Lecture 16: Multimedia

AP Computer Science Principles

Images

File Formats

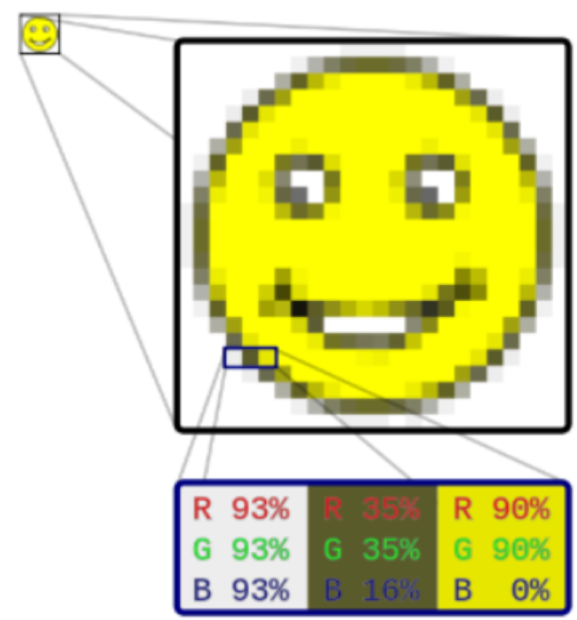
- A **multimedia** file(audio, video, image) is simply a sequence of 0's and 1s.
- In Lecture 2, with character encoding, this sequence can be translated into text.
  - For example, ASCII, UTF-8. In ASCII, 65 is an A, etc.
- Encoding such as ASCII, however, doesn't specify, for example, font color.
  - HTML provides one way of specifying font, background color.
- But a Word document(file ends in .doc and .docx) doesn't use HTML.
  - .doc is called a **filename extension**, indicates file's format.
- The purpose of a file format is to define a standardized way of representing information.

File Formats

- A file's format simply tells your computer how the bits of a files should be interpreted.
  - Those 0s and 1s could represent an image, a sound, or a video, but without any kind of rules for interpreting a file's bits, they're essentially meaningless.
- A file's extension usually indicates the format of the file, but isn't a definitive answer.
  - Microsoft word can open both important.doc to important.whatever
  - a PPM file for example is a text file if open with Notepad or TextEdit but is an image if open with GIMP.
- Popular image file formats: BMP, JPEG, PNG, GIF.
  - All of these acronyms are just different ways of representing images using binary data.

Bitmap

- An image is 2D grid of squares(**bitmap**), where each square is filled with only one color; this is kind of structure is called a **raster graphics image.**
  - similar to a 2D array.
- Each of these small squares, called a **pixel**, can be filled with exactly one color.
- Computers use **additive color mixing** to produce colors
  - primary colors are red, green and blue(RGB)
- Typically, bitmaps use 24-bit(3 bytes) for colors.
  - 8 bits for each of the three primary colors.
  - each color has values 0 to 255.
  - 24-bit equals aproximately 16.8 million colors.



Bitmap

- Colors often are represented in hexadecimal.
- Each color is stored using 8 bits, need six hexadecimal digits to express a 24-bit color
  - a hexadecimal is 4 bits.
- as a convention, hexadecimal colors are prefixed with either the hash character (#) or sometimes (0x)
  - the color #FF0000 (red) is a lot of red, no green, and no blue, while the color #0088FF (sky blue) is no red, a bit of green, and a lot of blue
- .bmp is a bitmap file format.
  - The first two bytes of any bitmap file are the same: the magic number 0x42 0x4D. 0x42 is the same as the decimal number 66, while 0x4D is the same as the decimal number 77.
  - then some **metadata**(data that describes a file's data) such as width, height, size.
  - then the bytes that describe the pixels of the image.

Resolution

- raster graphics image has to be made up of a grid of squares, but how big should each square be?
  - the **resolution** of an image, or the amount of information stored in the image
- same image size but smaller pixel size equals higher resolution(more detail) and therefore larger number of pixels in the image.
- for example, the left image below is 16 pixels by 16 pixels resolution

Lossless Compression

- what would the bitmap data look like for the image below?
  - a lot of the bytes are exactly the same!
- Is there a way to encode the image with an abbreviation for “500 black pixels” rather than listing each pixel individually?
- CompuServe develops GIF, or the Graphics Interchange Format, in 1987.
  - can be pronounced “GIF” or “JIF”
- Unlike bitmaps, GIFs are compressed,
  - can represent exactly the same information as some bitmaps using a smaller number of bits.
  - essentially, if two pixels that are horizontally adjacent are exactly the same, then GIF compresses the data.

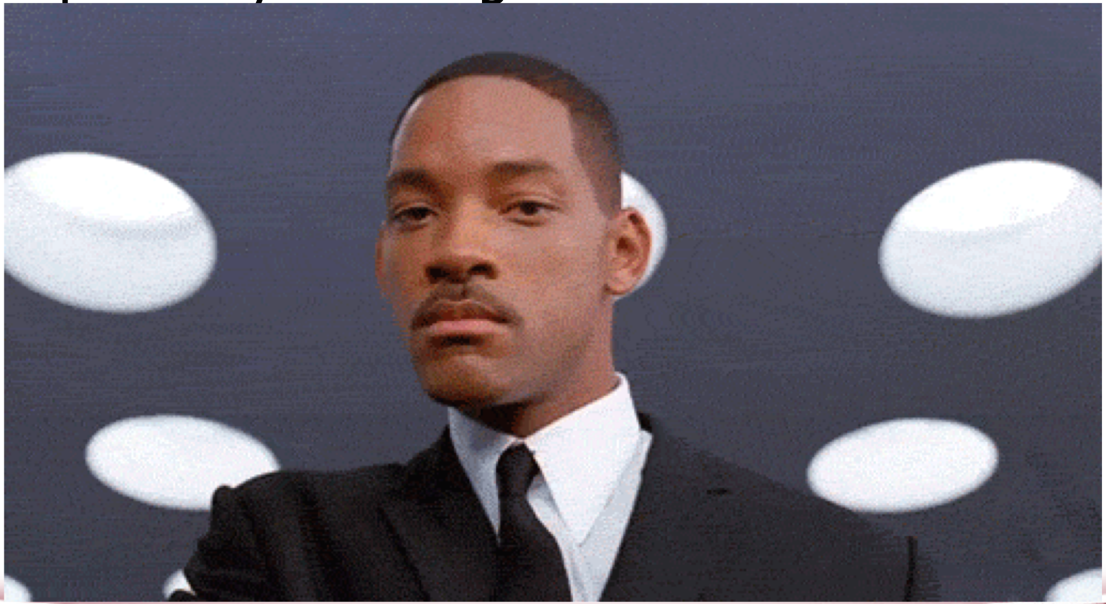
- this method is called **lossless compression**, i.e, no data is lost.
- besides images, RAR, ZIP, GZIP, LZW compress any file.

Lossy Compression

- while lossless compression preserves all of a file's data, **lossy compression** throws away some data in the interest of compressing the file even more.
- effective lossy compression should preserves the essential data
- you use effective lossy compression every time you text someone!
  - Hey wat r u doin tmrw? I wntd 2 go 2 c Ben.(43 characters)
  - Hey, what are you doing tomorrow? I wanted to go to see Ben.(57 characters)
  - compressed by about 25%
  - you can “probs” do even better.
- GIF uses lossless compression
- JPEG(Joint Photographic Experts Group), for example, uses lossy compression.
- PNG(Portable Network Graphic) is another popular format.
  - the original creators of PNG actually picked the acronym for "PNG is Not GIF" as a reaction to some of the dubious licensing issues with the GIF format.

Comparison

- Like GIF, PNG uses lossless compression, and like JPEG, uses 24 bits for color.
- Another feature of the PNG format is support for alpha, or transparency. BMP and JPEG, on the other hand, don't have a channel devoted to **transparency**.
- GIF does have the distinct advantage of supporting animation achieved by repeatedly showing a series of frames.



Comparison

- Summary of these formats.

Name	Extension	Compression	Color	Alpha
Bitmap	.bmp	No	24-bit	No
GIF	.gif	Lossless	8-bit	Yes
JPEG	.jpg, .jpeg	Lossy	24-bit	No
PNG	.png	Lossless	24-bit	Yes

Vector Graphics

- If a raster graphics image is scaled larger, the image will appear pixelated.
- **Vector graphics** can scale without losing quality.
- Unlike raster graphics, vector graphics doesn't involve turning an image into a grid and storing the values of individual pixels.
- Instead, vector graphics stores images using mathematics.
  - For example, let's say we want to represent a circle. In raster graphics, we'd create a grid of squares and then trace out a circle by filling in squares with some color.
  - Using vector graphics, we'd instead say that the equation for a square looks something like x^2 + y^2 = r^2.
  - We can create a raster graphic simply by picking a size for the image, then using the equation to figure out which pixels should be colored.
  - This representation doesn't depend on any pixels. If we want to create a larger circle, we can just pick a larger value for r, and our equation will create a larger circle without any loss in quality.

Scalable Vector Graphics

- **Scalable Vector Graphics(SVG)** is an example of a vector graphic.
- Notice the difference below between a JPEG and a SVG after some rescaling.

$$\sqrt{b^2 - 4ac}$$

$$\sqrt{b^2 - 4ac}$$

Sound

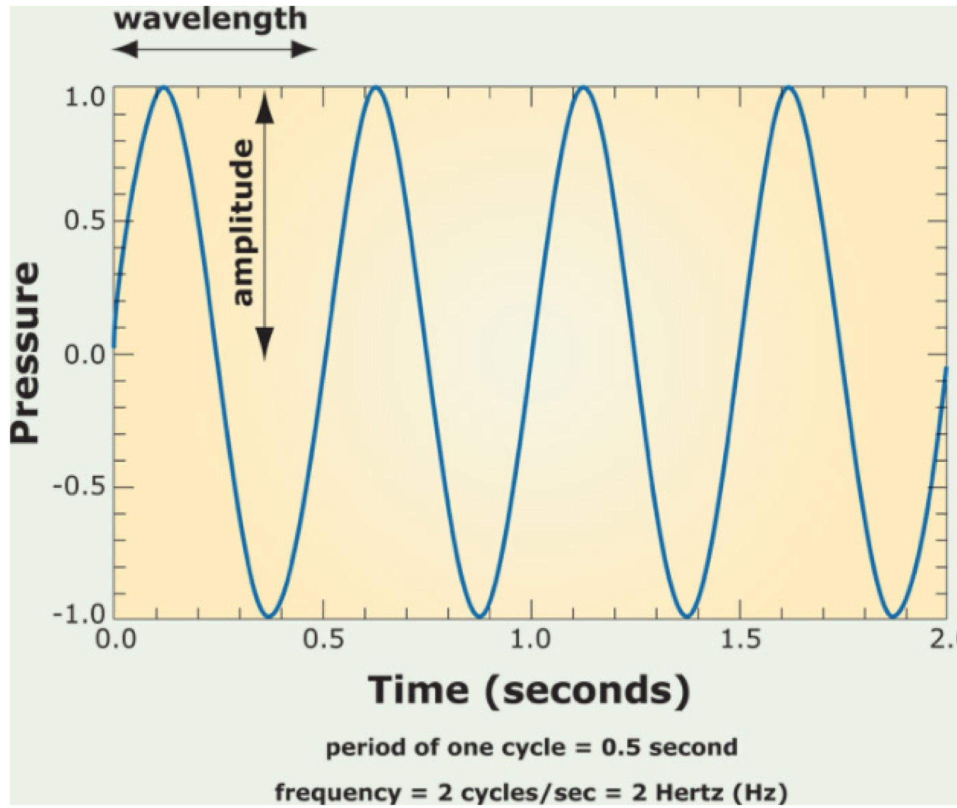
Sound Wave

- sound is oscillating pressure travelling through some medium(air/water).
- Important sound quantities:

amplitude: maximum displacement from equilibrium.

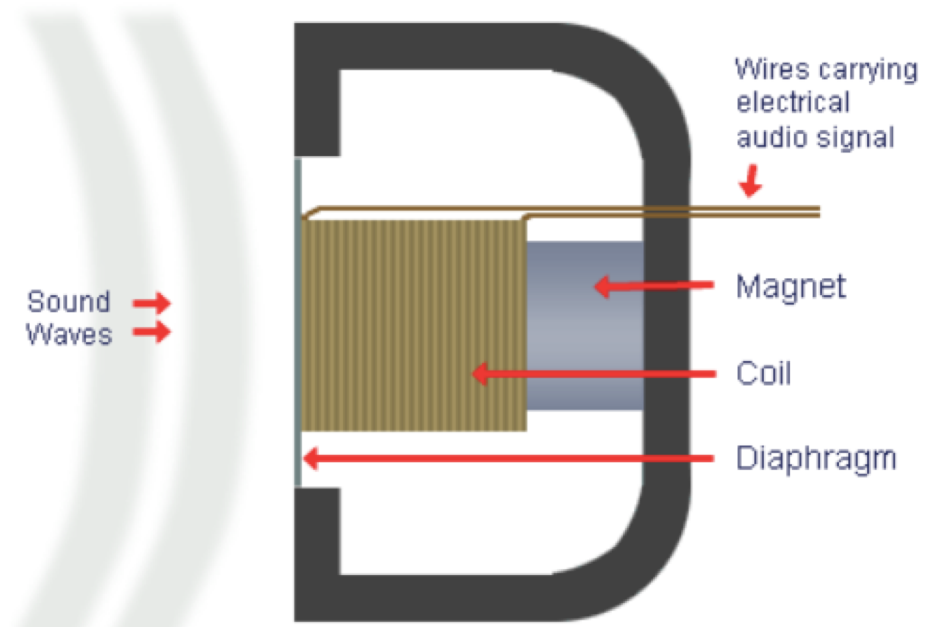
period: time to complete one cycle.(sec or sec per cycle)

frequency: reciprocal of period (cycle per second or hert(hz))
(frequency of a note determines its **pitch**)



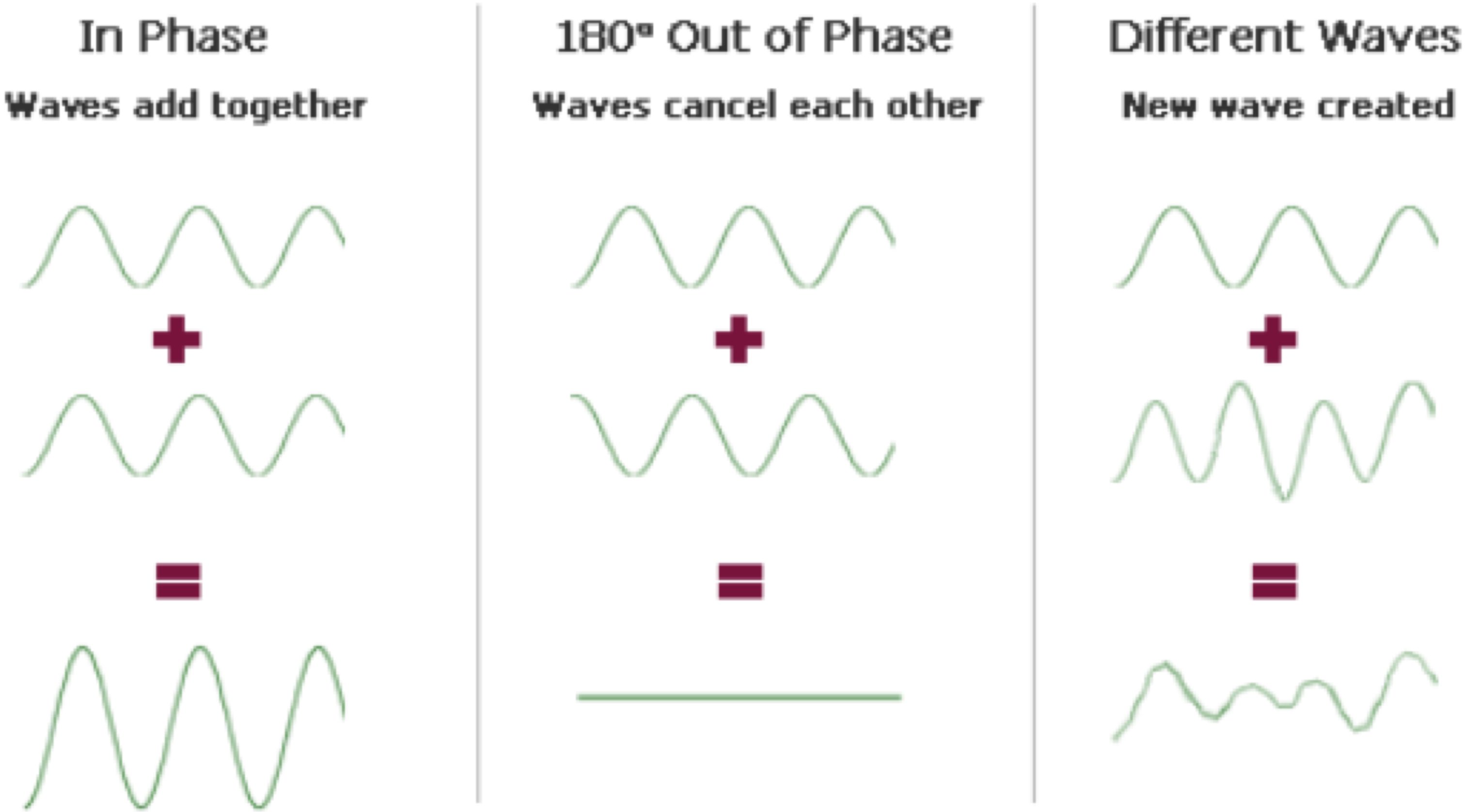
Microphone vs. Speaker

- Sound(pressure waves) vibrate a diaphragm(plastic or paper)
  - which moves the electromagnet(coil) back and forth
  - and causes the magnetic field to fluctuate
  - Inducing a current = audio signal
- Speakers work in reverse: An electrical signal cause the magnetic field to change and vibrate the diaphragm to produce the sound waves.



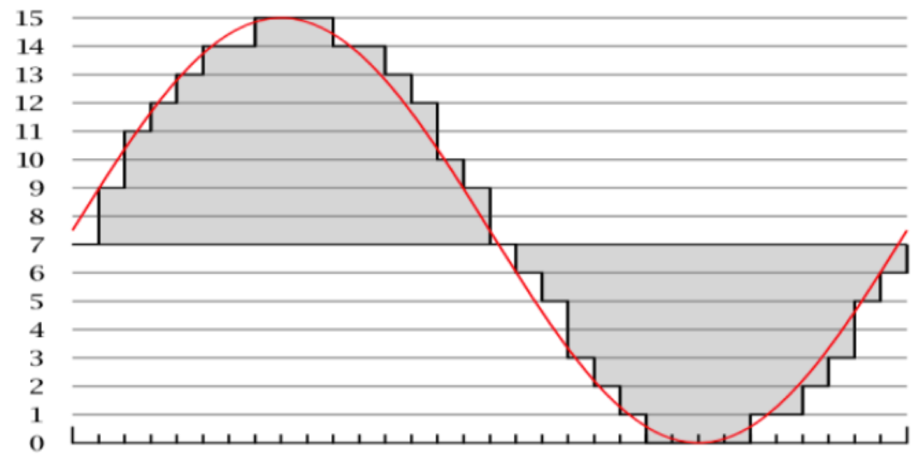
Sound Wave

- sound waves interact to produce different waves.



Digital vs. Analog

- sound can be modeled as a continuous, **analog** wave or signal.
- computers can only understand **digital** signals (discrete or finite signal(0s and 1s).)
  - analog signals are continuous and can take on an infinite possible values(the real numbers)
  - digital signals are finite.
  - For example, 8 bit colors can take on one of 256 discrete, finite possibilities. But actual colors can take on any of an infinite possible values or shades.
- **sampling** allow computers to approximate analog signals such as sound.
- the number of samples is the **sampling rate**, the higher the rate the better the quality.



Sampling

- CD-quality has a rate of 44,100 samples per second(44,100 Hz or 44.1 kHz).

Analog

Low sampling rate

High sampling rate

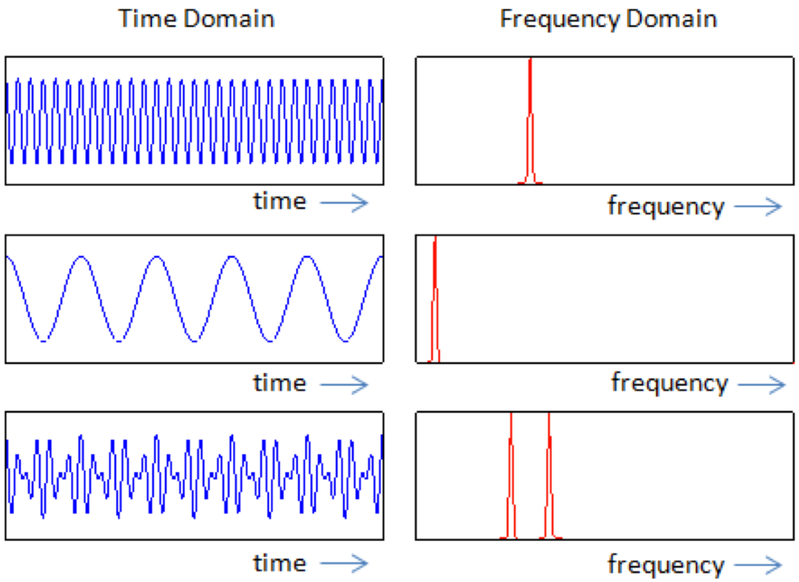
- the possible values for each sample is determined by the **bitrate** or bit depth.
  – CD usually has 16-bit audio which means it can take on one of 2^16 or 65,536 values.
  – This is analogous to 8-bit color for example taking on 2^8=256 possible values.
  – More bits means more accuracy but more memory.

MP3

- A popular format to store audio is MP3(MPEG-2 Audio Layer III)
- CD-quality(**PCM, pulse code modulation**) sounds require a lot of memory.
  - Each of 44,100 samples require 16 bits(32 bits for left and right channel), that's approximately 0.2 MB per second. A 3-mins song would require about 30MB of memory! MP3 is a lossy compression format and only need 3MB!
- MP3 compression uses **psychoacoustics** to take advantage of limitations in our biology and psychology of sound.
  - human hearing range is 20 Hz to 20 kHz
  - hard to differentiate between similar sounds(similar frequencies).
  - humans are only good at differentiating pitches in the 2 kHz and 5 kHz range.
  - in compression, sounds outside of human hearing range is thrown out and similar sounds can be combined.
  - **frequency masking**: if we hear a pair of sounds with similar frequencies at the same time, then the higher frequency will "mask" the lower frequency, which means that we won't hear it.

Compression

- In order to determine which data from the PCM encoding can be thrown away, your computer uses two different tools.
- **Fast Fourier Transform(FFT)**: convert sound from the time domain to frequency domain and filter out unnecessary frequencies.
- The other is called the **Modified Discrete Cosine Transform**, or MDCT for short, which analyzes the entire spectrum of frequencies of the audio.
- The result of this process is a visualization called a **spectrogram** which is essentially a fingerprint of the sound's frequencies. (Shazam app uses these techniques to identify songs.)



Popular Formats

- Other popular audio formats include AAC, WMA, WAV, FLAC.
- AAC(Advanced Audio Coding) was developed by Apple for iTunes but is currently adopted by many devices.
  - more efficient compression and better quality than MP3
- Windows Media Audio(WMA) is a proprietary format developed by Microsoft but has not been as popular as AAC or MP3.
- Waveform Audio(WAV) was also developed by Windows
  - uncompressed
- FLAC uses lossless compression and is an open format.
  - popular as an archival format

Video

Video Codecs

- When dealing with video on your computer, we're concerned with two different things: **codecs** and **containers**.
- We saw that PCM audio can take up a lot of space, but uncompressed video is even larger. Without any compression, an hour of high quality video can be hundreds of gigabytes large!
- The role of the codec is to do just that, as the codec (a combination of the words "encode" and "decode") is a program responsible for compressing and decompressing video.
- Today, the most popular standard for compressing videos is called **H.264**, sometimes referred to as AVC, or advanced video coding.
  - For example, all Blu-Ray players must be able to decode video that has been encoded according to the H.264 standard.
  - There are lots of codecs out on the Internet that compress video using H.264, including x264 and DivX.
  - software like VLC has built in codecs that can generally play any video file.

Containers

- video files are packaged up into a **container** which include video/audio files, captions, DVD menus, etc..
- popular video containers today include AVI, MKV, MP4, and MOV, and different container formats support different codecs for audio and video.
- Containers give us a bit of flexibility in creating videos, since we get to choose which codec we want to use for the video and which codec we want to use for the audio, though not all containers support all codecs.
- However, because codecs and containers are separate concepts, it could be the case that your media player supports a certain container format but not the codec used for the video in the container, which means you won't be able to play the video!

Video Compression

- When we looked at lossless compression for images, the basic idea was that compressed images avoided storing redundant data for similar pixels.

- This same principle is applied to video compression.

- In a movie or television show, it's usually the case that the entire scene isn't changing all at once. Instead, there's usually some subject in the video that's moving on a more static background.

- So, if nothing in the background is changing between frames, then we don't need to redundantly store that data!

Video Compression

- notice the chair on the left is static, it never moves in these frames.
- we don't need to save those data over and over again.
- that means we can actually save something that looks more like this:

Streaming

- **Streaming** gives us the ability to watch a video (or listen to an audio track) as it's being downloaded.
- a YouTube video file will be downloaded by the browser in small chunks (think less than 10 seconds of video).
- So that the video remains smooth as you're watching it, your browser actually needs to make sure it stays a bit ahead of you as the viewer.
  - the browser download ahead into a **buffer**.
  - "buffering" related errors are due to slow internet connection.
  - for slow connections, the server can choose a lower resolution video.(**adaptive bitrate streaming**)
- Common video resolutions are 1920x1080, 1280x720, and 854x480.
  - These are common denoted by 1080p, 720p and 480p(height of video).
  - Notice that in all three cases, the ratio between width and height is 16:9, which is known as the image's aspect ratio. (4:3 is an older ratio)

GPU

- Back in our hardware section, we briefly mentioned the role of the graphics card in drawing things to the screen.
- One component of the video card is the **GPU, or Graphics Processing Unit**, which is essentially a CPU dedicated just to graphics-related computations.
- We saw earlier how parallelism can lead to more efficient solutions to computing problems. GPUs take parallelism to a whole new level by running thousands of computations at the same time.
- It also helps if these problems are independent of one another, so everything can be solved at the same time with minimal need to combine everything back together. Luckily, both of these conditions are really common in graphics!
  - For example, figuring out the color of the pixel in the top-left of your screen doesn't depend on figuring out the color of the pixel in teh bottom-right of your screen.

Homework

1)Read and reread these lecture notes.

2)Watch or Read(Required):

a)MP3 Compression.

https://arstechnica.com/features/2007/10/the-audiofile-understanding-mp3-compression/

b)CPU vs GPU Painting.

https://youtu.be/-P28LKWTzrI

c) Sampling

https://www.youtube.com/watch?v=zC5KFnSUPNo

d) Analog vs Digital.

https://www.youtube.com/watch?v=btgAUdbj85E

References

This lecture is a summary of a lecture from an OpenCourseWare course below.

Computer Science E-1 at Harvard Extension School

Understanding Computers and the Internet

by Tommy MacWilliam

https://www.youtube.com/watch?v=-R2uBgcw600