A-level
Computer Science

Compression

Lesson Objectives

Students will learn about:

- Why compressing files is important.
- How text, image, audio, and video files are compressed.
- Effects of compressing a file.
- What the various file formats are.
- Compression algorithms: Run-length encoding and Huffman coding.

1.

Content

Introduction

- File handling is one of the primary functions of a computer system.
- Based on the type of data that needs to be stored, several types of file formats are available.
- Each file format occupies a certain amount of storage space.
- An image file with good quality occupies around 1 MB, and a video file needs to store 25 frames per second, occupying a large amount of storage space. Thus, compression methods are used to reduce the size of the files.
- Compression is also helpful in reducing the download time of image, audio, and video files from the Internet.

Compressing Image files

- Image compression is the reduction in file size to reduce download times and storage requirements.
- Compressing an image also changes its attributes, such as file type, resolution, dimensions and bit depth.
- There are two types of compression:
  - ✓ Lossless compression
  - ✓ Lossy compression

Lossless compression

- When the file is compressed, the quality of the image remains the same.
- The image can be reconstructed into its original form.
- In this case, information is very important and cannot be lost.

Lossless compression

- Let us consider a text file with the following sentence:

“See a pin and pick it up, all the day you'll have good luck; see a pin and let it lie, bad luck you'll have all day”.

- This text file can be compressed by making a table for this information.
- A character occupies a byte of memory. Whereas, numbers occupy comparatively less memory. The number 16 is represented in 5 bits (10000).

Index	word
1	see
2	a
3	pin
4	and
5	pick
6	it
7	up
8	all
9	the

Index	word
10	day
11	you’ll
12	have
13	good
14	luck
15	let
16	lie
17	bad

“See a pin and pick it up, all the day you'll have good luck; see a pin and let it lie, bad luck you'll have all day”

- The sentence can be coded in the form of numbers in the table and stored in the computer:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 1 2 3 4 15 6 16 17 14 11 12 8 9 10

- This saves memory by using codes for words that are repeated.
- With the code and the index table, the complete sentence can be recreated.

Index	word
1	see
2	a
3	pin
4	and
5	pick
6	it
7	up
8	all
9	the

Index	word
10	day
11	you’ll
12	have
13	good
14	luck
15	let
16	lie
17	bad

Lossy compression

- When a file is compressed, the unnecessary bits of information are removed permanently.
- This information is less likely to be noticed by humans.
- This type of compression is used for photographs where the information to be compressed cannot be predicted.

Uncompressed file formats

- TIFF (.tif) and BMP (.bmp) refers to raw bitmap, that is, uncompressed image files.
- As these file formats are uncompressed, they represent images with the highest image quality.

Types of compressed files

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 gif images is used to store animated graphics. Used for small images such as logos, icons, etc.
PDF	Lossless	Encodes text and graphics.

Streaming audio files: MP3

- Lossy file compression technique.
- A user does not hear any difference while listening to the MP3 file because of perpetual music shaping in file compression algorithms that only removes details that humans cannot hear.
- Therefore, only a few parts of the sound file are removed and sound quality is not compromised too much.
- The bit rate affects the quality of the MP3 file. It ranges from 80 kbps to 320 kbps. An original CD-quality audio has a bit rate of 1411 kbps.

Codecs and Compression Algorithms

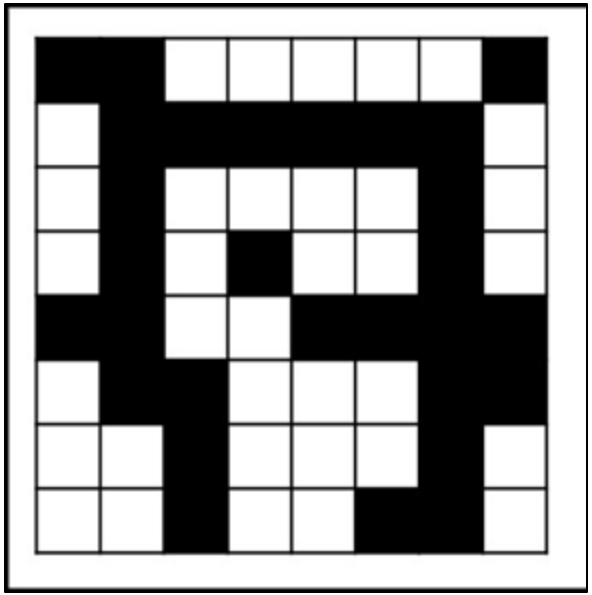
- Codecs are programs that encode or decode an audio, image, or video file.
- Compression codecs are aimed at reducing the size of a file without affecting its quality.
- Algorithms decide the amount of data that can be removed to reduce the file size.

Run-length encoding (RLE)

- Run-length encoding (RLE) is an example of a compression algorithm that converts the consecutive similar values into code.
- This code consists of the identical value and the number of times this value is repeated.
- This is a lossless type of compression.

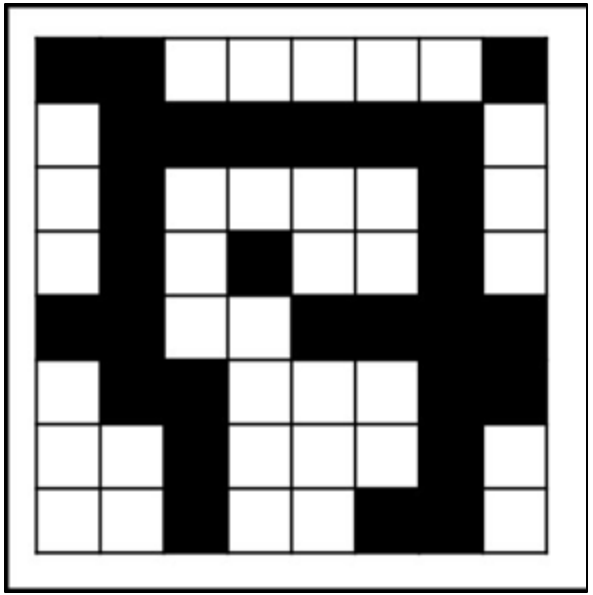
Run-length encoding (RLE)

- The computer stores binary value 1 for white and binary value 0 for black for each row of the image.
- The first row in the image can be represented as 2 0 5 1 1 0. This code represents 2 black pixels, 5 white pixels and 1 black pixels.
- Similarly, the second row in the image is represented as 1 1 6 0 1 1.

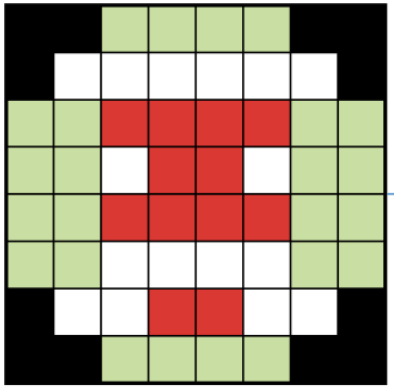


Run-length encoding (RLE)

- This type of coding is not efficient if the file does not have many runs.
- In such cases, the file size may increase instead of getting compressed.
- RLE is therefore used only in simple images with large areas of the same colour.



Run-length encoding (RLE)



Square colour	Red	Green Components	Blue
	0	0	0
	255	255	255
	0	255	0
	255	0	0

2 0 0 0 4 0 255 0 3 0 0 0 6 255 255 255 1 0 0 0 2 0 255 0 4 255 0 0 4 0 255 0 1
255 255 255 2 255 0 0 1 255 255 255 4 0 255 0 4 255 0 0 4 0 255 0 4 255 255 255
2 0 255 0 1 0 0 0 2 255 255 255 2 255 0 0 2 255 255 255 3 0 0 0 4 0 255 0 2 0 0 0

Run-length encoding (RLE)

2 0 0 0 4 0 255 0 3 0 0 0 6 255 255 255 1 0 0 0 2 0 255 0 4 255 0 0 4 0 255 0 1
255 255 255 2 255 0 0 1 255 255 255 4 0 255 0 4 255 0 0 4 0 255 0 4 255 255 255
2 0 255 0 1 0 0 0 2 255 255 255 2 255 0 0 2 255 255 255 3 0 0 0 4 0 255 0 2 0 0 0

- The size of original image file?
- How many bytes are used for the compressed file?
- Compare the size of original and compressed image file.

2.

Activity

3.

End of topic questions

End of topic questions

1. Why is file compression important in computer systems?
2. What is the difference between lossy and lossless compression?
3. What are the advantages and disadvantages of using an MP3 file format to compress audio and video files?
4. What is run-length encoding?
5. What type of compression algorithm (lossy/lossless) will you use for sending a computer program as an email attachment? Why?

7. Encode the image given, using run-length encoding.

00: black

10: blue

01: gray

11: white