File formats and Compression

File formats

- orderly sequences of data used to encode digital information for storage or exchange
- begin with an introductory 'header' section followed by a 'body'

Common file formats

- Tagged Image File Format (TIFF)
- Portable Network Graphic (PNG)
- Joint Photographic Experts Group File Interchange Format (JPEG or JFIF)
- Graphic Interchange Format (GIF)

Compression

- a way of encoding a file's data more concisely or efficiently
- Should reduce file size, but open easily
- Two strategies: Redundancy, Irrelevancy

Redundancy reduction

- used when the image is being encoded (or re-encoded)
- looks for patterns and repetitions that can be expressed more efficiently
- for example, there are 25 black pixels in a row, it is clearly better to record the information for one pixel and state that the next 24 are all the same, than to record each pixel separately

Irrelevancy reduction

- remove or alter information that makes little or no difference to the perception of the image

Lossless compressions

- generally based on redundancy reduction and typically concentrate on more efficient ways of encoding the image data.
- no information is irretrievably lost in the process

LZW compression

- Lempel, Ziv and Welch
- looks for recurrent patterns in the image's raster data, replacing these with codes, giving the most common patterns the shortest codes and storing all the definitions in a separate dictionary
- gif files

Lossy compressions

- based on irrelevancy reduction strategies but will usually also employ some redundancy strategies, particularly in their encoding
- gives much larger reductions in file size
- irreversible – it permanently disposes of information

Jpeg compression

- First step Discrete Cosine Transform to shift the image's colour values into a mode that can be more efficiently compressed and coded
- Next step, quantisation, simplifies and rounds the colour values before they are encoded, throwing away real information

Blockiness

- JPEG compresses small 8x8 pixel blocks of the picture at a time, working from top left to bottom right. Because the simplification (quantisation) of each 8x8 (64 pixel) block is done independently, at a high compression (i.e. low quality) the boundaries between the blocks will begin to show, causing the 'blockiness'

Newer approaches

- Fractal compression
- Wavelet compression

Fractal compression

- fractal compression approaches the image as a picture rather than a series of numbers to be processed.
- fractal compression looks for recurrent shapes and patterns in the image and replaces these with equations
- more computing power required
- STN format

Wavelet compression

- the wavelet transform regards the image as a signal or wave. It organises the image information into a continuous wave and decomposes it by repeated averaging
- jp2

Imaging Geometry