Representing Images and Sound

Dr. Abdallah Mohamed

Acknowledgement: Original slides provided courtesy of Dr. Lawrence.

Key Points

1) It is possible to digitize the naturally analog information of sound, images, and video.

2) Due to the large size of digitized images/video, compression is needed to make it more efficient to use and store the information.

Overview

Most of the information in the real world is not digital by nature.

Although we saw some reasonable encodings for numbers and characters, it is a little more complex to store images and sounds on a computer.

Images and sound are analog by nature. To convert to digital, we must sample the original, encode it, and then compress it to make it usable.

The increasing power of computers has made the virtual reality that can be produced more and more realistic.

Review: Digitizing Color

Recall that computers represent different colors by giving the intensities of red, green, and blue (RGB).

- Each red, green, and blue value was a number from 0 to 255 that we can represent in decimal or hexadecimal.
- Black is no color (all values are 0).
- White is full intensity of RGB (all values are 255).

Color Question

Question: What is the best description of color code: #B3009F?

A) a shade of purple

B) a shade of yellow

C) a shade of blue

D) a shade of green

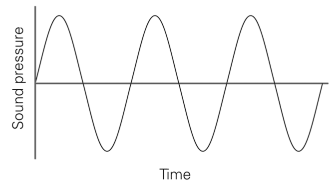
Digitizing Sound

An object creates sound by vibrating in a medium such as air.

Vibrations push the air and pressure waves emanate from the object and vibrate our eardrums.

The force, or ***intensity*** of the push determines the volume.

The ***frequency*** (number of waves per second) is the pitch.



Converting Analog Sound to Digital

Sound is analog (continuous) by nature.

To digitize sound into bits, we need to record with a binary number the amount by which the wave is above or below the zero line. (positive or negative sound pressure)

However, we cannot possibly record a value at every point in time, so we use ***sampling*** to collect information at certain intervals (points in time).

The ***sampling rate*** is the number of measurements per second. The higher the rate, the more accurate the digitization (but more space is required).

How Fast a Sampling Rate?

Sampling rate should be related to the wave's frequency.

- Too slow a rate could allow waves to fit between the samples causing us to miss segments of sound.

Guideline is **Nyquist Rule**: Sampling rate must be at least **twice as fast as the fastest frequency**.

Human perception can hear sound up to 20,000 Hz, so 40,000 Hz sampling rate is enough.

- The standard for digital audio is 44,100 Hz.

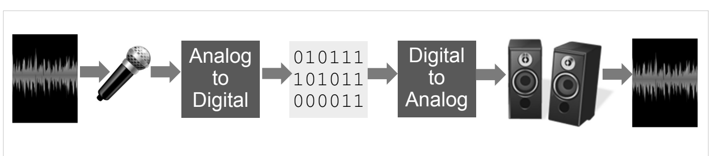
Analog-to-Digital Converters

An ***analog-to-digital converter*** (ADC) samples a signal at regular intervals and outputs binary numbers into memory.

- Microphone is an example of an ADC.

To play sound, ***digital-to-analog converters*** (DAC) are used that receive binary numbers as input and output an electrical wave by filling in between the digital values.

- The electrical signal is output to a speaker which converts it to a sound wave.



How Many Bits per Sample?

How accurate must the samples be?

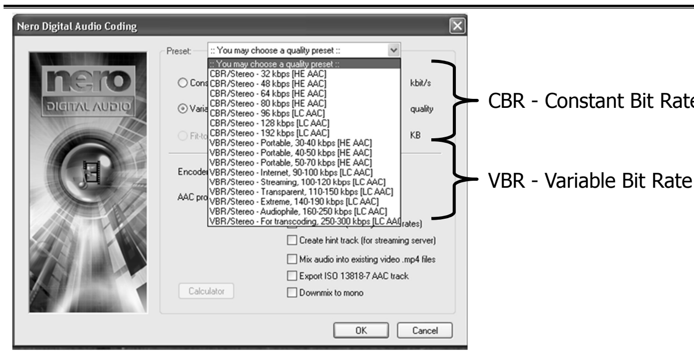
- Bits must represent both positive and negative values.
- The more bits, the more accurate the measurement.
- The digital representation of audio CDs uses 16 bits (records 65,536 levels, half above and half below the zero line).

Advantages of Digital Sound Representation

The advantages of digital representation:

- 1) All digital representations can be computed on (**manipulated digitally**). This makes it easier to edit and change them.
- 2) Reproducing the data can be done **exactly**.
  - Bit file can be copied without losing any information.
  - Original and copy are exactly the same.
- 3) **Compression** - Compression techniques such as (MP3 compression) allow for more compact representation.
  - Remove waves that are outside range of human hearing.
  - MP3 usually gets a compression rate of 10:1.
  - MP3 stands for MPEG level 3 ("sound track" of MPEG digital video).

Example: Digitizing CDs (CD ripping)



Quality is expressed in kbps. (kilo bits per second).
- If sampling rate is 44.1 kHz, what is the kbps for CD audio (16-bit samples)?
- For 192 kbps, what is the sample size? Does that sound right?

Digital Compression

Question: A music digitization program provides the two encodings below. Which encoding has the largest size (assuming no compression)?

A) Sample at 50 kHz and encode 16 bits per sample

B) Sample at 10 kHz and encode 32 bits per sample

Digital Quality

Question: A music digitization program provides the two encodings below. Which encoding has the best sound quality?

A) Sample at 50 kHz and encode 16 bits per sample

B) Sample at 10 kHz and encode 128 bits per sample

Digitizing Images and Video

Just like sound, images and video is encoded by sampling. For images, a sample consists of how many measurements are taken over an area.

- For instance, when scanning, you can determine how many pixels (samples) per inch you will take.
- The more samples (higher pixels per inch) the finer detail the image will be encoded.

Movies are sequences of individual images.

Two Ways to Encode Digital Images

Bitmap representation stores a 2D matrix (width x height) with a color intensity at each pixel.

- Examples: PNG, JPEG, GIF, TIFF

Vector representation describes an image as a sequence of lines or shapes each with a color.

- Examples: fonts (as scale better), Postscript (eps), PDF, Scalable Vector Graphics (svg)

Bitmap format is good for complex images and can be compressed. However, it does not scale well.

Vector representation is good for line art and text and will have smaller sizes for those types of images.

Digitizing Images and Video Compression

Without compression, storing images would be impractical.

Compression may be ***lossless*** (no information is lost) or ***lossy*** (information may be lost during compression).

JPEG compression can compress images.

- JPEG is a lossy compression scheme that makes images much smaller and the picture quality is controllable.
- Since our eyes are not very sensitive to small changes in hue, (but are sensitive to small changes in brightness), stores a less accurate description of hue (fewer pixels).
- Gets a 20:1 compression ratio without eyes being able to perceive the difference.
- The actual compression algorithm is beyond our scope.

PNG is a lossless compression method.

- Best for text and line art.

JPEG Example

Full quality – 83,261 bytes (2.6:1)

Average quality – 15,138 bytes (15:1)

Medium quality – 9,553 bytes (23:1)

Low quality – 4,787 bytes (46:1)

Source: Wikipedia - http://en.wikipedia.org/wiki/JPEG

JP

Aside: JPEG in Digital Cameras

Most digital cameras use some form of JPEG compression and often provide you with a setting that indicates image quality.

Smaller images (and thus more images on the camera) come at the cost of lower quality. Probably better to select high quality!

Example:

- Nikon D5000 – 12.3 megapixel sensor (4288 x 2848)
- RAW format is 12-bits per pixel: 18.45 MB (uncompressed) and 10.6 MB (compressed)
- JPEG: high quality: 5.9 MB, medium: 3.3 MB, low: 1.5 MB
- Source: Nikon

MPEG Compression Scheme

The **MPEG** compression scheme follows the same idea as JPEG, but is applied to motion pictures.

Two "levels" of compression:

- 1) JPEG-like compression is applied to each frame.
- 2) Then "**interframe coherency**" is used so that only record and transmit the differences between one frame and the next. This results in huge amounts of compression.

Digital Compression

Question: True or false: An MP3 performs lossy compression.

A) true

B) false

Optical Character Recognition

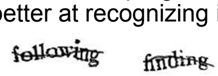
Optical Character Recognition (OCR) is the process of analyzing captured images to determine its contents.

- Example: Scan in a document and have the computer convert the document into a text document rather than an image.

OCR is used in other areas:

- auto-mail sorters
- photo radar cameras (license plate numbers)
- handwriting recognition
- fingerprinting technology (biometrics - eye, fingerprint, etc.)

Captchas used on web sites to stop automated programs are based on the idea that humans are better at recognizing image patterns that computer algorithms.



Virtual Reality: Fooling the Senses

Input and output devices can use all senses to engage the user in the virtual reality experience.

- Sound and sight we have seen already.
- Smells - it has been done, but not well.
- Taste - not really..

Touch has been increasingly used to communicate realism. Examples including vibrating controllers and interactive devices that provide motion and vibration that mimics real world cues.

- These ***haptic devices*** engage our sense of touch.

Advanced: The Challenges of Bandwidth and Latency

Although images, sound, and video are represented digitally, two issues challenge the construction of a virtual reality.

- ***Latency*** is the time it takes for information to be delivered.
  - Too long a latency period ruins the illusion as we can sense the delay.
  - Absolute limit to how fast information can be transmitted—speed of light.
- ***Bandwidth*** is the rate at which information can be delivered.
  - Bandwidth is important as digital encodings, even with compression, consume a lot of space.

Conclusion

Sounds, images, and video are digitally encoding by ***sampling*** the analog input and encoding each sample in bits.

The raw samples consume significant amounts of space, so they are ***compressed*** to make them faster to process and smaller to store.

Although increasing computer power has made virtual reality more realistic, continuing work is performed on compression and techniques to improve ***bandwidth*** and reduce ***latency***.

Objectives

- Define: intensity, frequency
- Define: sampling, sampling rate
- Define: Nyquist Rule
- Explain the purpose of analog-to-digital and digital-to-analog converters.
- List two advantages of digital sound.
- Compare and contrast: lossy and lossless compression
- Define: JPEG, MPEG, haptic device, OCR
- Compare the difference between representing images using bitmaps or vectors.
- Define: bandwidth, latency