- Bandwidth limitation
- **Image compression**
  - A process in which amount of data used to represent image is reduced to meet a bit rate requirement, while the quality of reconstructed image satisfies the requirement of the application, and the complexity is affordable.
  - Saves both transmission bandwidth and storage space

Why Compressible?

- Statistical redundancy
  - Spatial redundancy
  - Temporal redundancy
  - Coding redundancy
- Psycho-visual redundancy
  - Luminance/contrast masking
  - Texture masking
  - Frequency sensitivity
  - Color sensitivity
  - Temporal masking

Why Compressible?

- **Statistical redundancy**
  - Inter pixel redundancy
    - Spatial redundancy
      - Autocorrelation along row/column is high (near 1)
      - The intensity value of a pixel can be guessed from that of its neighboring pixels
    - Temporal redundancy
      - Similarity between two neighboring frames is strong
      - Predict a frame from its neighboring frame
  - **Coding redundancy**
    - Data redundancy (a file can be zipped)
      - Efficient representation of symbols
    - Information theory results

Why Compressible?

- Psycho-visual redundancy
  - HVS (Human Visual System)
    - Non-linear
    - Visual information is not perceived equally
    - Coding vision-sensitive information
    - Masking:
      - The detectability of one stimulus depends on the detectability of another

Why Compressible?

- **Luminance/contrast masking**

$$\frac{\Delta I}{I} \approx const$$

  - ΔI: threshold for an object to be noticed with a 50% chance.
  - It implies that
    - When the background is bright, a larger difference in gray levels is needed for the HVS to discriminate the object
  - It directly affects quantization in video and image compression
    - A non-uniform quantization is implied.

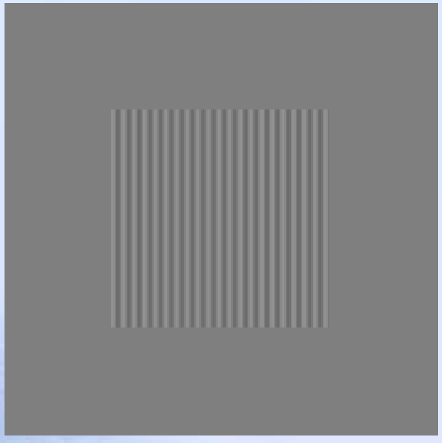
Self-contrask-masking

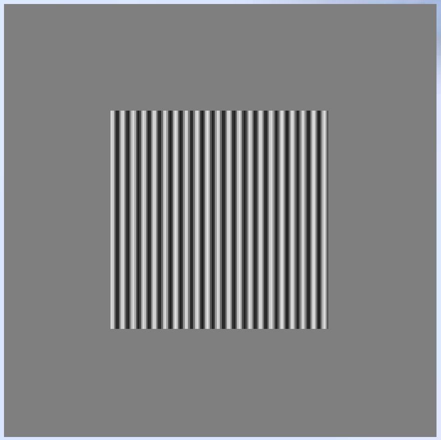
coefficient amplitude determines the masking level of itself

Low amplitude

High amplitude

(a) (mask)

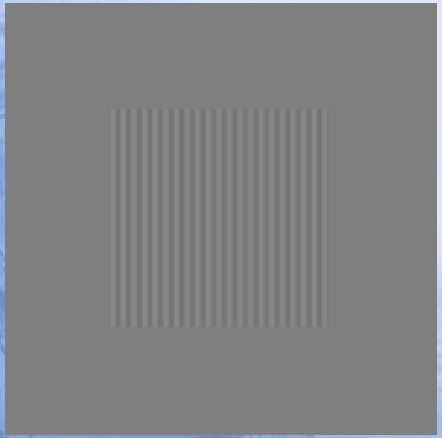


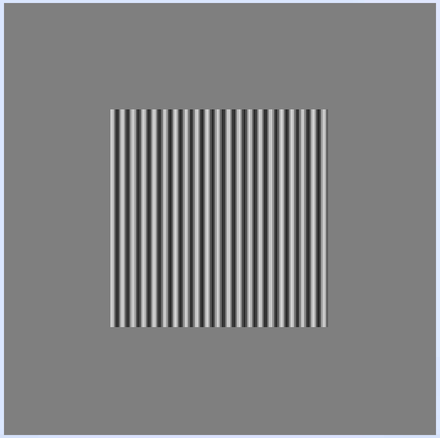


(a) (mask)

Same amplitude difference /distortion between (a) and (b)

(b) (with distortion)





(b) (with distortion)

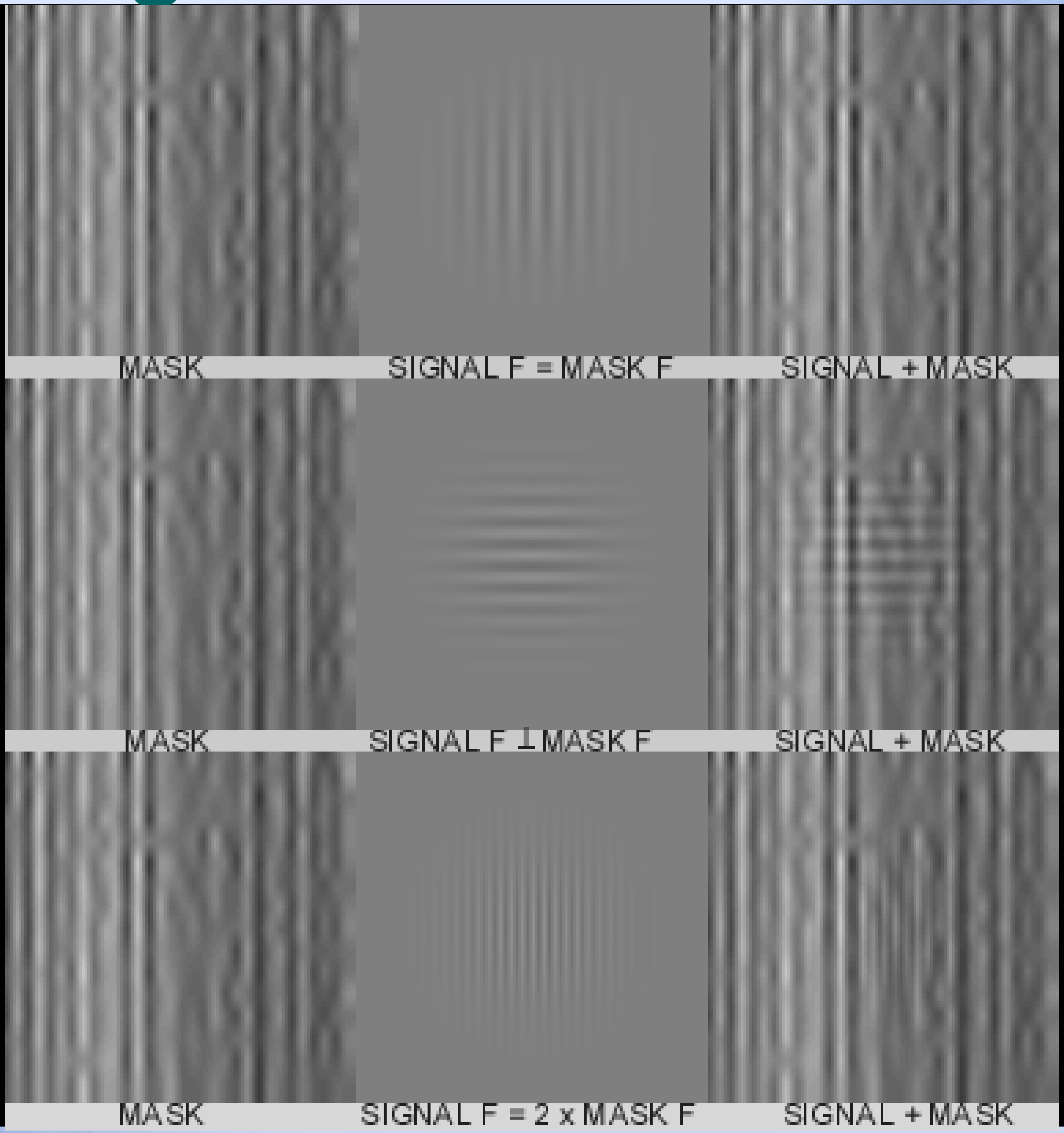
Why Compressible?

- **Texture masking**
  - The discrimination threshold increases with increasing picture details.

S_i

Visual Masking: Demonstration

- Typical result from masking by noise
- Noise is narrowband
  - **limited radial freq**
  - **limited orientation**
- Little masking unless frequencies of mask are close to those of signal
  - both radially and in orientation
- In compression, the **image itself acts as the mask**

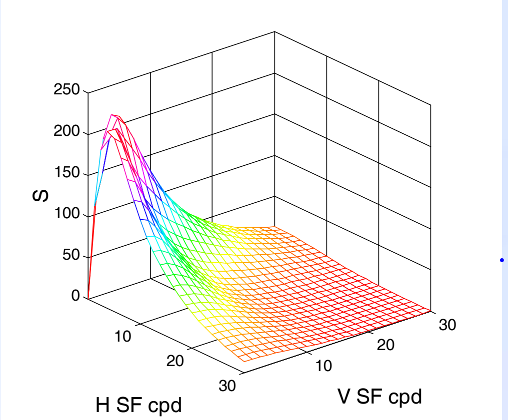


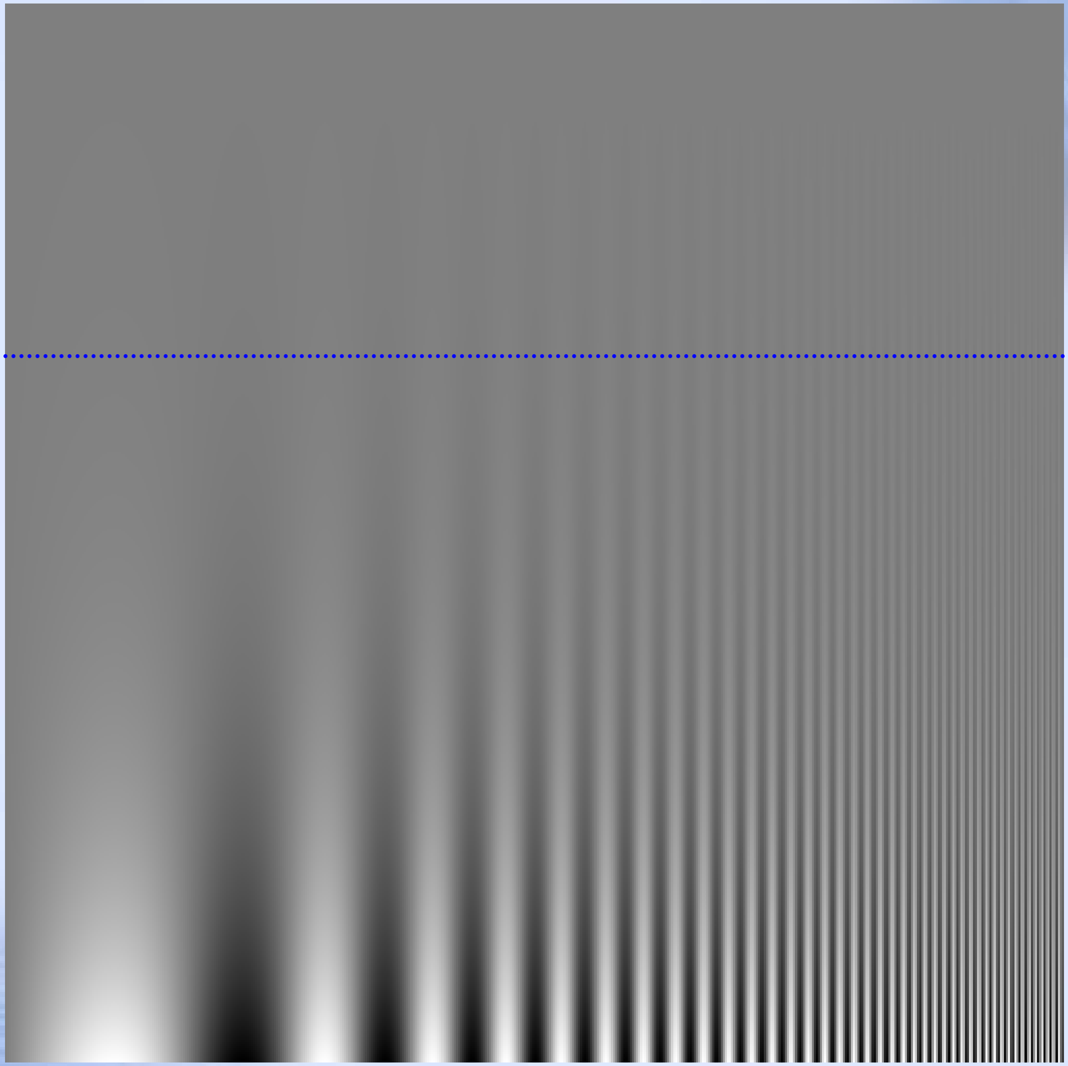
(Courtesy of Scott Daly)

Why Compressible?

- **Frequency sensitivity**
  - The discrimination threshold increases with increasing frequency
  - Exploited in transform coding
- **Temporal masking**
  - HVS is not sensitive to details when there is an abrupt scene change

Contrast Sensitivity Function (CSF)





Decreasing amplitude

Increasing frequencies

(Courtesy of Scott Daly)

Why Compressible?

- **Frequency sensitivity**
  - The discrimination threshold increases with increasing frequency
  - Exploited in transform coding
- **Temporal masking**
  - HVS is not sensitive to details when there is an abrupt scene change

Why Compressible?

- **Color sensitivity**
  - RGB Model
    - Red, green and blue
  - HSI Model
    - Hue, saturation and intensity
    - Not used
  - YUV Model
    - Y: luminance, U/V: chrominance, used in PAL TV system
  - YIQ Model
    - Used in NTSC TV system

Why Compressible?

- **Color sensitivity**
  - YDbDr Model
    - Used in SECAM
  - YCbCr Model
    - Used in JPEG and MPEG
    - U and V are shifted to non-negative
  - All YCC value has a one-to-one correspondence with RGB, with conversion matrix

Why Compressible?

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

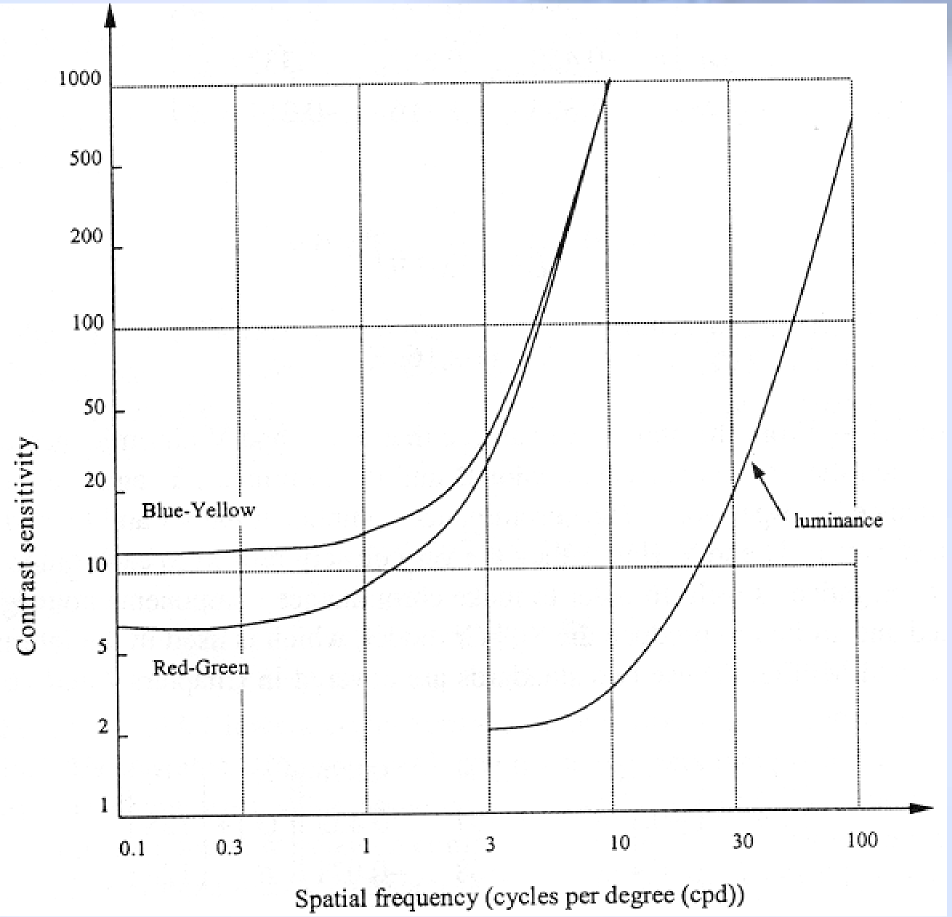
$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

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

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

Why Compressible?

- Color sensitivity
  - HVS much more sensitive to luminance than to chrominance
  - Chrominance can be down-sampled



Visual Quality Measurement

- **How to evaluate the results of two compression methods?**
- **Subjective quality measurement**
  - Five scale rating system used by Bell Labs
    1. Impairment is not noticeable
    2. Impairment is just noticeable
    3. Impairment is definitely noticeable, but not objectionable
    4. Impairment is objectionable
    5. Impairment is extremely objectionable
  - Subjective quality measurement is costly

Visual Quality Measurement

- **Objective quality measurement**
  - (Peak) Signal-to-noise ratio (PSNR)

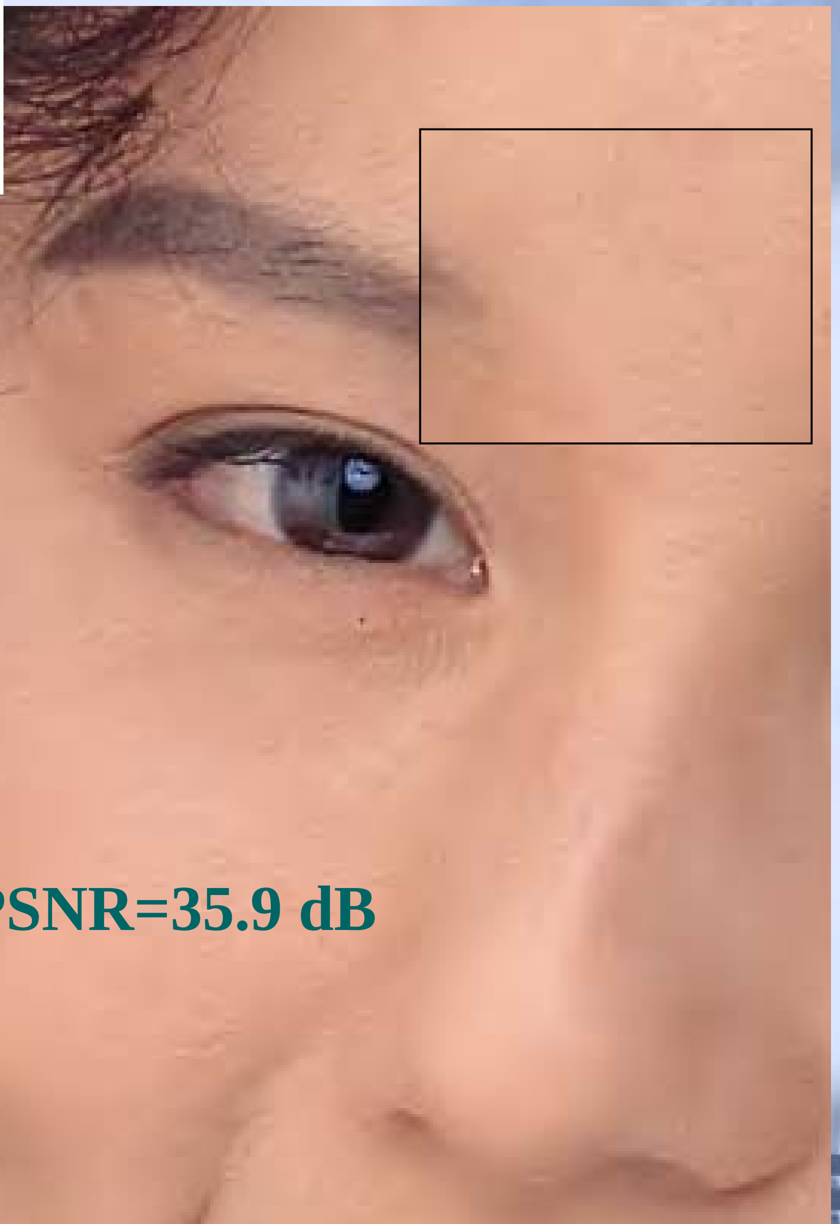
$$e(x,y) = f(x,y) - g(x,y) \qquad E_{ms} = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} e(x,y)^2 \qquad E_{rms} = \sqrt{E_{ms}}$$

$$SNR_{ms} = 10\log_{10}\left[\frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} g(x,y)^2}{MN \cdot E_{ms}}\right] \qquad SNR_{rms} = \sqrt{SNR_{ms}}$$

$$PSNR = 10\log_{10}\left[\frac{255^2}{E_{ms}}\right]$$

JPEG2000 (at 0.75 bpp)

Frequency & Color Weighting vs. No Weighting



Information Theory Results

- **Entropy**
  - Information Measure

$$I = \log_2 \frac{1}{p} bits$$

  - Average Information per Symbol (entropy)

$$H = -\sum_{i=1}^{m} P_i \log_2 P_i \quad bits$$

  - Summarize
    - Entropy is the measure of randomness
      - Reach maximum when all symbols equally probable
    - Information comes from the decrease of entropy

Information Theory Results

- **Shannon's Noiseless Source Coding Theorem**
  - For a discrete, memoriless, stationary information source, the *minimum* bit rate required to encode a symbol on average is equal to the entropy of the source
  - Coding efficiency is measured by H/L_{avg}
    - L_{avg} *: average length per symbol*

Information Theory Results

- **Shannon's Noisy Channel Coding Theorem**
  - It is possible to transmit symbols over a noisy channel without error if the bit rate *R* is below the channel capacity *C*
  - For AWGN channel, we have

$$C = W \log_2 (1 + \frac{S}{N})$$

- **Shannon's Source Coding Theorem**
  - For a given distortion *D*, there exists a rate distortion function *R*(*D*), which is the minimum bit rate required to transmit the source with distortion less than or equal to *D*

Information Theory Results

- Information Transmission Theorem
  - $C >= R(D)$
  - If the channel capacity of a noisy channel C is larger than the rate distortion function $R(D)$, then it is possible to transmit an information source with distortion D over a noisy channel
- Current coding schemes are far from these theoretical limits
  - E.g., the theoretical source coding limit for human speech is 300 bits/s, but currently the commercially used speech codecs are still using 4kps-8kbps

Summary

- Motivations for compression
  - Saves transmission bandwidth and storage space
  - Enabling technology
- Principles of compression
  - Statistical redundancy
  - Psycho-visual redundancy
- Visual quality measurement
  - HVS plays a very important role
  - Should take advantage of HVS properties
- Information theoretical results