JPEG

JPEG

- Joint Photographic Expert Group
- Voted as international standard in 1992
- Works with color and grayscale 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.

Features

- Sequential encoding
  - Block-by-block encoding in scan order
- Progressive encoding
  - geared at progressive transmission, or successive approximations: to achieve higher order resolution pictures it uses more and more DCT coefficients, or more and more bits/coefficient
  - Note: Different quality resolutions are *not* available on the same bitstream
- Hierarchical encoding
  - A lower resolution image is encoded first, upsampled and interpolated to predict the full resolution and the prediction error is encoded with one of the other JPEG coders (pyramidal coder using the JPEG coder)
- Lossless encoding using a different strategy
  - Predictive coding on a neighborhood of three samples is used instead of DCT

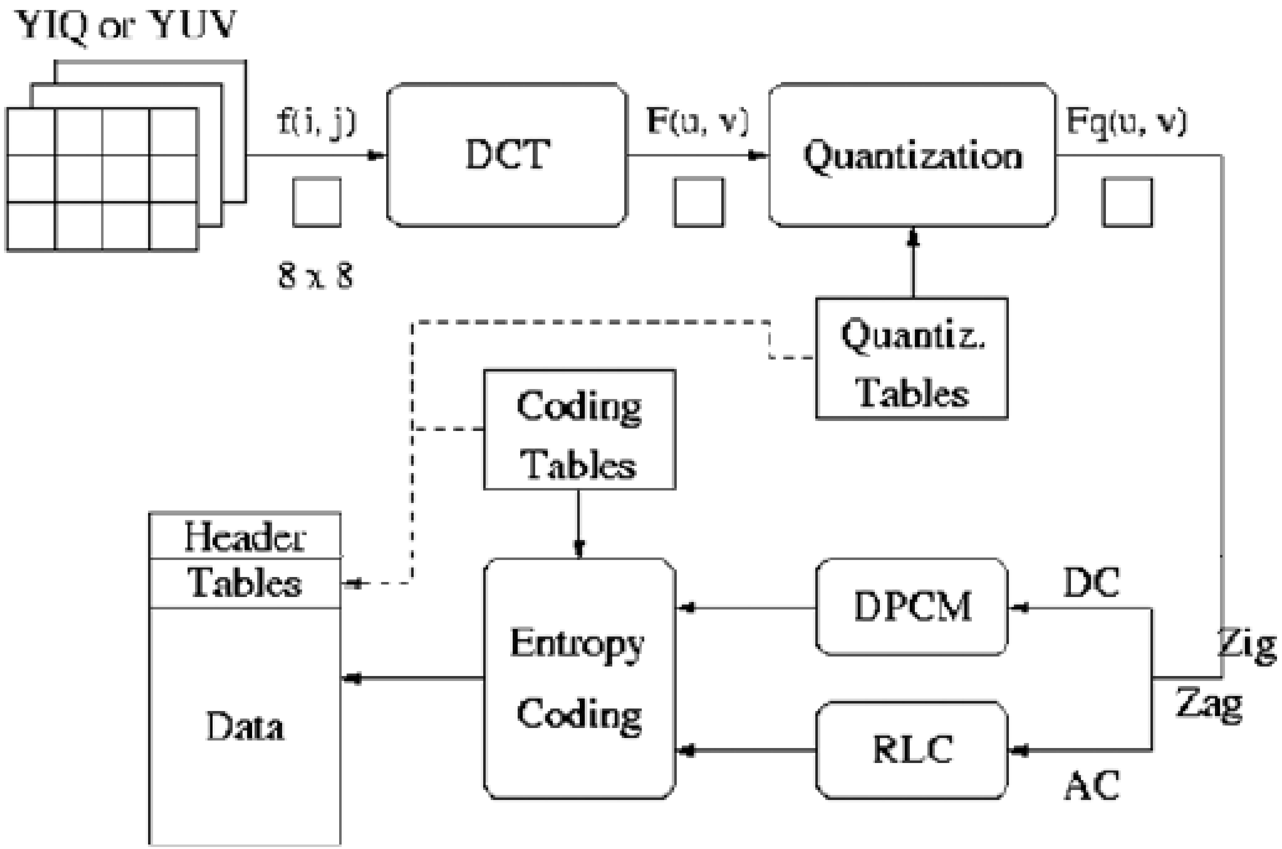
Structure

- Block-based transform coding using DCT
- Rationale
  - 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.

Main steps

- Colorspace conversion
- DCT (Discrete Cosine Transformation)
- Quantization
- Zigzag Scan
- DPCM on DC component
  - DPCM: differential pulse code modulation
  - DC: zero-frequency component
- RLE on AC Components
  - Run Length Encoding
  - AC: other spectral components
- Entropy Coding

Block diagram

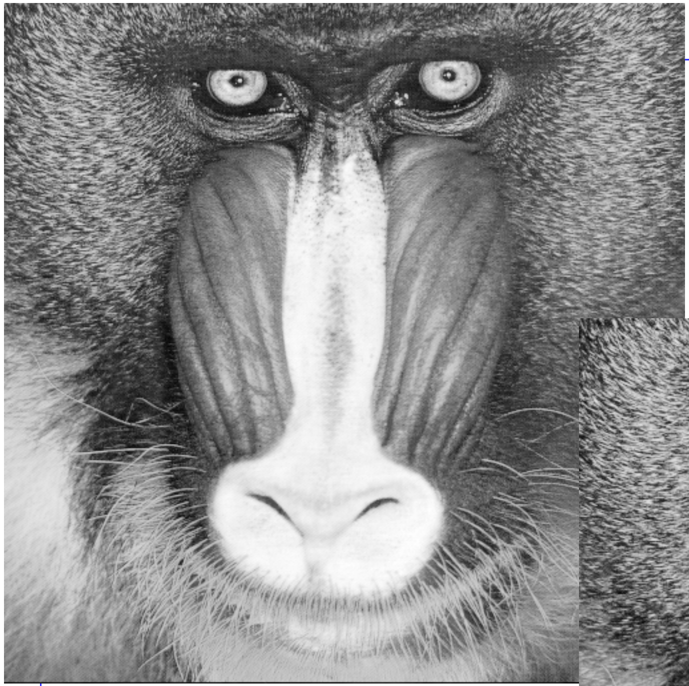


Color

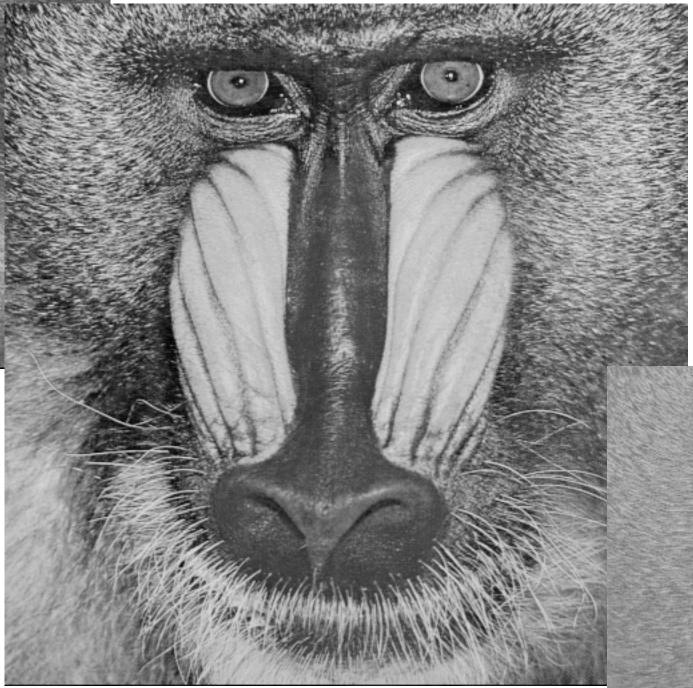
- Images are converted from RGB to YCbCr
- Cb and Cr components are downsamples
  - Possible formats:
    - 4:4:4
    - 4:2:2
    - 4:2:0

```
Y  =           0.299*R +      0.587*G  +      0.114*B
Cb = 128 - 0.168736*R -  0.331264*G  +        0.5*B
Cr = 128 +       0.5*R -  0.418688*G  - 0.081312*B
```

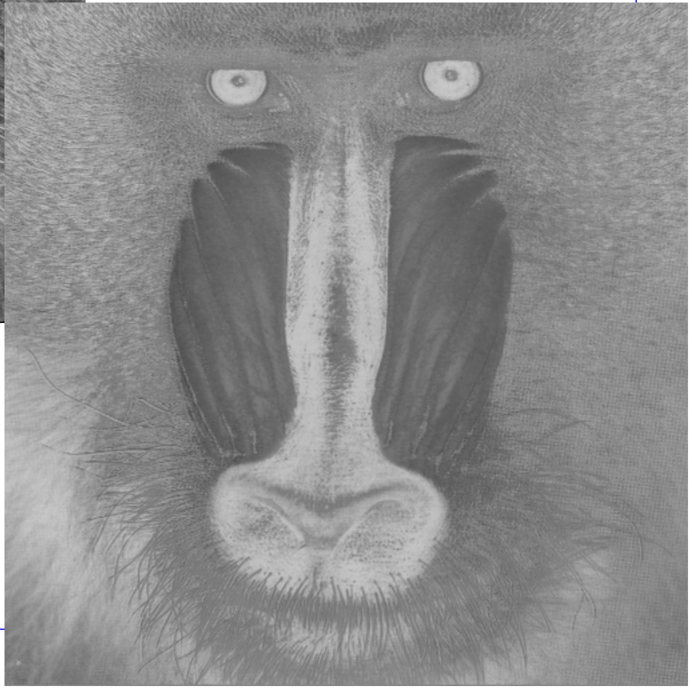
RGB



rho(red-green)=0.3565

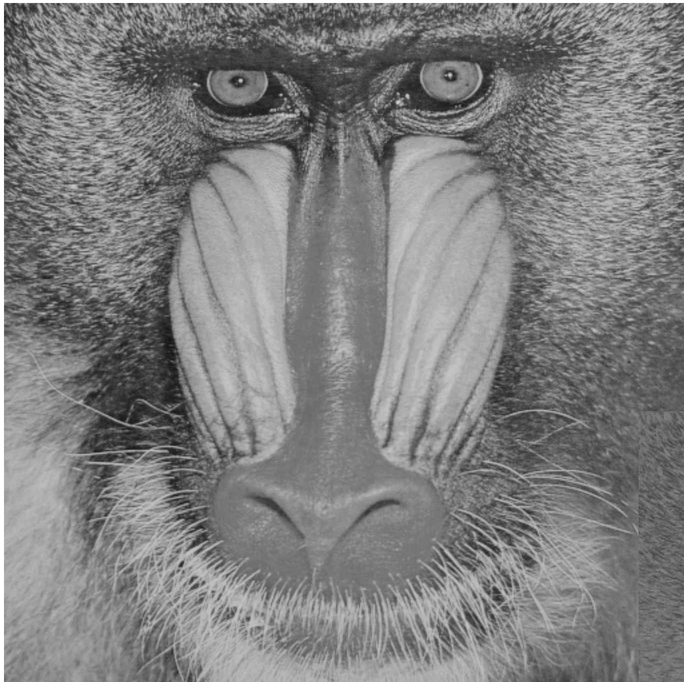


rho(green-blu)=0.85

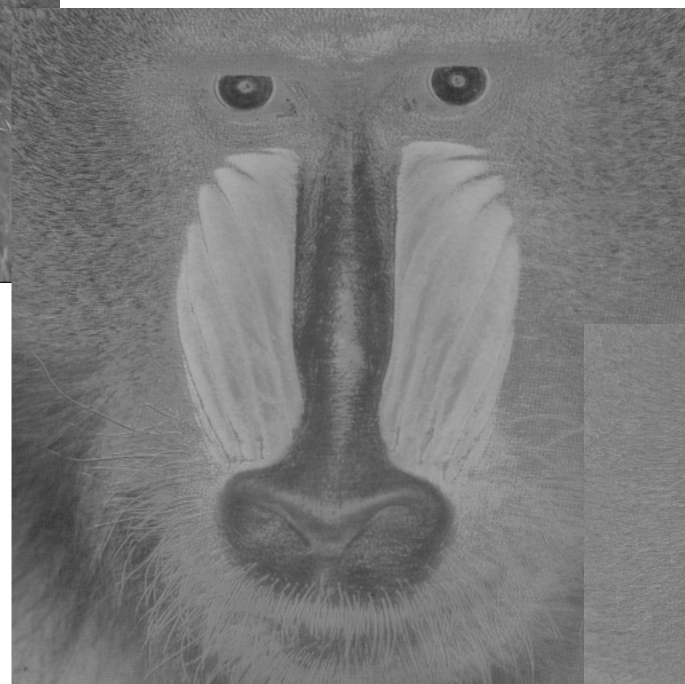


YCbCr

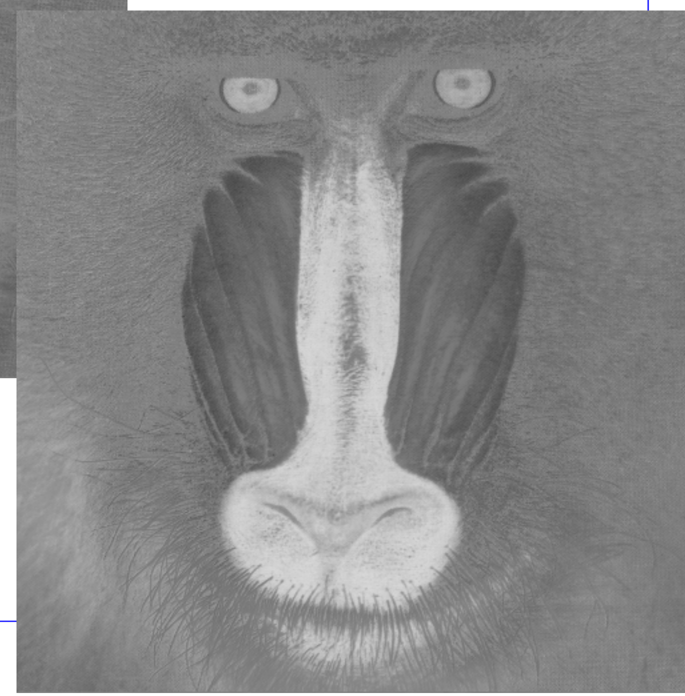
rho(YCr)=-0.1474



rho(YCb)=0.09



rho(YCb)=-0.72



2D DCT

- Separable product (equivalently, a composition) of DCTs along each dimension

$$X_{k_1,k_2} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1\right] \cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2\right]$$

- *Row-column* algorithm
- The inverse of a multi-dimensional DCT is just a separable product of the inverse(s) of the corresponding one-dimensional DCT
  - e.g. the one-dimensional inverses applied along one dimension at a time

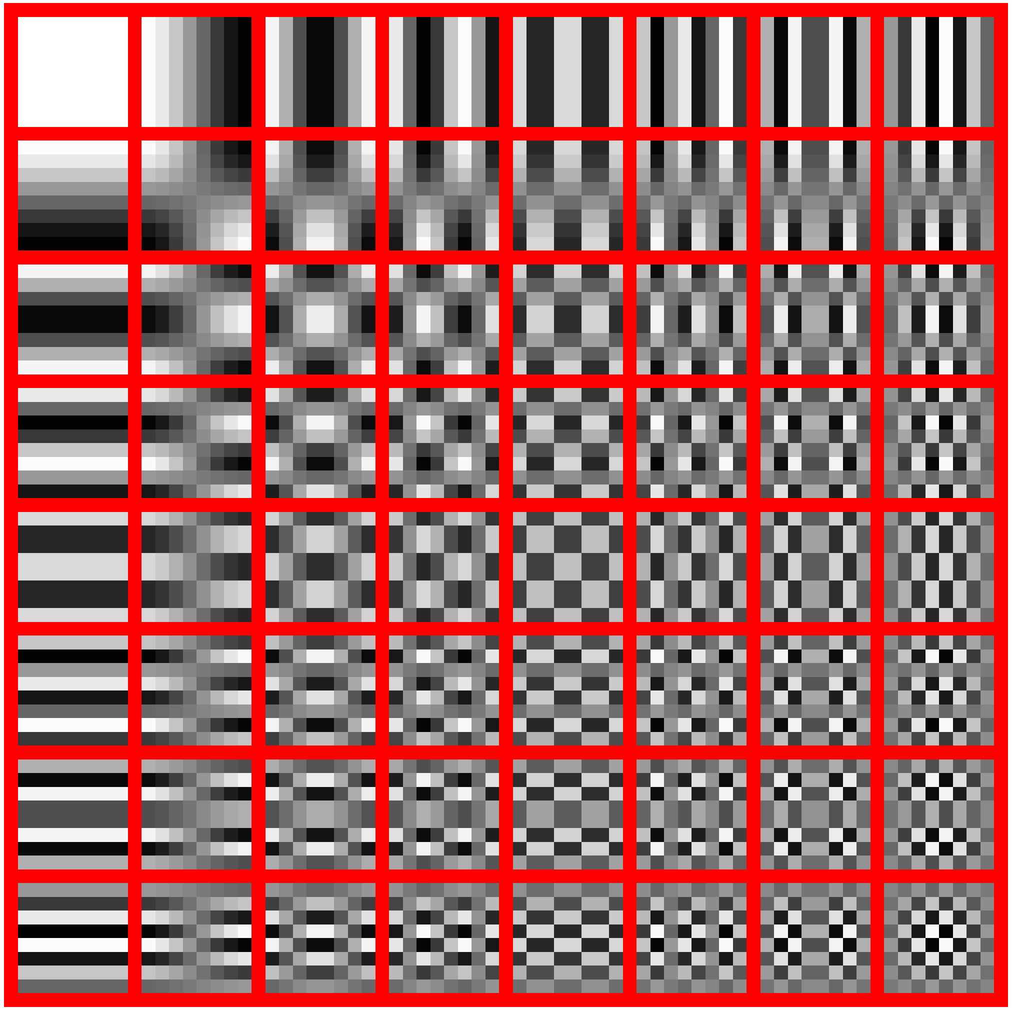
DCT: basis functions

Block-based transform

Block size N1=N2=8

This has proved to be a good compromise between coding efficiency (large blocks) and avoidance of blocking effects (small blocks).

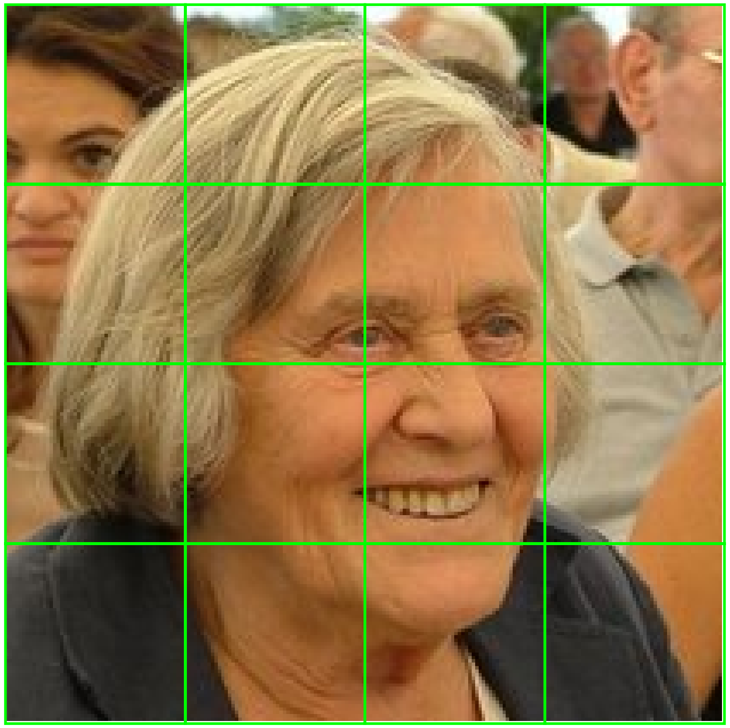
This applies to typical target images: 512x512, target rate: 0.5bpp, bit depth=8 or 12 bits



JPEG preview

Block-based DCT

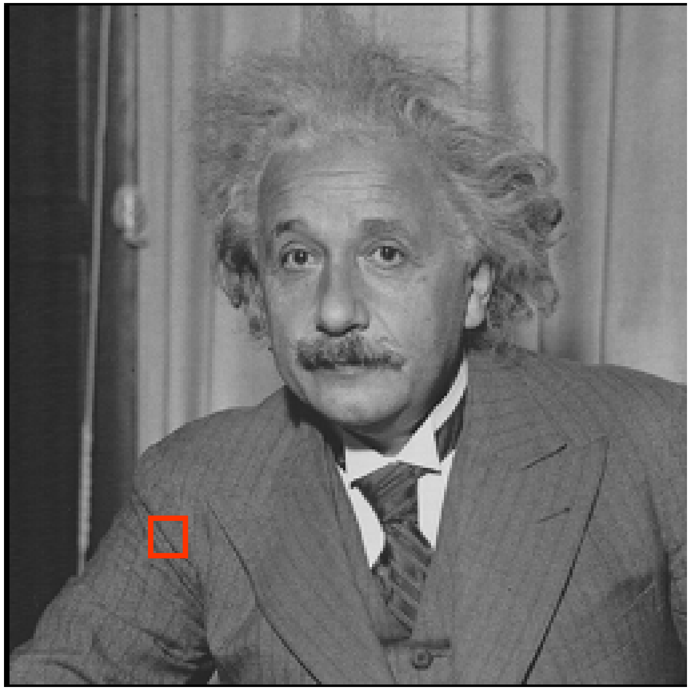
Block size
$N_1 = N_2 = 8$

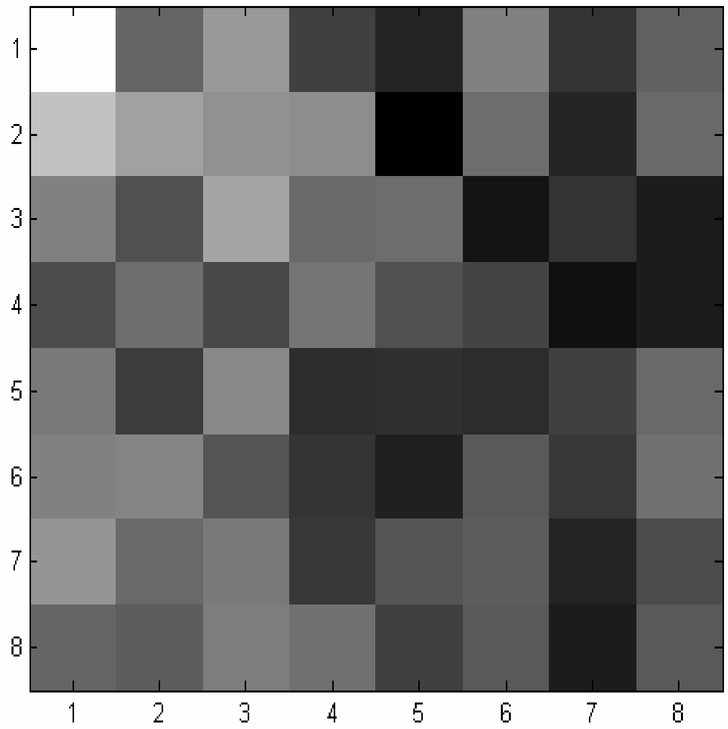


The source data (8x8) is transformed to a linear combination of these 64 frequency squares.

Block-based DCT

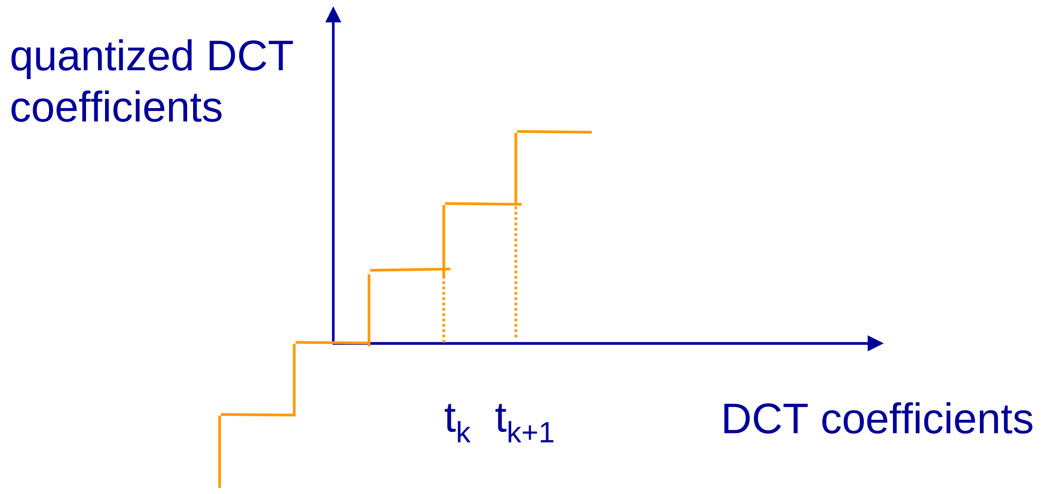
- The DCT coefficients are rounded to the nearest integer and 128 is subtracted. Then they are divided by the corresponding weight in the quantization matrix (see quantization)





Quantization

- JPEG quantizes the block cosine coefficients uniformly



Weighted quantization

- "Perceptual" quantization is also possible
  - Based on the observation that the human eye is more sensible to low than to high frequencies (luminance and color CSF)
  - The quantization tables are given but the standard allows the user to define his own quantization tables
  - quantization interval $\propto$ wights

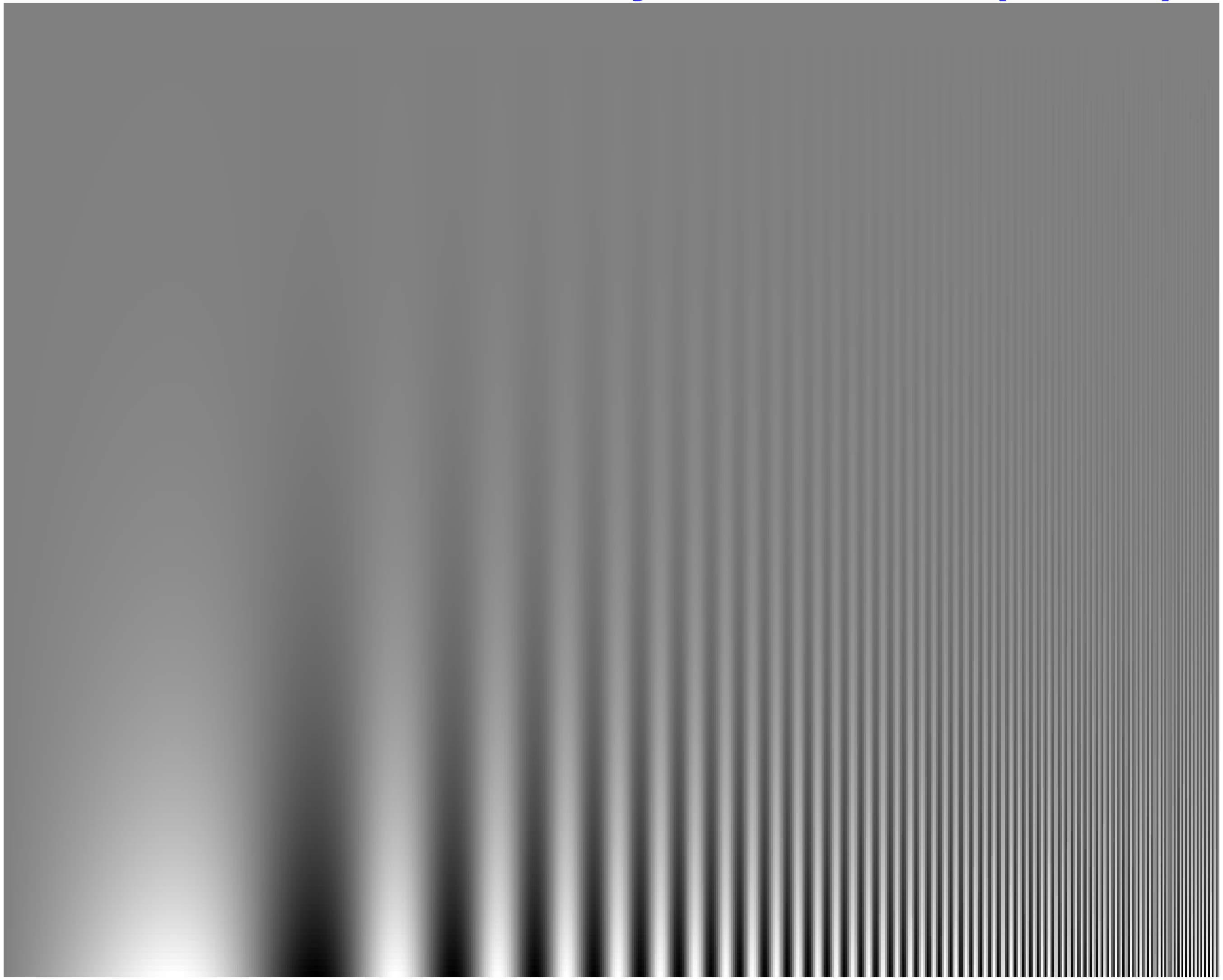
Luminance

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

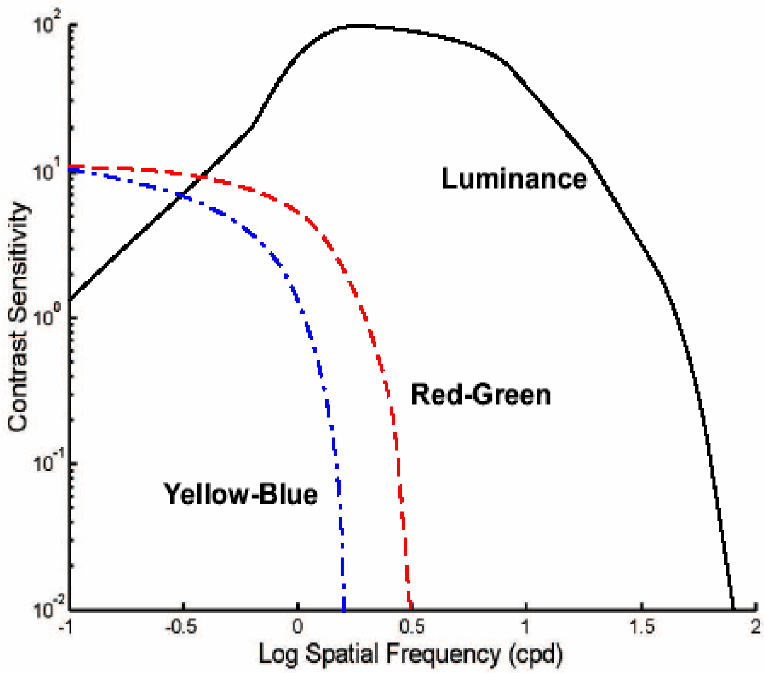
Chromaticity

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Contrast sensitivity function (CSF)



Luminance and Chromatic CSF



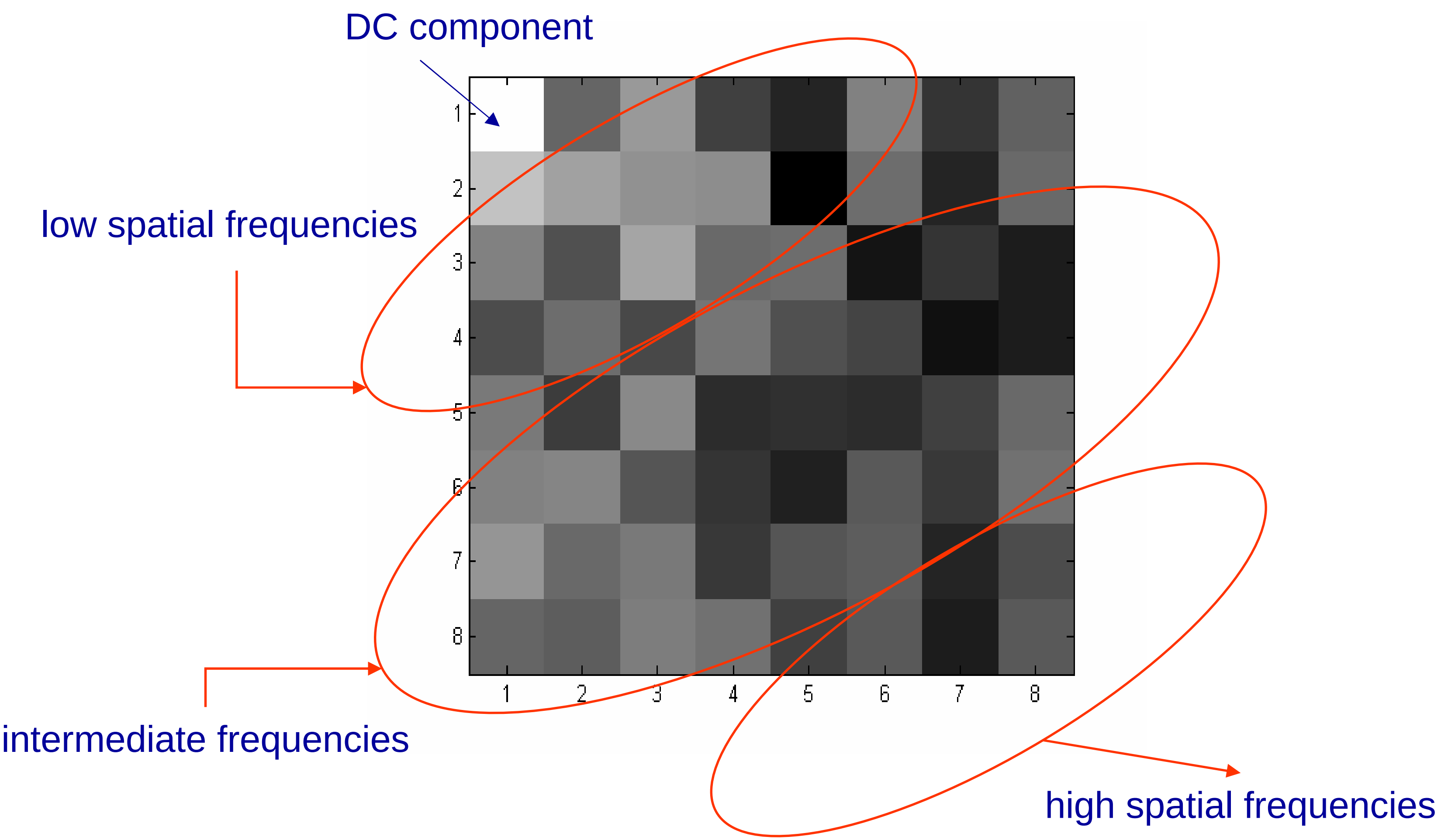
The spatial CSF for luminance contrast is band-pass in nature, with a **peak sensitivity around 5 cycles per degree**. This function approaches zero at zero cycles per degree, illustrating the tendency for the visual system to be insensitive to uniform fields. It also **approaches zero at about 60 cycles per degree**, the point at which details
can no longer be resolved by the human eye.

The chromatic mechanisms are of a low-pass nature and have ***significantly lower*** **cut-o frequencies**. This indicates the reduced availability of chromatic information for the details.

The yellow-blue CSF has a lower cutoff frequency then the red-green one.

Note also that the **luminance CSF is much higher** than the chromatic CSFs. **This denotes a greater sensitivity of the visual system to small changes in luminance contrast**
compared to chromatic contrast.

DCT block of coefficients

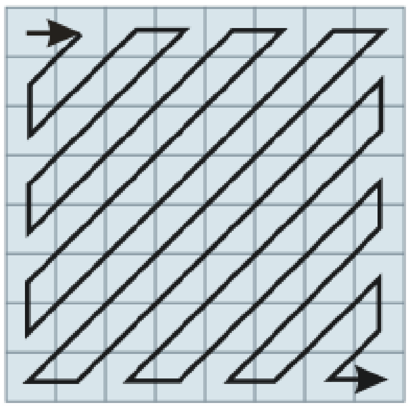


Raster scan of the coefficients

- In each block the coefficients are scanned in zig-zag order, which groups similar frequencies

quantized DCT coefficients

−26	−3	−6	2	2	−1	0	0
0	−2	−4	1	1	0	0	0
−3	1	5	−1	−1	0	0	0
−3	1	2	0	0	0	0	0
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

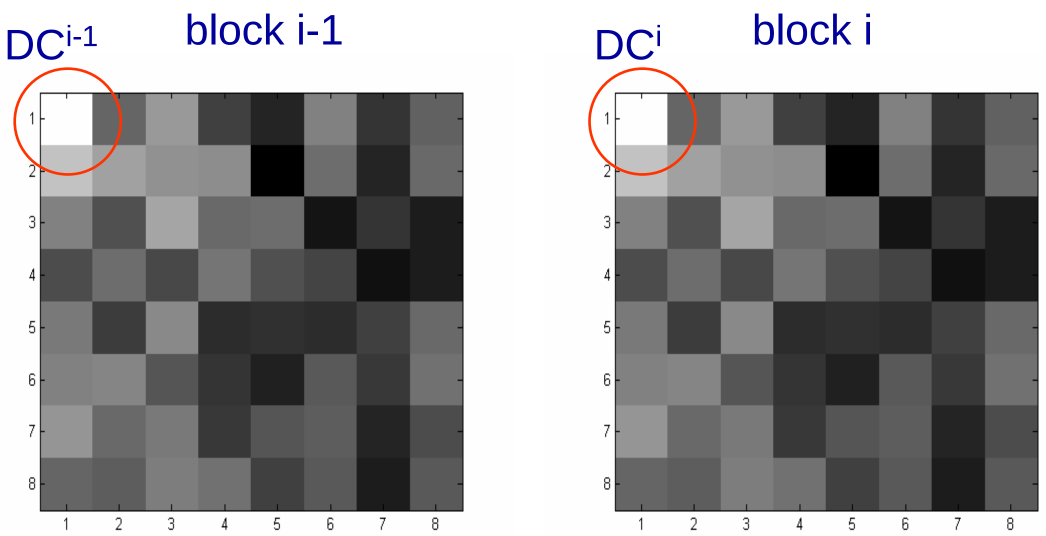


-26,
-3, 0,
-3, -2, -6,
2, -4, 1, -3,
1, 1, 5, 1, 2,
-1, 1, -1, 2, 0, 0,
0, 0, 0, 0, -1, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

DPCM on DC component

- The DC component changes slowly across the image
  - Since the bit depth is 8, it can be shown that the DCT coefficients are in the range $[-2^{10}, 2^{10}-1]$
- The residual with respect to the previous block is bounded

$$DC^{i} - DC^{i-1} \in \left[-2^{11}, 2^{11} - 1\right]$$

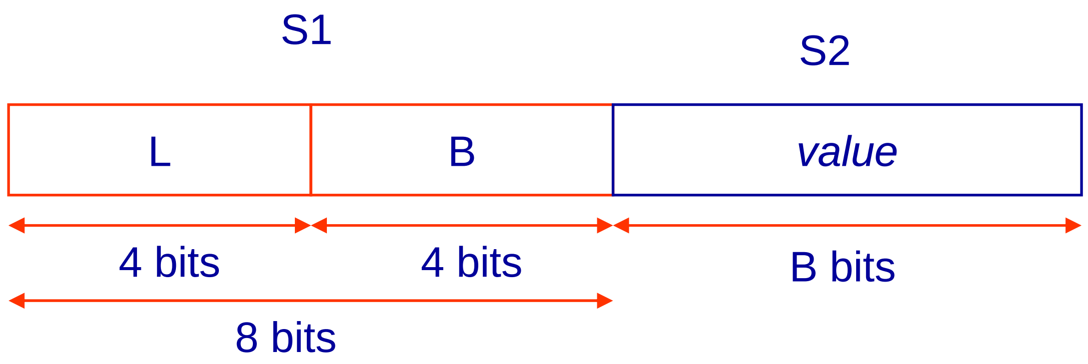


DPCM

- Predictive coding: The difference between the actual sample value and its predicted value (predicted value is based on previous sample or samples) is quantized and then encoded forming a digital value.
  - DPCM code words represent differences between samples unlike PCM where code words represented a sample value.
- Based on the fact that most source signals show significant correlation between successive samples so encoding uses redundancy in sample values which implies lower bit rate.
- Realization of basic concept
  - Predict current sample value based upon previous samples (or sample) and encode the difference between actual value of sample and predicted value (the difference between samples can be interpreted as prediction error).

JPEG codeblock

- Succession of symbols S1=(L,B) of eight bits followed by symbol S2
  - L: length of a consecutive run of zeros, coded on 4 bits
    - It takes values between 0 and 15
  - B: number of bits used to code the value of the next non-zero coefficient
  - The symbol S1=(15,0) is interpreted as a run-length of size 16 followed by another run length
    - Usually more than 15 zeros are present
  - If B is not zero, the subsequent symbol (of size B bits) specifies the amplitude of the next non zero coefficient



Example

0	0	0	4	0	0	0	1

$(3,3)(100)(3,1)(1)$

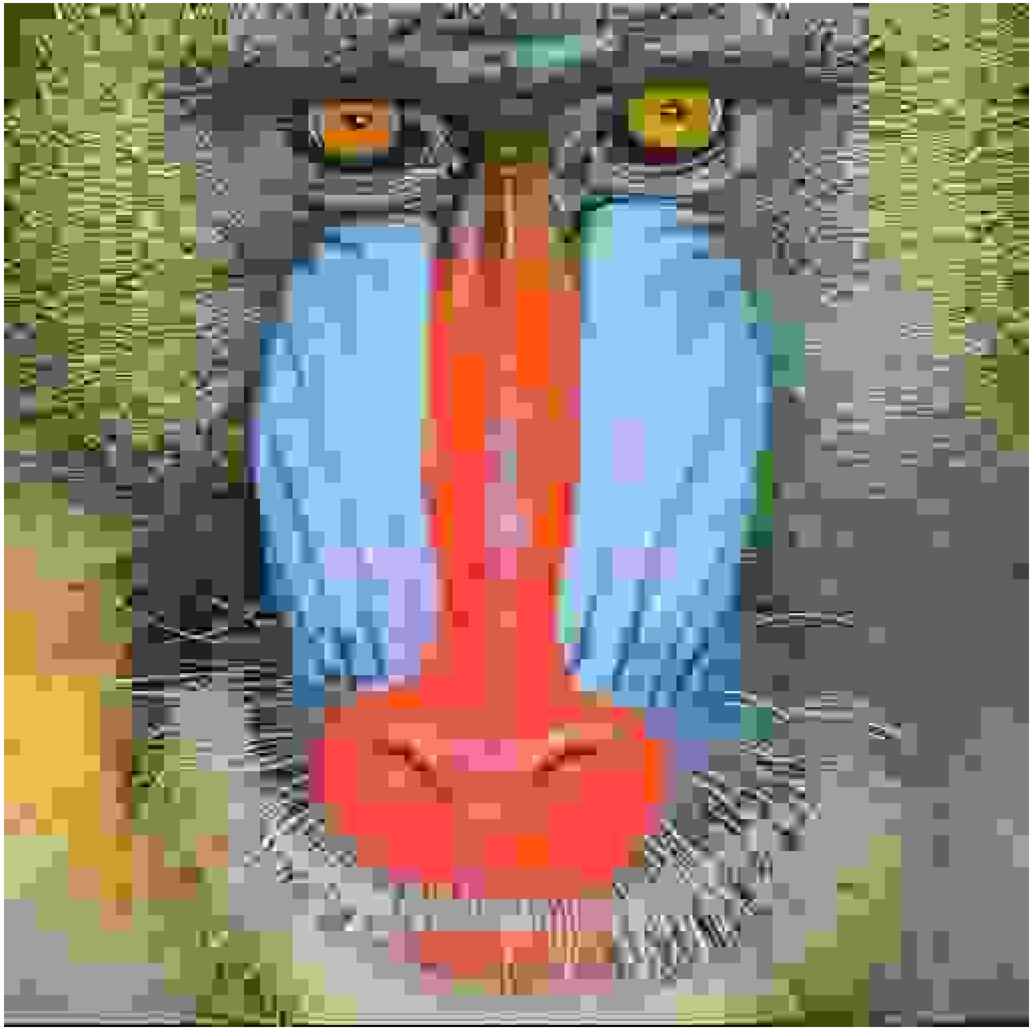
L B
S1 S2

The value of the coefficient coded on B bits belongs to a set of 2B values as indicated in the second column

B	Range of values
1	–1 1
2	–3 –2 2 3
3	–7 ... –4 4 ... 7
4	–15 ... –8 8 ... 15
....	
10	–1023 ... –512 512 ... 1023

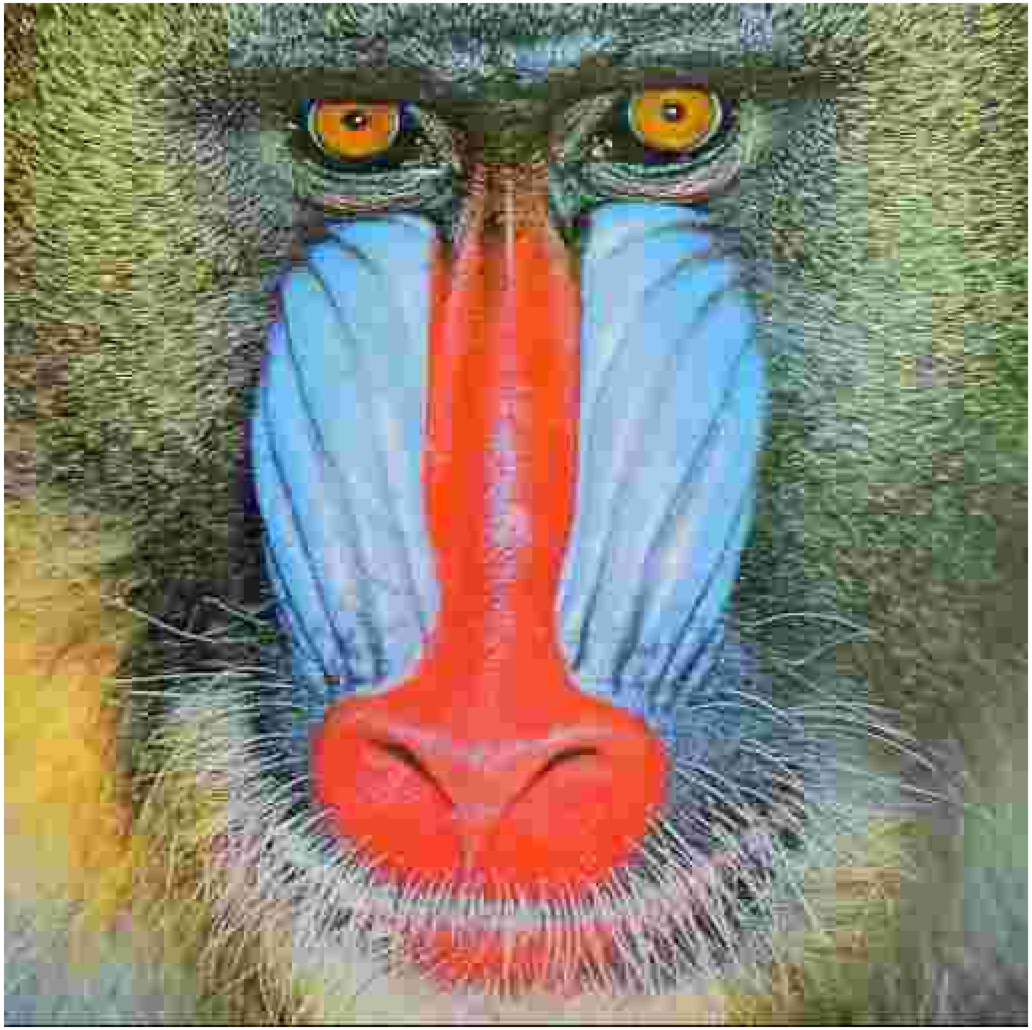
Example

0.2bpp



Example

0.5bpp



Example

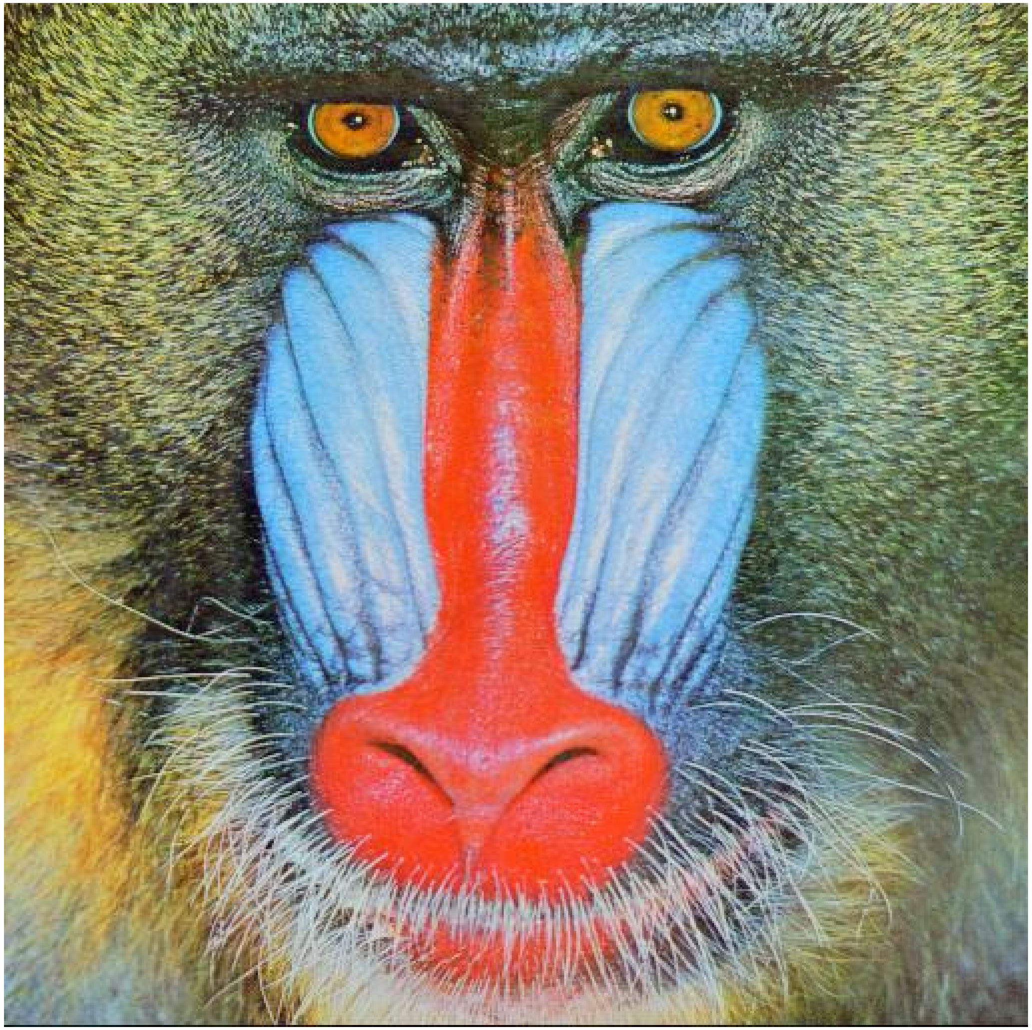
0.97bpp

Example

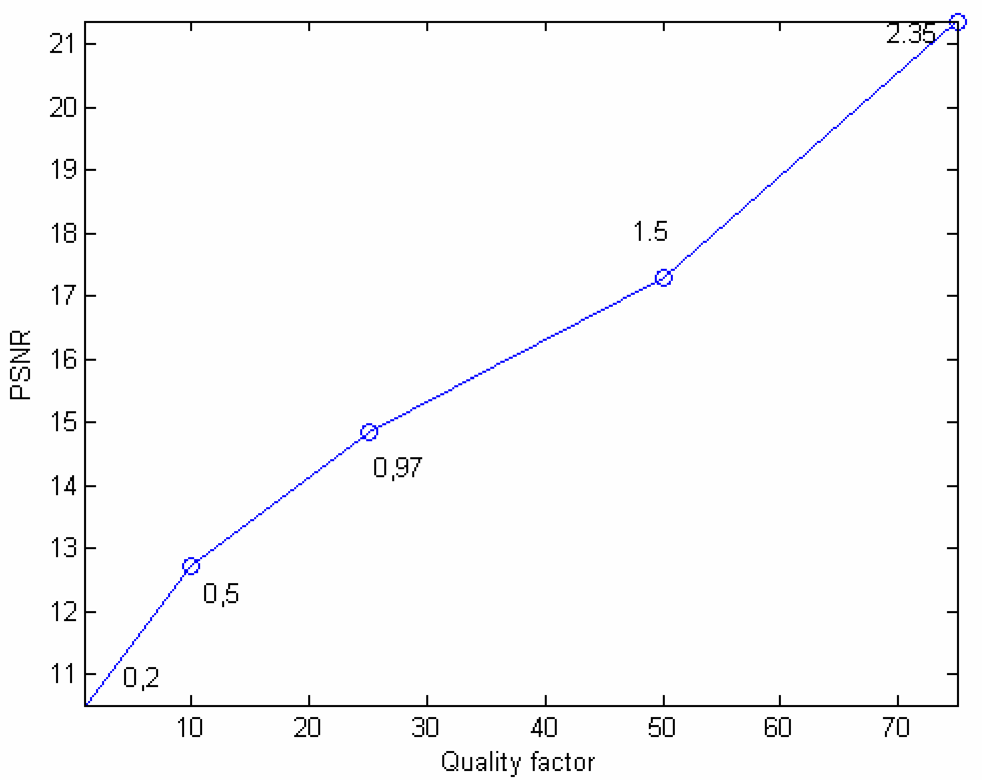
1.5bpp

Example

2.5bpp



Rate/Distortion plot



Pros&Cons

Pros

- Low complexity
- Memory efficient
- Reasonable coding efficiency
- Good image quality
- Adjustable image quality

Cons

- Single resolution
- Single quality
- No target bit-rate
- Blocking artifacts at low rates
- No region of interest (ROI) functionality
- No lossy and lossless representations on the same bitstream