Cambridge IGCSE

Computer Science

Section 1

Compression

Unit 1: Data representation

Objectives

- Understand the purpose and need for data compression
- Understand how files are compressed using lossy and lossless compression methods

Data compression

We use data compression to **reduce the size of files**. *This is especially important for sound and image files which can be very large.*

Reducing the size of files means :

-> they take **less storage space** on our digital devices

-> **less bandwidth** will be required to download and upload the data

-> **faster transmission** of data, so reducing the time to send and receive music, videos etc.

Data compression

Mostly used with large sound, image and video files.

There are two types:

- **Lossy** compression
- **Lossless** compression

Lossy compression

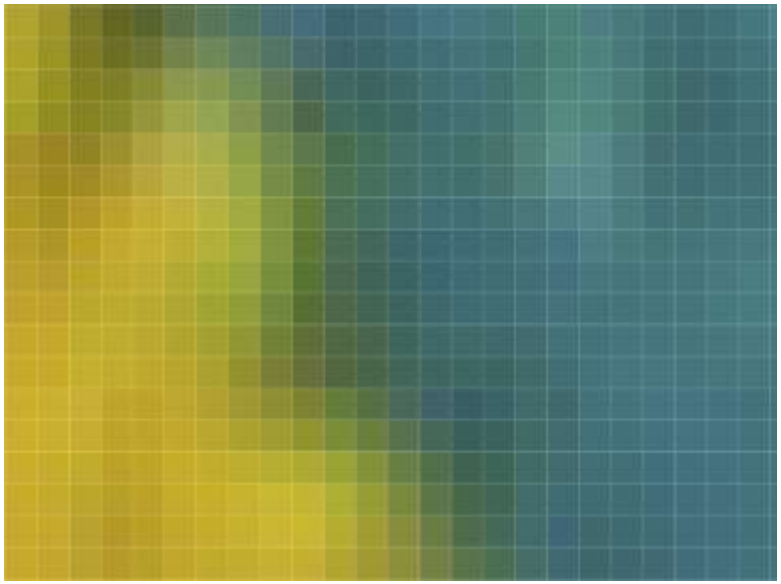
- When a file is compressed, the **unnecessary bits** of information are **removed permanently**
- The original file cannot be reproduced once it has been compressed -> there is **some loss of data**
- So, the file after compression is **NOT** exactly the same as the original
- The compression algorithm tries to remove the data that is **not likely to be noticed** by humans.

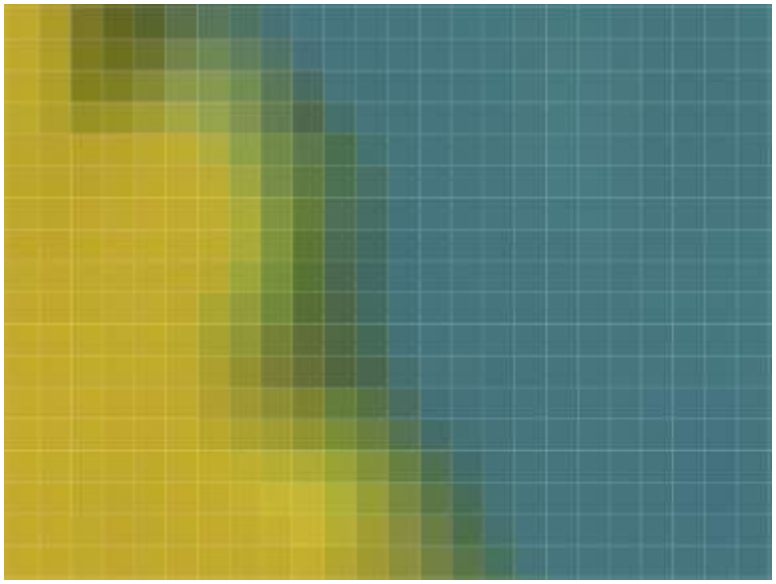
Lossy compression - images

- Removes data permanently
- Tries to reconstruct an image without the missing data
- Produces much smaller file sizes than lossless but also, some loss of quality (eg .jpg)

Lossy compression - images

- With images, similar coloured pixels are all made the same colour by the compression algorithm.





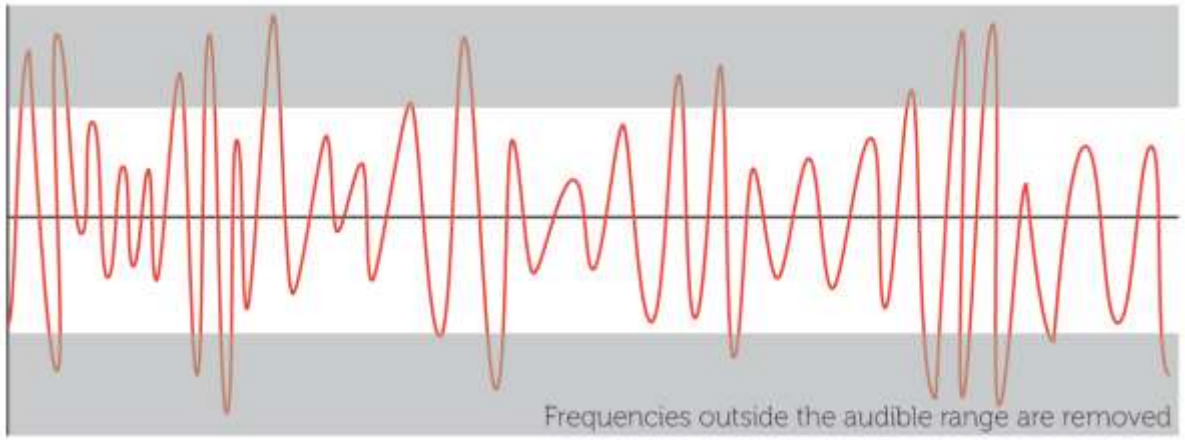
- The result after compression is NOT exactly the same as the original

Our hearing range

- Between 20-20,000 Hz
- Varies depending on age
  - Take the test!

Lossy compression - sound

- Compression algorithm removes the sounds in the frequency ranges that we cannot easily hear or that least affect the playback quality
- Lossy compression leaves out some data - this can affect the quality of the sound recording



Lossy compression - text

- Lossy compression **does not work** with **text** - a computer program, or a page of writing.

```
# S CTIO  2
w ile gues    =    rr ctPas  or :
    guess = i put "Try to guess t      sswor    ")
    g   ses   gue  es + 1

rint("Pa swor   uessed c r  ctly
```

Lossy compression - summary

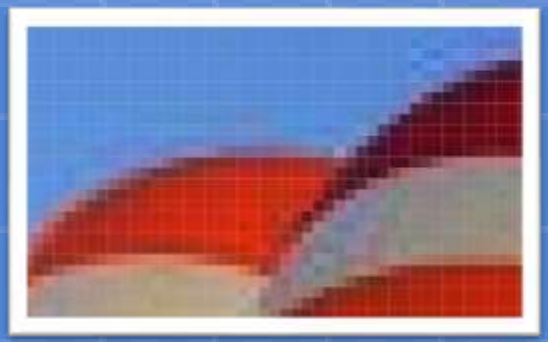
- Compression algorithm reduces file size by permanently removing some data
- In Image files, the resolution (number of pixels) and colour depth (number of colours) are reduced
- In Sound files, the sample rate and resolution are reduced *(also take away unnecessary sounds … frequency test)*
- The file is recreated using the remaining data and algorithms guess the removed information
- The compressed file is **not the same** as the original
- Not be used with text files, because all data must be kept.

Lossless compression

- When the file is compressed, the **quality of the data remains the same**.
- When the file is uncompressed, it is restored exactly as it was in its original form, **none of the detail from the original file is lost**
- This is used when all of the information is important and needs to be kept the same.

Lossless compression - images (eg .png)

- Finds areas of the same colour and records them as 15 blue pixels rather than blue pixel, blue pixel, blue pixel etc.
- 11011010,11011010,11011010,11011010 becomes 00000100-11011010
- This is called Run Length Encoding - RLE - looking for repeated patterns



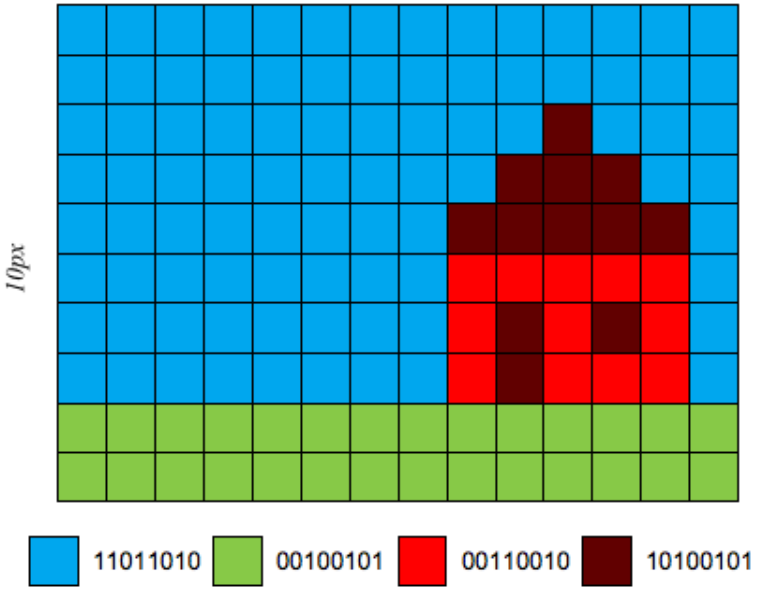
Lossless compression

- Lossless compression leaves out repeated data and instead makes a note of how many times it is repeated
  - *E.g.* ***10 x 5*** *takes less space to store than:*

    5+5+5+5+5+5+5+5+5+5

- This is an example of **RLE - Run Length Encoding**

Lossless image compression



The first and second rows of the image above would be stored as 14 lots of 11011010:

11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010, 11011010

This could be stored as *00001110-11011010* (14 of 11011010) without losing any of the information required to put the image together again. This is known as RLE or Run Length Encoding. Record the data for the third, fourth and fifth lines in the image:

Line no.	Binary image data
1	00001110-11011010
2	00001110-11011010
3	

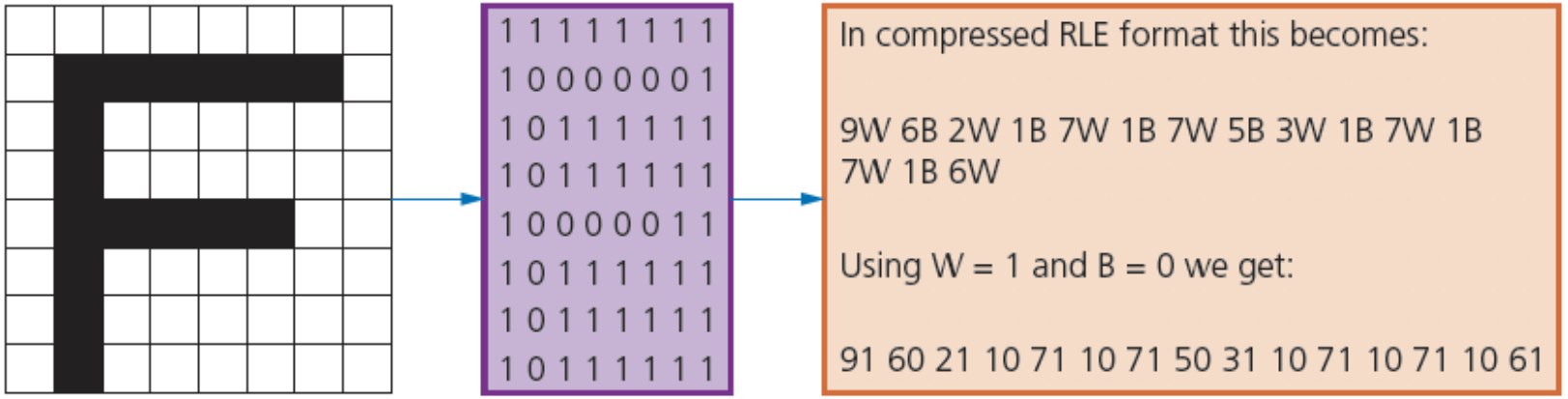
Sound file formats

- **.WAV** – uncompressed files
- **.FLAC** or **.M4A** lossless compression, slightly smaller files
- **.MP3** – Lossy compression, much smaller files

Lossless image compression

Example of Run Length Encoding (RLE)

Each square in the grid requires 1 byte of storage. A white square has a value 1, and a black square a value of 0.



Uncompressed, the 8x8 grid would need **64 bytes**.

The **compressed RLE format** has 30 values, and therefore only needs **30 bytes** to store the image.

Lossless text compression

- Finds patterns in the original text
- Encodes each pattern in a dictionary

An eye for an eye,

a tooth for a tooth

.	0	0000
An_	1	0001
eye	2	0010
for	3	0011
an_	4	0100
,_	5	0101
a_	6	0110
tooth	7	0111

- 38 Characters including spaces = **38 bytes** (assuming an 8-bit ASCII table is used)

Dictionary compression

An eye for an eye,

a tooth for a tooth.

.	0	0000
An_	1	0001
eye	2	0010
for	3	0011
an_	4	0100
,_	5	0101
a_	6	0110
tooth	7	0111

1	2	3	4	2	5	6	7	3	6	7	0
0001	0010	0011	0100	0010	0101	0110	0111	0011	0110	0111	0000

When compressed ...

48 bits = **12 bytes** = 32% of original size (including codes)

Transmission of data

- Why use compression?
  - Download speeds are increased
  - Data allowances are saved
  - Voice can be transmitted fast enough to keep up with speech

Advantages of compression

- Smaller files = fewer packets = faster transmission
  - Quicker to complete
  - Reduces traffic over the network Internet
  - Less chance of collisions or transmission errors
- Improves download speed of video, sound (including speech used for VOIP systems) and image files
- Speeds up download of webpages that use images
- Takes less space on disk / servers
- Enables better streaming of music and video

Buffering

- Video or music streaming causes buffering if the download speed is slower than the playback speed

  - How could you reduce the chances of people experiencing buffering if you were the website owner?

Uncompressed file types

- TIFF (.tif) and BMP (.bmp) refers to raw bitmap, that is, uncompressed image files.

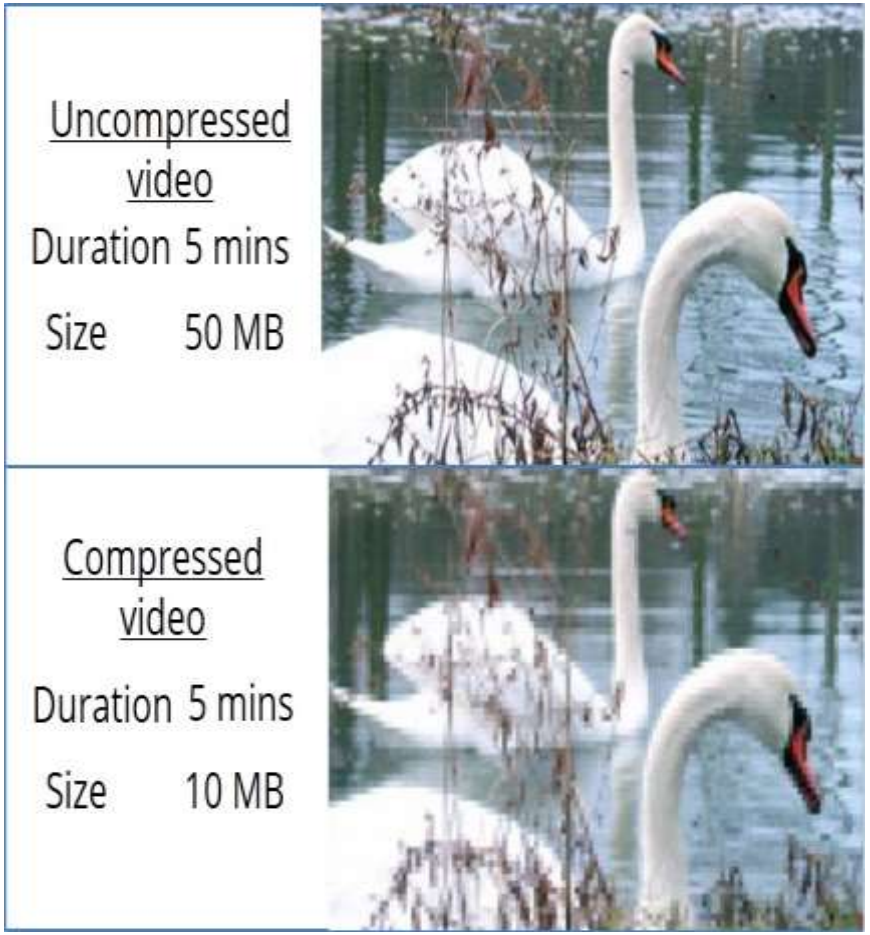
- These file formats are uncompressed, they represent images with highest image quality.

Compressed file types

Format	Type of compression	Application
PNG	lossless	Used for transferring images over the internet
JPEG	lossy	Higher compression rate than a PNG. Used in digital cameras and web pages.
GIF	lossless	Compresses images to a maximum of 8-bit depth. Not used for high quality images. A sequence of gifs are used to store animated graphics. Used for small images such as logos, icons, etc.
PDF	lossless	Encodes text and graphics

Compressing video files

- Compressing a video file reduces the resolution, dimensions and bit rate.
- Compressing a video file may also lead to poor quality and random coloured blocks on the screen. These blocks are called artefacts.
- MP4 and MOV are examples of lossy video file formats.



Streaming audio and video files

- Compression is very helpful in streaming and downloading audio and video files.
- MP3 file format is used for audio compression.
- MP3 ensures up to 90% of compression.
- MP3 files are used for storing files in computers, MP3 players, mobile phones, etc. The CD files are converted to MP3 file format using file compression software.
- Even though, the quality of MP3 file cannot match the original CD file, it is still satisfactory for various purposes

Summary

Compression exists to reduce the size of the file

-> less bandwidth required

-> shorter transmission time

-> less storage space required

- Lossless compression reduces the file size without permanent loss of data, e.g. run length encoding (RLE)

- Lossy compression reduces the file size by permanently removing data, e.g. reducing resolution, colour depth, sample rate …

Key Vocabulary

- compressed
- uncompressed
- bandwidth
- algorithm
- streaming
- storage
- lossy
- lossless
- transmission
- downloading
- buffering

Past paper question examples ...

(c) Carla wants to reduce the file size of the photos she has transferred to her computer. She does not want the quality of the photos to be reduced, so she uses lossless compression.

Describe how lossless compression reduces the file size of the photos.

.....

.....

.....

.....

.....

.....

.....

..... [4]

Past paper question examples ...

(d) Priya shares her sound files with other students. Before sharing the sound files, she compresses the files using lossless compression.

Describe how lossless compression reduces the size of a sound file.

.....

.....

.....

..... [2]

Past paper question examples ...

5 Audrey wants to send a sound file to Nico using email.

The file is too large to attach to an email so Audrey decides to compress the file.

She uses lossy compression to reduce the size of the sound file.

(a) Describe how lossy compression reduces the size of the sound file.

.....

.....

.....

.....

.....

.....

.....

..... [4]

Past paper question examples ...

(b) Nico asks Audrey why she used lossy compression rather than lossless.

(i) State **one** advantage Audrey could give of using lossy rather than lossless to compress the sound file.

.....

..... [1]

(ii) State **one** disadvantage Nico could give of using lossy rather than lossless to compress the sound file.

.....

..... [1]