Overview of JPEG

What is JPEG?
- Works with colour and greyscale images, e.g., satellite, medical, ...

Motivation

- The compression ratio of lossless methods (e.g., Huffman, Arithmetic, LZW) is not high enough for image and video compression, especially when the distribution of pixel values is relatively flat.
- JPEG uses transform coding; it is largely based on the following observations:
  - Observation 1: A large majority of useful image contents change relatively slowly across images, i.e., it is unusual for intensity values to alter up and down several times in a small area, for example, within an 8 x 8 image block. Translate this into the spatial frequency domain, it says that, generally, lower spatial frequency components contain more information than the high frequency components which often correspond to less useful details and noises.
  - Observation 2: Psychophysical experiments suggest that humans are more receptive to loss of higher spatial frequency components than loss of lower frequency components.

JPEG Decoding

How to decode the JPEG Pictures?
Reverse the Process!

- i.e. The flow arrow will point backwards
- This will imply that the transform function needs to have an equivalent inverse.

Image Compression Process

- There are FOUR (4) modes:
  - Lossy sequential DCT (supported by all)
  - Extended lossy DCT (further enhancements)
  - Lossless mode
  - Hierarchical mode
- The baseline process takes the following techniques: Block, MCU, FDCT, RLE and Huffman. We shall discuss the details later.
Image Preparation

- A source image consists of 1..255 components/planes
- Color has its 3 components with equal resolution

Sampling factors

- X, Y are maximum value
- Hᵢ = Horizontal sampling ratio, Vᵢ = Vertical sampling ratio
- H_max = Horizontal sampling ratio, V_max = Vertical sampling ratio

Example:

Given X = 512, Y = 512, H_max = 4, V_max = 2

Level 0: H₀ = 4, V₀ = 1
  X₀ = X * H₀ / H_max = 512 * 4 / 4 = 512
  Y₀ = Y * V₀ / V_max = 512 * 1 / 2 = 256

Level 1: H₁ = 2, V₁ = 2
  X₁ = X * H₁ / H_max = 512 * 2 / 4 = 256
  Y₁ = Y * V₁ / V_max = 512 * 1 / 2 = 256

Level 2: H₂ = 1, V₂ = 1
  X₂ = X * H₂ / H_max = 512 * 1 / 4 = 128
  Y₂ = Y * V₂ / V_max = 512 * 1 / 2 = 256

Interleaving of data units

- Non-interleaved order of data

Strictly Left-to-Right and Top-to-Bottom
Interleaving of data units

MCU (Minimum Coded Units) = combination of interleaved data units of different components.

<table>
<thead>
<tr>
<th></th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

d1: H=2, V=2;  d2: H=2, V=1;  d3: H=1, V=2;  d4: H=1, V=1

Interleaving of data units

The MCU in this example are:

MCU1 = d200d201d210d211d300d301d400d401

MCU2 = d100d101d110d111d302d303d400d411

Note: The data units from the same component are shown with the same color.

Major Steps

1. DCT (Discrete Cosine Transformation)
2. Quantization
3. Zigzag Scan
4. DPCM on DC component
5. RLE on AC Components
6. Entropy Coding

Discrete Cosine Transform (DCT)

From spatial domain to frequency domain

Discrete Cosine Transform (DCT):

\[
F(u,v) = \frac{1}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} A(s)A(t) \cos \left( \frac{(2s+1)\pi u}{16} \right) \cos \left( \frac{(2t+1)\pi v}{16} \right) f(s,t)
\]

\[
A(s) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } s = 0 \\ 1 & \text{otherwise} \end{cases}
\]

Inverse Discrete Cosine Transform (IDCT):

\[
F(s,t) = \sum_{u=0}^{7} \sum_{v=0}^{7} A(u)A(v) \cos \left( \frac{(2u+1)\pi s}{16} \right) \cos \left( \frac{(2v+1)\pi t}{16} \right) \frac{1}{4} F(u,v)
\]

\[
A(u) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u = 0 \\ 1 & \text{otherwise} \end{cases}
\]
Discrete Cosine Transform (DCT)

Example:

![DCT Example Matrix]

The 64 (8 x 8) DCT basis functions:

![DCT Basis Functions]

Factoring reduces problem to a series of 1D DCTs:

\[
F[u, v] = \frac{1}{\sqrt{N}} \sum_j A[j] \cos \left( \frac{2\pi}{N} (u+1) \frac{j+1}{2} \right) G[v]
\]

Most software implementations use fixed point arithmetic. Some last implementations approximate coefficients so all multiplies are shifts and adds.

Quantization

Formula:

\[
F'[u, v] = \text{round} \left( \frac{F[u, v]}{q[u, v]} \right)
\]

Why?

To reduce number of bits per sample

Example: 101101 = 45 (6 bits).

\[q[u, v] = 4\] Truncate to 4 bits: 1011 = 11.

Quantization error is the main source of the Lossy Compression.

Uniform Quantization

Each \(F[u,v]\) is divided by the same constant \(N\).

Non-uniform Quantization using Quantization Tables

Eye is most sensitive to low frequencies (upper left corner), less sensitive to high frequencies (lower right corner).

The Luminance Quantization Table \(q(u, v)\)

<table>
<thead>
<tr>
<th>u</th>
<th>16</th>
<th>11</th>
<th>10</th>
<th>16</th>
<th>24</th>
<th>40</th>
<th>51</th>
<th>61</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>12</td>
<td>14</td>
<td>19</td>
<td>26</td>
<td>58</td>
<td>60</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td></td>
<td>24</td>
<td>40</td>
<td>57</td>
<td>69</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>22</td>
<td>29</td>
<td>51</td>
<td>87</td>
<td>80</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>22</td>
<td>37</td>
<td>56</td>
<td>68</td>
<td>109</td>
<td>103</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>35</td>
<td>55</td>
<td>64</td>
<td>81</td>
<td>104</td>
<td>113</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>64</td>
<td>78</td>
<td>87</td>
<td>103</td>
<td>121</td>
<td>120</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>92</td>
<td>95</td>
<td>98</td>
<td>112</td>
<td>100</td>
<td>103</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>
Quantization

The Chrominance Quantization Table \( q(u, v) \)

<table>
<thead>
<tr>
<th></th>
<th>17</th>
<th>18</th>
<th>24</th>
<th>47</th>
<th>99</th>
<th>99</th>
<th>99</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>21</td>
<td>26</td>
<td>66</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>24</td>
<td>26</td>
<td>56</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

The numbers in the above quantization tables can be scaled up (or down) to adjust the so called quality factor. Custom quantization tables can also be put in image/scan header.

Zig-zag Scan

Why?
1. To group low frequency coefficients in top of vector.
2. Maps 8 x 8 to a 1 x 64 vector.

Differential Pulse Code Modulation (DPCM) on DC component

- DC component is large and varied, but often close to previous value.
- Encode the difference from previous 8 x 8 blocks \( \rightarrow \) DPCM

\[
\text{DIFF}(i-1,i) = D_{C_i} - D_{C_{i-1}}
\]

Run Length Encode (RLE) on AC components

- 1 x 64 vector has lots of zeros in it
- Keeps \( \text{skip} \) and \( \text{value} \), where \( \text{skip} \) is the number of zeros and \( \text{value} \) is the next non-zero component.
- Send \((0,0)\) as end-of-block sentinel value.

Entropy Coding

- Categorize DC values into \( \text{SIZE} \) (number of bits needed to represent) and actual bits.

<table>
<thead>
<tr>
<th>Size</th>
<th>BCD</th>
<th>Diff DC Coefficient Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0000</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0001</td>
<td>-1, 1</td>
</tr>
<tr>
<td>2</td>
<td>0010</td>
<td>-3, -2, 2, 3</td>
</tr>
<tr>
<td>3</td>
<td>0011</td>
<td>-7, -6, 6, 7</td>
</tr>
<tr>
<td>4</td>
<td>0100</td>
<td>-15, -14, 14, 15</td>
</tr>
<tr>
<td>5</td>
<td>0101</td>
<td>-31, -30, 30, 31</td>
</tr>
<tr>
<td>6</td>
<td>0110</td>
<td>-63, -62, 62, 63</td>
</tr>
<tr>
<td>7</td>
<td>0111</td>
<td>-127, -126, 126, 127</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
<td>-255, -254, 254, 255</td>
</tr>
<tr>
<td>9</td>
<td>1001</td>
<td>-511, -506, 506, 511</td>
</tr>
<tr>
<td>10</td>
<td>1010</td>
<td>-1023, -1012, 1012, 1023</td>
</tr>
</tbody>
</table>

Example: If DC value is 4, 3 bits are needed. Send off \( \text{SIZE} \) as Huffman symbol, followed by actual 3 bits.

Entropy Coding

- For AC components two symbols are used:
  Symbol_1: \((\text{skip}, \text{SIZE})\),
  Symbol_2: actual bits.
- Symbol_1 \((\text{skip}, \text{SIZE})\) is encoded using the Huffman coding, Symbol_2 is not encoded.
- Huffman Tables can be custom (sent in header) or default.
A Glance at the JPEG Bitstream

A "Frame" is a picture, a "scan" is a pass through the pixels (e.g., the red component), a "segment" is a group of blocks, a "block" is an 8 x 8 group of pixels.

Frame header:
- sample precision
- (width, height) of image
- number of components
- unique ID (for each component)
- horizontal/vertical sampling factors (for each component)
- quantization table to use (for each component)

Scan header
- Number of components in scan
- component ID (for each component)
- Huffman table for each component (for each component)

Misc. (can occur between headers)
- Quantization tables
- Huffman Tables
- Arithmetic Coding Tables
- Comments
- Application Data

The Four JPEG Modes (Revisited)

The four JPEG modes:
- Sequential Mode
- Loss-less Mode
- Progressive Mode
- Hierarchical Mode

Note: In "Motion JPEG", Sequential JPEG is applied to each image in a video.

Sequential Mode JPEG

Each image component is encoded in a single left-to-right, top-to-bottom scan. Baseline Sequential Mode, the one that we described above, is a simple case of the Sequential mode:
- It supports only 8-bit images (not 12-bit images)
- It uses only Huffman coding (not Arithmetic coding)

Lossless Mode JPEG

A special case of the JPEG where indeed there is no loss. Its block diagram is as below:

- It does not use DCT-based method! Instead, it uses a predictive (differential coding) method:
  - A predictor combines the values of up to three neighbouring pixels (not blocks as in the Sequential mode) as the predicted value for the current pixel, indicated by "X" in the figure on next page.
  - The encoder then compares this prediction with the actual pixel value at the position "X", and encodes the difference (prediction residual) losslessly.

Lossless Mode JPEG

It can use any one of the following seven predictors:

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>A + B - C</td>
</tr>
<tr>
<td>5</td>
<td>A + (B - C) / 2</td>
</tr>
<tr>
<td>6</td>
<td>B + (A - C) / 2</td>
</tr>
<tr>
<td>7</td>
<td>(A + B) / 2</td>
</tr>
</tbody>
</table>
Lossless Mode JPEG

- Since it uses only previously encoded neighbours, the very first pixel \( I(0, 0) \) will have to use itself. Other pixels at the first row always use \( P_1 \), at the first column always use \( P_2 \).
- Effect of Predictor (test result with 20 images is shown on next page)
- Note: “2D” predictors (4-7) always do better than “1D” predictors.

Progressive Mode JPEG

- Goal: display low quality image and successively improve.
- Two ways to successively improve image:
  1. Spectral selection: Send DC component and first few AC coefficients first, then gradually some more ACs.
  2. Successive approximation: send DCT coefficients MSB (most significant bit) to LSB (least significant bit). (Effectively, it is sending quantized DCT coefficients first, and then the difference between the quantized and the non-quantized coefficients with finer quantization stepsize.)

Hierarchical Mode JPEG

- Down-sample by factors of 2 in each dimension, e.g., reduce 640 x 480 to 320 x 240
- Code smaller image using another JPEG mode (Progressive, Sequential, or Lossless).
- Decode and up-sample encoded image
- Encode difference between the up-sampled and the original using Progressive, Sequential, or Lossless.
- Can be repeated multiple times.
- Good for viewing high resolution image on low resolution display.