



1.3.1 Compression

Lossy and Lossless

Compression is the process of reducing the size of a file.

This can be done to:

Reduce the storage requirements for the file

Reduce the bandwidth and time taken for a file to be transmitted over a network.

Compression methods can be either:

Lossless – The file size is reduced with no loss of data. The original data can be retrieved when the file is decompressed.

Lossy – The file size is reduced with some loss of unnecessary data. This data is permanently lost.

Lossless compression is suitable for compressing text files and simple bitmap images.

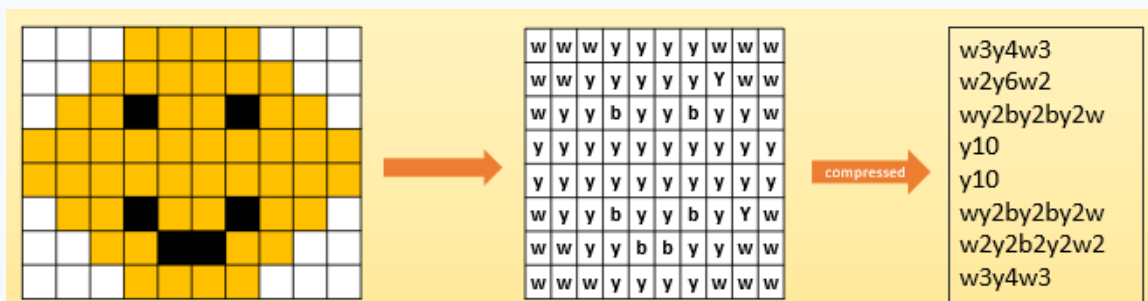
Lossy compression is suitable for compressing photos and videos. A loss in the quality of the image or video is acceptable in order to reduce the time taken to transmit over a network.

Lossless: RLE

Run length encoding (RLE) is a lossless compression technique that uses sequences of repeating data to reduce the size of a file.

The data and the number of times it appears in a sequence is recorded in a format that is smaller than the original bitmap image.

This bitmap image has many repeating sequences of pixels of the same colour.



When required the compressed file can be decompressed to reconstruct the original bitmap image again.

If the image did not have many repeating sequences of pixels of the same colour, the compression would be ineffective.





Lossless: Dictionary Encoding

Dictionary coding is a lossless compression method that substitutes the items in a file with an index to an entry for that item in a dictionary.

Example: Dictionary coding

Nory was a Catholic because her mother was a Catholic, and her mother was a Catholic because her father was a Catholic, and her father was a Catholic because his mother was a Catholic, or had been

compressed

```
1 2 3 4 5 6
7 2 3 4 8 9 6 7
2 3 4 5 6 10 2 3
4 8 9 6 10 2 3 4
5 11 7 2 3 4 5 12 13 14
```

1	Nory
2	was
3	a
4	Catholic
5	because
6	her
7	mother
8	,
9	and
10	father
11	his
12	or
13	had
14	been

Each repeated word is stored only once in the dictionary. Occurrences of the word are replaced by a shorter index, reducing the amount of storage required.

Dictionary encoding is particularly effective when a file contains many repeated words or sequences.

