JPEG

Agenda

- JEPG is lossy
- JPEG is an example of transform encoding
- Images are pre-processed: colour conversion, chroma-subsampling, 8x8 blocks
- Pixel data is converted to frequency coefficients
- Compression is done by quantisation of the frequency values
- Lossless encoding is also used to compress further the frequency values

JPEG

JPEG = Joint Photographic Experts Group

Standardised set of algorithms so that the various implementations all produce files in the same format

- Standard for compressing and decompressing still images
- Based on DCT and applies to colour and grey-scaled still images
    - Independent of image sizes
    - Applicable to any aspect ratio (W/H)
    - Colour space and number of colours independent
    - Image content of any complexity, with any spatial and statistical properties
    - Run on many standard processors
    - Compression ratio dictated by user or application

JPEG is lossy, and it can cause an issue called **artifacting**, in which parts of the image because noticeably **blocky**

Transform Encoding (Lossless)

Main idea: changing the representation (domain) of data can sometimes make it possible to extract details that won't be missed because they are beyond the acuity of human perception.

Two of the most commonly used transforms in digital media are DCT and DFT

High frequency components correspond to quick fluctuations of colour in a short space—changes that aren't easy for the human eye to see

Once the high frequency components have been seperated out, remove them.

Frequency in Digital Images

frequency (digital image): the rate at which color values change

The Discrete Cosine Transform

DCT tells how to transform an image from the spatial domain (i.e. pixel values), to the frequency domain (coefficients by which the frequency components should be multiplied)

The DCT is generally applied to 8x8 pixel blocks

KEY EQUATION

Let $f(r, s)$ be the pixel value at row r and column s of a bitmap. $F(u, v)$ is the coefficient of the frequency component at (u, v), where $0 \leq r, u \leq M - 1$ and $0 \leq s, v \leq N - 1$.

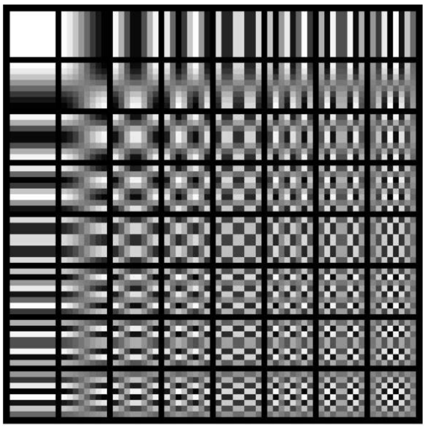
$$F(u, v) = \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} \frac{2C(u)C(v)}{\sqrt{MN}} f(r, s) \cos\left(\frac{(2r + 1)u\pi}{2M}\right) \cos\left(\frac{(2s + 1)v\pi}{2N}\right)$$

$$\text{where } C(\delta) = \frac{\sqrt{2}}{2} \text{ if } \delta = 0 \text{ otherwise } C(\delta) = 1$$

Base Frequencies

Each graph at position (u,v) represents the function:

$$\cos\left(\frac{(2r + 1)u\pi}{16}\right) \cos\left(\frac{(2s + 1)v\pi}{16}\right)$$



DCT takes a bitmap image in the form of a matrix of color values – $\boldsymbol{f}$ (*row*, *column*) – and returns the frequency components of the bitmap in the form of a matrix of coefficients – $\boldsymbol{F}$ (u, v)

The coefficients give the amplitudes of the frequency components

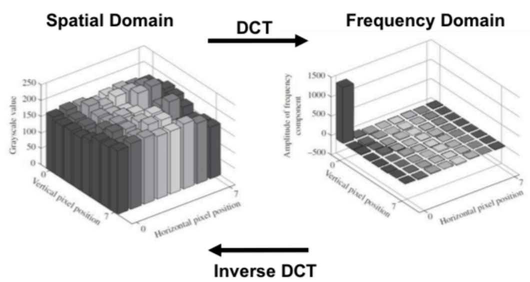
DC component: The first element $\boldsymbol{F}(0,0)$

AC component: All the other components

A negative coefficient amounts to adding the inverted waveform.

The discrete cosine transform is invertible

(Note: the encoding process is **lossless**)



Spatial Domain

255	255	255	255	255	255	159	159
255	0	0	0	255	255	159	159
255	0	0	0	255	255	255	255
255	0	0	0	255	255	255	255
255	255	255	255	255	255	100	255
255	255	255	255	255	255	100	255
255	255	255	255	255	255	100	255
255	255	255	255	255	255	100	255

DCT ↓ ↑ Inverse DCT

1628	−61	39	234	173	−128	171	22
−205	−163	74	222	1	74	−30	111
81	150	−95	−82	−42	−11	−6	−53
188	231	−135	−188	−53	−36	−2	−103
96	71	−42	−78	−32	2	−17	−32
25	−42	25	3	−14	27	−26	19
70	15	−6	−51	−17	7	−16	−13
94	72	−38	−87	−18	−14	−4	−40

Frequency Domain

Steps of JPEG compression

```
algorithm jpeg
/*Input: A bitmap image in RGB mode. Output: The same image, compressed.*/
{
(Convert image to Y'CbCr and do chroma sub-sampling)
Divide image into 8 x 8 pixel blocks
Use Discrete Cosine Transform (DCT) to transform the pixel data from the spatial domain to the frequency domain
Quantise frequency values
DPCM and zigzag order
Do run-length encoding
Do entropy encoding (e.g., Huffman)
}
```

Images are pre-processed

Colour Conversion: RGB to Y'CbCr

$$
\begin{aligned}
Y' &= 0 + (0.299 \cdot R'_D) + (0.587 \cdot G'_D) + (0.114 \cdot B'_D) \\
C_B &= 128 - (0.168736 \cdot R'_D) - (0.331264 \cdot G'_D) + (0.5 \cdot B'_D) \\
C_R &= 128 + (0.5 \cdot R'_D) - (0.418688 \cdot G'_D) - (0.081312 \cdot B'_D)
\end{aligned}
$$

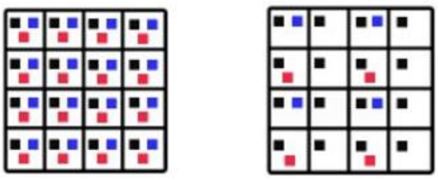
And back:

$$
\begin{aligned}
R'_D &= Y' + 1.402 \cdot (C_R - 128) \\
G'_D &= Y' - 0.344136 \cdot (C_B - 128) - 0.714136 \cdot (C_R - 128) \\
B'_D &= Y' + 1.772 \cdot (C_B - 128)
\end{aligned}
$$

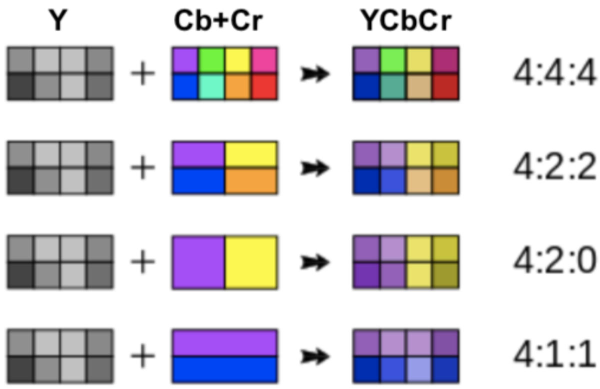
Chroma Sub-Sampling

The human eye is more sensitive to changes in light (i.e., luminance) than in colour (i.e., chrominance)

Chroma sub-sampling (also called chrominance downsampling) is a process of throwing away some of the bits used to represent colour information



choose to save only one Cb value and one Cr value but four Y values for every four pixel values



Y Cb,Cr	Y	Y	Y
Y Cb,Cr	Y	Y	Y
Y Cb,Cr	Y	Y	Y
Y Cb,Cr	Y	Y	Y

4:1:1

Y Cb,Cr	Y	Y Cb,Cr	Y
Y	Y	Y	Y
Y Cb,Cr	Y	Y Cb,Cr	Y
Y	Y	Y	Y

4:2:0

Y Cb,Cr	Y	Y Cb,Cr	Y
Y Cb,Cr	Y	Y Cb,Cr	Y
Y Cb,Cr	Y	Y Cb,Cr	Y
Y Cb,Cr	Y	Y Cb,Cr	Y

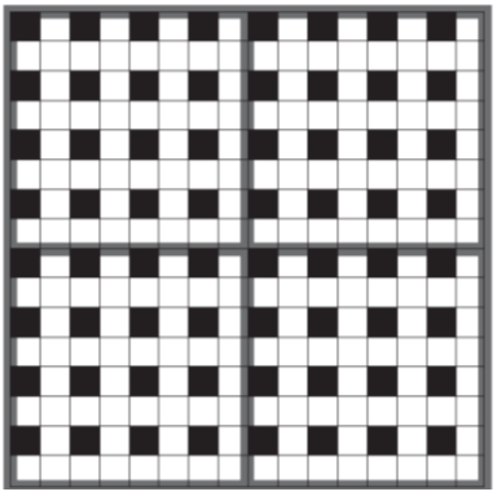
4:2:2

In JEPG, we use 4:2:0 form for Chroma sub-sampling

Blocks

• With YCbCr colour mode, we begin by dividing the image into 16x16 pixel macroblocks.

• Then, with 4 : 2 : 0 chroma downsampling, we get four blocks of 8x8 Y data for every one block of 8x8 Cb and one block of 8x8 Cr data.



What is the compression rate achieved?

Original: $16 \cdot 16 \cdot 3 = 768$

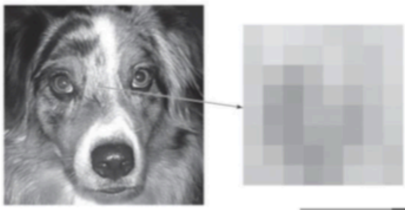
New: $16 \cdot 16 \ + \ 8 \cdot 8 \cdot 2 = 384$

Compression rate = 768 : 384 = 2 : 1

Pixel data is converted to frequency coefficients

Shifting Pixel Values

On an intuitive level, shifting the values by −128 is like looking at the image function as a waveform that cycles through positive and negative values



94	103	101	96	88	85	92	96
88	101	89	87	93	82	81	95
83	74	55	70	90	79	81	93
86	52	36	60	75	65	77	89
81	43	38	62	62	50	71	87
78	49	38	51	52	50	71	82
84	69	45	38	51	70	78	75
80	80	67	46	56	82	86	78

585.7500	−24.5397	59.5959	21.0853	25.7500	−2.2393	−8.9907	1.8239
78.1982	12.4534	−32.6034	−19.4953	10.7193	−10.5910	−5.1086	−0.5523
57.1373	24.829	−7.5355	−13.3367	−45.0612	−10.0027	4.9142	−2.4993
−11.8655	6.9798	3.8993	−14.4061	8.5967	12.9151	−0.3122	−0.1844
5.2500	−1.7212	−1.0824	−3.2106	1.2500	9.3595	2.6131	1.1199
−5.9658	−4.0865	7.6451	13.0616	−1.1927	1.1782	−1.0733	−0.5631
−1.2074	−5.7729	−2.0858	−1.9347	1.6173	2.6671	−0.4645	0.6144
0.6362	−1.4059	−0.7191	1.6339	−0.1438	0.2755	−0.0268	−0.2255

Quantise frequency values (*)

DCT coefficients matrix

- Values are arranged from lowest frequencies to highest frequencies
  - Lowest frequencies represent average value for the block
  - Highest frequency represent fine detail (they can be dropped)
    - Eye unable to perceive brightness levels above or below thresholds

- Gentle gradation of brightness of colour are more important to the eye than abrupt changes

Quantisation = Division

Quantisation involves dividing each frequency coefficient by an integer and **rounding off**

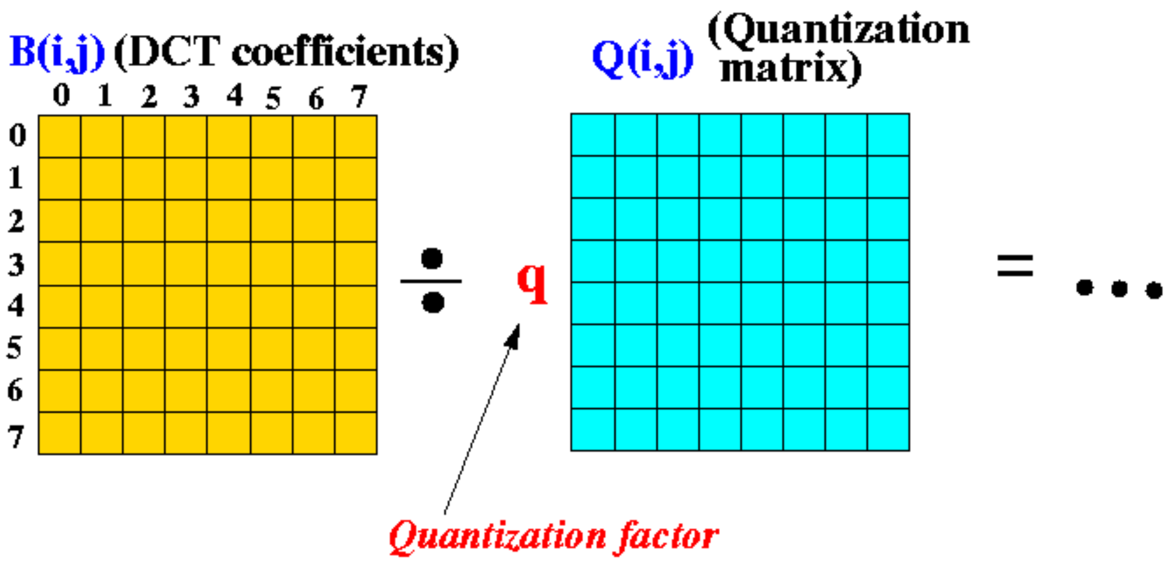
Rounding during quantisation makes JPEG lossy

The coefficients for high-frequency components are typically small, so they often round down to 0—which means, in effect, that they are thrown away

A quantisation table will be stored with the compressed image

Not every value in the matrix needs to be divided by the same integer

$$F^Q(u,v) = Integer\ Round\left(\frac{F(u,v)}{Q(u,v)}\right)$$



Quantisation factor: can be used to increase or decrease the quantisation value.

Higher q value means greater compression and less quality; Lower q value means less compression and greater quality.

Note: In the top left area the quantisation value are very small, which indicates that the coefficient of that area in DCT frquency matrix should not be lost as they contain fundamental information about image

Question 1

- What known limitations of the human visual system are used in JPEG to achieve compression?
- Is JPEG always lossy?
- Ingeneral, does JPEG achieve greater compression rates on colour images or on gray scale images?

1) Our eyes are more sensative to the gentle gradation of brightness of color than the abrupt changes

(fine variations of colour are not easily perceivable and so can be removed to achieve compression)

2) Yes
3) On color images (Chroma sub-sampling)
Question 2
In JPEG, what is the effect of using quantisation tables that contain small values?
Solution: achieves little compression and preserves much quality

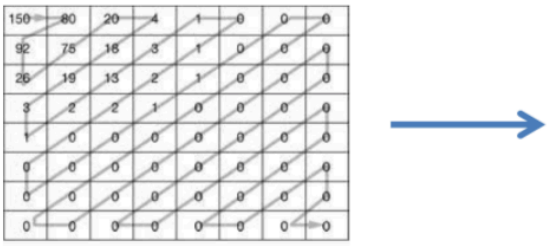
Lossless encoding is also used to compress further the frequency values

DPCM (differential pulse code modulation)

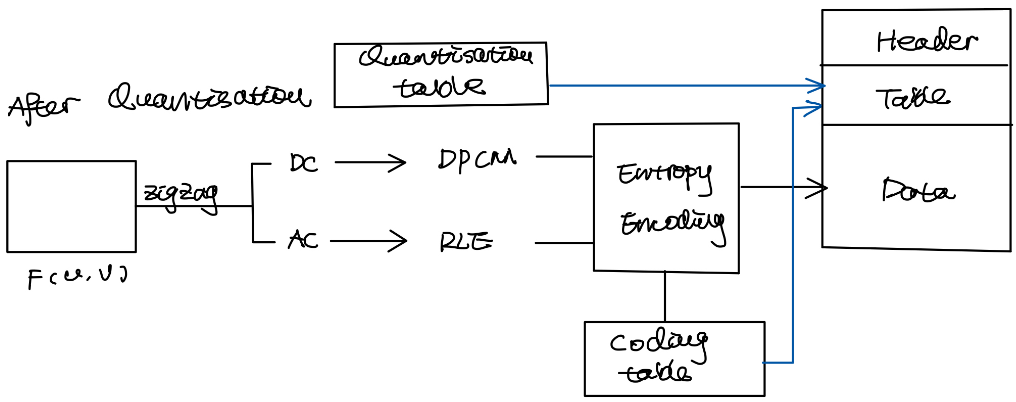
DPCM is a compression technique that works by recording the difference between consecutive values rather than the actual values
It is effective in cases where consecutive values don't change very much because fewer bits are needed to record the change than the value itself
In JPEG, the upper leftmost value in an block (the DC component) is stored as the difference from the DC component in the previous block

Zigzag Ordering

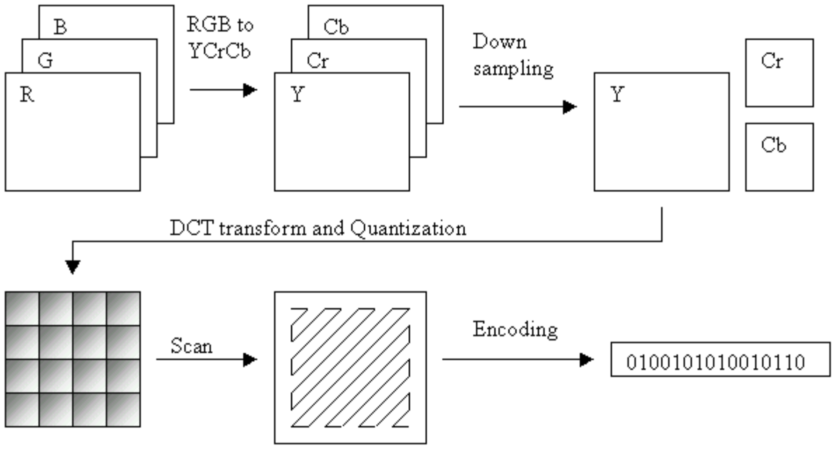
The zigzag reordering sorts the values from low-frequency to high frequency components.
- The high-frequency coefficients are grouped together at the end.
- If many of them round to zero after quantisation, run-length encoding is even more effective



150	80	92	26	75	20	4	18	19	3	1	2	13	3	1	0	1	2
2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0								



JPEG overview



Question
In JPEG, which of the following steps contribute the most to compression?
Answer: DCT coefficient quantization

JPEG encoding

- After these steps have been performed, the compressed file is put into a standardised format that can be recognised by the decompressor
- A header contains global information such as: the type of file, the width and height, one or more quantizsation tables, Huffman code tables, and an indication of any pixel-padding
- An alternative JPEG compression method is called **JPEG2000**, noted for its high compression rate and good quality, without some of the blocky artifacts of standard JPEG. This method represents digital image data as wavelets as an alternative to the DCT

Exercise 1
Consider the 4x4 block of pixels in Table 1, where:
r corresponds to the fully saturated bright red colour; g to the fully saturated bright green colour; b to the fully saturated bright blue colour; w to white; bk to black; and gy to mid-grey.

r	b	g	b
gy	r	gy	gy
gy	w	r	w
gy	bk	bk	r

Table 1

Extract from the block of pixels shown in Table 1, the 4x4 matrix of values, comprised between 0 and 255, that correspond to the R channel.

1000	800	100	40
600	400	30	10
150	80	10	3
20	10	5	2

Table 2

10	20	50	99
20	50	99	99
50	80	99	99
99	99	99	99

Table 3

255	0	0	0
127	255	127	127
127	255	255	255
127	0	0	255

Exercise 2
Consider that the matrix shown in Table 2 is a matrix of DCT coefficients.

- Apply quantisation to this matrix, using the quantisation matrix shown in Table 3.
- Arrange the quantised values in zig zag order and do Run Length encoding.
- To obtain a better quality image after compression, what would you change to the quantisation matrix of Table 3?

Solution:
(1)

100	40	2	0
30	8	0	0
3	1	0	0
0	0	0	0

(2) Zig-zag order: 100 40 30 3 8 2 0 0 1 0 0 0 0 0 0 0
With RLE: (100,1) (40,1) (30,1) (3,1) (8,1) (2,1) (0,2) (1,0) (0,7)
(3) To obtain a better quality, decrease the quantisation value in Table 3

Questions in test

c) Consider the block diagram shown in Figure 2 and answer the questions below.

[10 marks]

i) Describe what is contained inside the block marked with the label "A".

(2 marks)

ii) What step of the JPEG compression process is marked with the label "B"? Explain this step.

(3 marks)

iii) Explain the use of quantisation tables in the JPEG compression process.

(3 marks)

iv) Which phenomenon of human vision is exploited in chroma sub-sampling?

(2 marks)

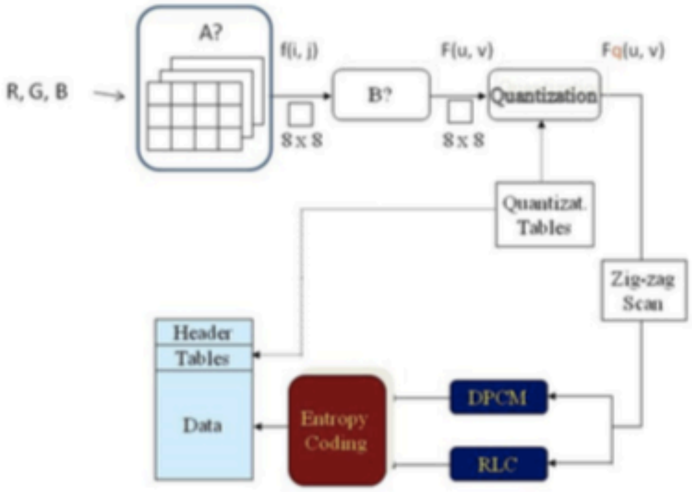


Figure 2: JPEG block diagram

Solution:

(1) It contains the pixel value respectively for Y, Cb and Cr (After chroma sub-sampling)

(2) DCT transformation. In this step, first, shifting pixel values by -128, then applying DCT transform, an 8*8 block of DCT coefficients are obtained. The purpose of this step is to extract the high frequency components, which are not easy to be perceived by human, from the rest.

(3) The quantisation table is used to store different quantisation value for different DCT coefficient. Since human is more sensitive to the gentle gradation of brightness of colous, in the quantisation table, small values are set to low frequency components, which we do not want to loss, large values are set to high frequency components.

(4) The Human eye is more sensitive to changes in brightness than in chrominance

ii) Why is JPEG said to be lossy? Give detailed explanations.
Solution: In the whole JPEG process, their are two places that contributes to JPEG's lossy characteristic. First, when pre-processing the image, chroma sub-sampling is applied to down-sample the YCbCr values. In JPEG we apply the chroma sub-sampling with form 4:2:0, which means in each 4 pixels, we preserve all 4 Y value whereas only 1 Cb and 1 Cr value (thrown away some chrominance information), since human eye is more sensitive to changes in brightness than Chrominance. Second, when quantisizing the DCT coefficients, the DCT coefficients are devided by the quantisation values in the quantisation tables and round off. Actually, it is rounding off makes JPEG lossy, since some of the high frequency components with small values will be rounded to 0 and loss information.

iii) In general, does JPEG achieve greater compression rates on colour images or on gray scale images? Explain your answer
Solution: It achieves greater compression rates on colour images. This is because chroma sub sampling, which only applies to colour images, can achieve extra compression. It works by removing some of the colour information in the image data while preserving all the brightness information. It does this after converting the mage from RGB to YUV.