Image Compression and Video Compression
Topics

- Introduction to Image Compression
- Transform Coding
- Subband Coding, Filter Banks
- Haar Wavelet Transform
- SPIHT, EZW, JPEG 2000
- Motion Compensation
- Wireless Video Compression
Standards

- JPEG
  - Joint Photographic Expert Group
  - Still Image Compression Standard
- JBIG
  - Bi-level Compression (Not used much)
- JPEG-2000
  - Wavelet Based
- H.261, H.263 and H.264
  - Interactive Video Standards.
- MPEG-1, MPEG-2
  - Moving Picture Expert Group
  - Video Compression Standards for CD ROM’s, internet and television.
- MPEG – 4
  - Low Bit Rate Video Compression
- MPEG - 7
  - (follow-on to MPEG-4) Content searchable video
Contents

- Capture and represent digital images
- Light and Color
- Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy
- Distortion Measures

Human Visual System
Image Orientation

- An image consists of a two-dimensional array of pixels.
- Neighbors of a pixel

<table>
<thead>
<tr>
<th>(x-1,y-1)</th>
<th>(x-1,y)</th>
<th>(x-1,y+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x,y-1)</td>
<td>(x,y)</td>
<td>(x,y+1)</td>
</tr>
<tr>
<td>(x+1,y-1)</td>
<td>(x+1,y)</td>
<td>(x+1,y+1)</td>
</tr>
</tbody>
</table>
Number of Gray Levels

- Decreasing gray level results in visual displeasure and loss of detail.
Contents

- Capture and represent digital images
- Light and Color
- Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy
- Distortion Measures

Human Visual System
Light and Color

- The wavelengths of electromagnetic radiation between roughly 370nm and 730nm account for light visible to the human visual system.

- The curve is the spectrum of the solar radiation (the WHITE light) reaching the earth’s surface.

- The white light is a mixture of all wavelengths.
Visible Light

- Color is the visual perception to different wavelengths of visible light.

- Different ranges give rise to different color response.

<table>
<thead>
<tr>
<th>Range* (in nm)</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>380 - 450</td>
<td>Violet</td>
</tr>
<tr>
<td>450 - 490</td>
<td>Blue</td>
</tr>
<tr>
<td>490 - 560</td>
<td>Green</td>
</tr>
<tr>
<td>560 - 590</td>
<td>Yellow</td>
</tr>
<tr>
<td>590 - 640</td>
<td>Orange</td>
</tr>
<tr>
<td>640 - 730</td>
<td>Red</td>
</tr>
</tbody>
</table>

*Ranges are approximate
*There is no sharp boundary
Almost any color can be described as the mixture of the basic color: R, G, B (YES, theoretically there are some colors that cannot be encoded as a combination of three colors).

This is the basis of the digital color techniques. Used in the digital camera, CRT/LCD monitor, projector, etc.

Directly due to the three photoreceptors on the retina (explained later).
Color Reproduction

- Each pixel can be represented by a spectrum (left). On the monitor it is represented by the [RGB] value.
Black and White Image

- Only two colors (black or white) for each pixel on the screen.
- Color Depth = 1 is black and 0 is white.
Although the exact colors that an 8-bit screen can display are not fixed, there can never be more than 256 unique colors on the screen at once. You may assign colors as follows and other variations are possible.

<table>
<thead>
<tr>
<th>R</th>
<th>G</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
In 8-bit images the 256 colors that make up the image are stored in an array called a "palette" or an "index".

As mentioned above, 8-bit images can never contain more than 256 unique colors.
"True-color" or "24-bit" color displays can show millions of unique colors simultaneously on the computer screen.

- 8 bits each for the red, green, and blue channel.

$2^{24} = 16,788$ million colors.

3-dimensional Color space
True Color Example

- True-color, or 24-bit, images are typically much larger than 8-bit images in their uncompressed state.
Capture and represent digital images
Light and Color
Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy
Distortion Measures

Human Visual System
Why can Images be Compressed?

- Image compression can be achieved primarily because image data are highly redundant.
- The degree of redundancy determines how much compression can be achieved.
- Four types of redundancy can be identified:
  - Spatial Redundancy
    - Correlation between adjacent data points
  - Temporal Redundancy
    - Correlation between different frames in an image
  - Spectral Redundancy
    - Correlation between different color planes or sensors
    - Limitation of Low-level Human Vision System
  - Psycho-visual Redundancy
    - Limitation of high-level Human Vision System
Spatial Redundancy

- Neighboring pixels are **highly correlated**.

Washington DC Image
Spatial Redundancy

- The histogram of the difference signal is highly peaked at origin which can be exploited by the Entropy Coder.

Spread of differences in neighboring pixels
Contents

- Capture and represent digital images
- Light and Color
- Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy
- Distortion Measures

Human Visual System
Temporal Redundancy

- Adjacent frames are highly correlated.

[Images of previous and current frames with displacement vectors and motion-compensated prediction error]
Contents

- Capture and represent digital images
- Light and Color
- Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy
- Low-level Vision
- Distortion Measures
The Eye
Eccentricity

- The fovea subtends 2 degrees of visual angle
- To stimulate the fovea the stimulus must be centered on the screen and cover a visual angle of 2 degrees
  - 2cm on the screen at 57cm
  - 4cm on the screen at 114cm
Cones and Rods Shapes

- The light enters the inner segment and passes into the outer segment which contains light absorbing photopigments.
- Less than 10% photons are absorbed by the photopigments.
Cones and Rods

- Converts light into signals
- Transmits signal to brain

# cones in each retina: $5 \times 10^6$
# rods in each retina: $10^8$
# optic nerves: $1.5 \times 10^6$

Therefore, the rods and cones must be interconnected to nerve fibers on a many-to-one basis.

http://white.stanford.edu/~brian,numbers,numbers.html
Rods

Number of rods in each retina: $10^8$

Photoreceptors which

- Generate **achromatic** response.

- Sensitive only in low light levels (scotopic range... night vision).
Cones

Number of cones in each retina: $5 \times 10^6$

Photoreceptors

- Responsible for **COLOR** vision.
- 3 Types of cones: **S, M, L** in Fovea
- Sensitive only in high light levels.
  - Human being have difficulty to see color in the dark.

Dogs have 2 type of cones.
Bees have 4, Mantis Shrimp has record 10.
Cones: S, M and L

- S: Blue response (peak: 445nm)
- M: Green or yellow response (peak: 535nm)
- L: Red response (peak: 575nm)
- Not uniformly sensitive
  - more sensitive to green than red (64%L, 32%M, 2%S).
Opponent Processing

- Retina carries out “matrix operation” to represent colors in the opponent color system (Y, Y-B, R-G)
- Cannot see BLACK and WHITE at the same time.
Spectral Redundancy

- The output of the three cone color is transformed into an achromatic channel (such as luminance) and two chromatic channels (opponent channels).

- High-level HVS (Human Vision System) is much more sensitive to the variations in the achromatic channel than in the chromatic channels (explained later).

- The chromatic components are often sub-sampled by a factor of 2:1 or 4:1 in both horizontal and vertical directions (as in JPEG, MPEG).
Many transformations, depending on the application, have been suggested to exploit the correlation in the R, G, B channel.

The JPEG standard does not provide a definition for the color space to be used; the choice has intentionally been left to the user.
An example of a transformation used for color image compression is that used in the NTSC national color television standard.

The RGB values are transformed into a new set of values known as Y (gray scale info), I (hue), and Q (saturation), (also called luminance, in-phase, and quadrature, respectively) according to

\[
\begin{pmatrix}
Y(i,j) \\
I(i,j) \\
Q(i,j)
\end{pmatrix} = \begin{pmatrix}
.299 & .587 & .114 \\
.596 & -.274 & -.322 \\
.212 & -.523 & .311
\end{pmatrix} \begin{pmatrix}
R(i,j) \\
G(i,j) \\
B(i,j)
\end{pmatrix}
\]

First row sum=1, second and third row sum=0. For gray scale R=G=B
Spectral Redundancy

The transformation to convert YIQ back to RGB values is:

\[
\begin{pmatrix}
R(i, j) \\
G(i, j) \\
B(i, j)
\end{pmatrix} =
\begin{pmatrix}
1.000 & 0.956 & 0.621 \\
1.000 & -0.272 & -0.647 \\
1.000 & -1.106 & 1.703
\end{pmatrix}
\begin{pmatrix}
Y(i, j) \\
I(i, j) \\
Q(i, j)
\end{pmatrix}
\]
A transform similar to the YIQ, but easier to implement in hardware is the color difference transform, which generates the Y, R-Y, and B - Y components. This is also denoted as \(YC_bC_r\) (Luminance, Chrominance-blue, Chrominance-red representation).

The perfect transform has yet to be agreed on, but with respect to data compression, the performance difference are insignificant.
Color transform: YCrCb <-> RGB

- The transformation is a matrix operation.

\[
\begin{bmatrix}
Y \\
U \\
V
\end{bmatrix} = \begin{bmatrix}
0.299 & 0.587 & 0.114 \\
-0.16875 & -0.33126 & 0.5 \\
0.5 & -0.41869 & 0.08131
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\[
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix} = \begin{bmatrix}
1.0 & 0 & 1.402 \\
1.0 & -0.34413 & -0.71414 \\
1.0 & 1.772 & 0
\end{bmatrix} \begin{bmatrix}
Y \\
U \\
V
\end{bmatrix}
\]
Contents

- Capture and represent digital images
- Light and Color
- Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy
  - High-level Vision
- Distortion Measures
Human visual perception

- The spectral redundancy occurs in the low-level vision system (first three layers of the retina in the eyes). At higher levels of human brain visual perception takes place.
- Human visual perception is very complex.
Human visual perception

- Can you believe your eyes?

Black dot?
White dot?
Human visual perception

- Can you believe your eyes?

Wheel circling?
“Look and you will find it -- what is unsought will go undetected”
Human visual perception

- Can you believe your eyes?

How many faces here?
Human visual perception

- Can you believe your eyes?

Halloween?
Can you believe your eyes?

Halloween?  
Lady?
Human visual perception

- Can you believe it?

Halloween!
Lady!
Human visual perception

- Human vision system is very complex.
- Only part of it is explored.
- Some can be applied to compression.
Our sensitivity to the resolution is a function of the distance.
Human visual perception

- This knowledge is exploited for the progressive geometry compression.
- Select resolution based upon the projected screen size (or area) of an element. Objects appear smaller as they move further away.
Human Visual Perception

- Spatial frequency components visible up to 60 cpd.
Results of Contrast Grating tests can be modeled with a Contrast Sensitivity Function.
CSF defines the bandwidth of vision.
Note that the axis is in the log form.
Human eyes are less sensitive to high frequencies.
Partially applied in image compression.
**Velocity CSF - Motion Sensitivity**

- **Motion Sensitivity**
  - Human eyes are less sensitive to detail moving across retina.
  - Fast moving objects become “blurred”.
- Partially applied in Video compression.

![Graph showing motion sensitivity](image)
Eccentricity CSF – Region of Interest

- Not fully supported in JPEG 2000.
Contents

- Capture and represent digital images
- Color
- Why can Images be Compressed?
  - Spatial Redundancy
  - Temporal Redundancy
  - Spectral Redundancy
  - Psycho-visual Redundancy

- Distortion Measures

Human Visual System
Distortion Measures

- Since most image compression algorithms are lossy, there is a need to develop standard metric to define the quality of the reconstructed image.
- The deviation of the reconstructed image with the original image is called distortion.
- If the reconstructed image is close to the original image, the distortion measure should indicate a low value or alternately, the quality metric must show a higher value.
Mean Square Error (MSE)

- Denoting the pixels of the original image by $x_i$, and the pixels of the reconstructed image by $y_i$, $i = 1, 2, \ldots, N$, where $N$ denotes the number of pixels in the image, the **mean squared error (MSE)** between the two images is defined as

$$MSE = \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2$$

- $\sigma$ is called the **root mean square error (RMSE)**.
Peak Signal-to-Noise Ratio (PSNR)

- Defined as

\[ PSNR = 20 \log_{10} \frac{\max(x_i)}{\sigma} \]

- Some authors use the following to define PSNR:

\[ PSNR = 10 \log_{10} \frac{[\max(x_i)]^2}{\sigma^2} = 20 \log_{10} \frac{\max(x_i)}{\sigma} \]

- Either definition is acceptable since only the relative PSNR values are used in practice.
If the reconstructed image is close to the original image, then $\sigma$ is small and PSNR takes a large value.

PSNR is dimensionless and is expressed in decibel, a unit originally defined to express sound intensity in a logarithmic scale.

Typical PSNR values for images range between 20 to 40, anything above 35 is a very good quality image.
RMSE and PSNR

- Assuming pixel values \([0,255]\).
- \(\text{RMSE}(\sigma)\) of 25.5 results in PSNR of 20.
- \(\sigma = 2.55\) results PSNR of 40.
- If \(\sigma = 0\), PSNR is infinity.
- If \(\sigma = 255\) then PSNR will be negative.
Signal-to-Noise Ratio

\[ SNR = 10 \log_{10} \frac{\sigma^2_x}{\sigma^2} = 10 \log_{10} \frac{1}{N} \sum_{i=1}^{N} x_i^2 \]

\[ = 10 \log_{10} \frac{1}{N} \sum_{i=1}^{N} x_i^2 \]