Digitisation

Sound and images are analogue — complex waveforms
Analogue signals need to be digitised — sampling and quantisation
Size of digitised data — sampling and quantisation rates

Analogue sound and images

sound: results from motion of particles through a transmission medium
Mechanical wave
产生|传输|机械波

Sound wave — Sinusoidal functions

Sound wave — Complex waveforms
Each complex wave can be regarded as the sum of simple sine waves with different amplitude and different frequency
Example: Visible light
SPD (Spectral Power Distribution) : how intensity of light from some source varies with the wavelength of light

Fourier analysis
原理： attempts to represent complex wave with a series of sine and cosine waves with different amplitudes, periods and phases
作用： the data in time domain are transformed into the frequency domain

Information in a wave
Pure tone: A sound represented by a completely regular sine wave
Pitch: the frequency information of sound
Loud: the amplitude information of sound
Color: the wavelength of a light ray (颜色与波长相关)

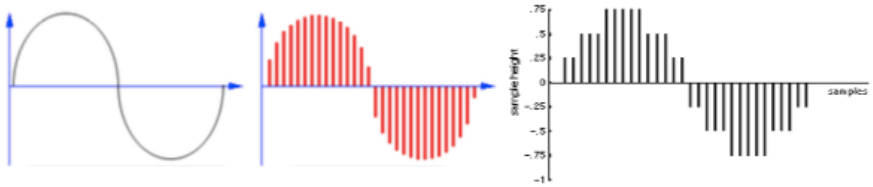
Sampling and quantisation

Analogue-to-Digital conversion (ADC): converting the continuous phenomena of images, sound, and motion into a discrete representation that can be handle by a computer 连续 —> 离散 — 能被电脑接受
Digital signal 的好处：
Digitised sound and image can be captured in fine details (能抓取细节)
Digital data communication is less vulnerable to noise (抗噪声)
Digital data can be communicated more compactly when compressed (更简洁、压缩)

Sampling
Sampling: chooses discrete points to measure a continuous signal

images: evenly sperated in space, sound: evenly seperated in time
Sampling rate: the number of samples taken per unit time or unit space

Quantisation
Quantisation: each sample should be represented by a fixed number of bits
bit depth | sample size | color depth (image) — limit precision
$m = 2^n$ — m different values



sampling — frequency | quantisation — amplitude

Aliasing (sampling error)
Undersampling causes aliasing, the original data cannot be reproduced

Nyquist theorem:
To avoid aliasing, assume the highest frequency of the signal that we interest is fm, we need to sample the signal with higher than twice of fm.
$fs \geq 2 \cdot fm$, $fs = 2 \cdot fm$ is the nyquist rate

Quantisation error
As bit depth increases, the precision increases. For image, it will lose the colors (loss details in general).

Signal to Noise Ratio (SNR)
SNR: the ratio of the meaningful content of a signal versus the associated noise
Analogue: SNR is the ratio of average power in signal versus the power in noise level
Digital: signal-to-quantisation ratio (SQNR)

SQNR (in decibels)
SQNR: the ratio of the maximum sample (quantisation) value versus the maximum quantization error
Dynamic range: related to SQNR, the ratio of the largest signal amplitude and the smallest that can be represneted with a given bit depth
$SQNR = 20 \log_{10}(2^n)$

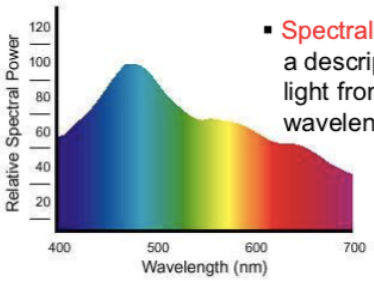
The size of the digitised data
Stereo — 2 channels — 32 bits
Mono — 1 channel — 16 bits
CD quality: 44.1 kHz

Colours

Colour — a property of light
Colour vision
Colour models — RGB | YUV/YCbCr | CMYK | HSV

Colour is a property of light

Light



- Spectral Power Distribution (SPD): a description of how the intensity of light from some source varies with wavelength.

1. The narrower the band, the pure the colour (If there is a pure green light, the SPD will only distributed at the certain wavelenght which the bandwidth will be narrow)
2. Pure spectral colour: Colours that correspond to narrower wavelength bands

Properties of Colours
Hue: the name of the colour
Saturation: a fully saturated colour is one with no mixture of white
Brightness: the extent to which an area appears to emit light
Color is a property of light, objects have no color of their own, but merely the ability to reflect a certain section of visible spectrum

Brightness and Luminance
Brightness: subjetive perception
Luminance: a mathematical definition, relates with wavelength and power
note: lights of equal power, different wavelength do not appear equally bright
The brightest wavelength are about 550nm

Color Vision

Color perception is subjective
Color perception is 1. determined by physical context of the object 2. Affected by previous experience (brain)
Color perception is difficult:
a. Varies from person to person
b. Affected by adaptation
c. Affected by surrounding colors

Retina — two parts:
rodes — night vision | cones — colour vision
Cones: (3 types) 1. L or R: red light (610 nm)

2. M or G: green light (560 nm)
3. S or B: blue light (430 nm)

Color blindness — resulting from missing cone types

The Tristimulus Theory
Human perception of colour derives from the eye's response to 3 different groups of wavelengths, i.e. red, green, blue. Therefore, any sensation of colour can be produced by mixing together suitable amounts of these colours.

Metamers
Metamers: two colours that produce the same tristimulus values are visually instinguishable though they may cause by different spectra
Identical perceptions of color can be caused by very different spectra
(同色异谱)

Colour models

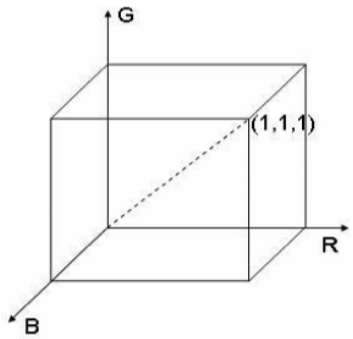
Colour model: an abstract mathematical model describing the way colours can be represented as sets of numbers

RGB: used in computer monitors
YUV/YCbCr: used for image compression
CMTK: used for printing
HSV: an intuitive colour model

RGB
Define colour value in terms of the proportions of R, G, B (r,g,b)
Colour depth: the number of bits used to hold a colour value
note: (0,0,100) is also a saturated dark blue, to be unsaturated, mixed up with R and G — (20,20,100)

RGB to grayscale : R = G = B — a straight line running from the origin (0,0,0) to (255,255,255)

Basic colours in RGB



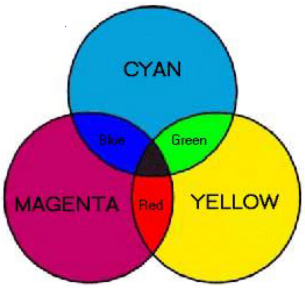
(0,0,0): black
(255,0,0): red
(0,0,255): blue
(255,0,255): magenta
(0,255,0): green
(255,255,0): yellow
(0,255,255): magenta
(255,255,255): white

Lower value — dark, higher value — bright, saturated and bright: 255
YUV

The YUV model is a simple translation of RGB model, seperating all the luminance information (Y) from the colour (chrominance) information (U,V)

CMYK

The CMYK model is a subtractive model that is used in printing



Subtractive Primary Colours

C = W-R
M = W-G
Y = W-B

Whereas the RGB model depends on a light source to create colour, the CMYK model is based on the light-absorbing quality of ink printed on paper. It uses the subtractive primaries Cyan, Magenta and Yellow.

显色原理:

颜色的本质是某物质吸收了特定波长光后剩余的光反射到眼睛里所产生的视觉，比如红色就是那种物质专吸收了绿色、蓝色和其他波长的光，反射了红色光，而黑色则是吸收了所有光色。

Adding K to CMY

To obtain black in CMY,

1. large quantity ink — waste of ink
2. Does not allow a perfect black in pratice

HSV

H: hue, S: saturation, V: value/brightness (hexagon)

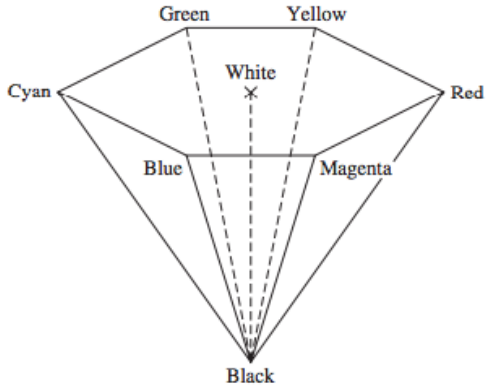


Figure 2.47 HSV color space, a hexacone

H	S	V	Color
0	1.0	1.0	Red
120	1.0	1.0	Green
240	1.0	1.0	Blue
*	0.0	1.0	White
*	0.0	0.5	Gray
*	*	0.0	Black
60	1.0	1.0	?
270	0.5	1.0	?
270	0.0	0.7	?

grayscale: the vertical axis represent grayscale colours

Digital Images

Images can be stored in two ways: bitmaps | vector-based
Picture elements (pixels)
Color bitmap images: true-colour | index-based

Bitmaps and vector-based

Image files
Images are displayed as a grid of "pixels" of various colours. The iamge files store the image data in one of two ways:
Bitmapped images (GIF or JPEG)
Vector-based images

Bitmapped image
Bitmap image: created with a pixel-by-pixel specification of points of color

Vector-Based image
Vector-Based image: use object specifications and mathematical equations to describe shapes to which colours are applied

Advantage & Disadvantage:
Bitmapped disadvantage:
1. Bigger
2. Less scalable (will loss quality)
3. When zooming, see individual pixels
4. Store color information for each pixel — too much data

Vector-based advantage:
1. Can be displayed in any size without losing quality
2. Esay to be edited by changing instructions

disadvantage:
1. Only contain instructions
2. Can only deal with simple images

Pixels

Image sampling and quantisation
A digital image is represented by a matrix of numeric values, each representing a quantized intensity value. e.g. I (row, column)
The intensity at each pixel is represented by an integer and is determined from the continuous image by averaging over a small neighbourhood around the pixel location (用矩阵表示图片，每一个矩阵元素都有一个值I，I由该像素点与临域值的平均值决定)

Image sampling
square sampling grid with pixels equally spaced along the two sides of the grid

(正方形的网格)

Quantization
Quantization: a matter of the color model used and the corresponding bit depth

Digital cameras
use the same digitisation process: sample + quantisation
Sampling rate: how many points of color are sample and recorded in each dimension of the image (typically, 1600x1200, 1280x960, 1024x768, 640x480)
Quantisation in Digital cameras: use RGB, saves each pixel in three bytes, one for each channel (24 bits — 2^{24} = 16,777,216 colors)

Pixel dimensions
Pixel dimension: for an image file, the number of pixels horizontally and vertically (e.g. 1600x1200) — logical pixels
Computer screen has a fixed maximum pixel dimensions (e.g. 1024x768 or 1400x1050) — physical pixels
When displaying image, the logical pixel is mapped to a physical pixel on a computer screen

Image resolution
Image resolution: The number of pixels per inch
Assumed the same number of pixels are used in the horizontal and vertical directions
Typically, monitors have a screen resolution of 72 ppi (pixels per inch)
Image resolution is depended both on pixel dimension and the physical size of the monitor

True-colour or index-based
Colour bitmap images can be true-colour or index-based

True-colour image
RGB formats: true-colour, use 8 bits of data for each R, G, B value, this forms a 24-bit pixel palette, which has 16.7 million colours

Index-based image
Indexed formats: mapped into a smaller colour pallete (CLUT), The indexed image's pallette contains all of the colours that are available for the image

Digital image file type
a. Not all color models can be accomodated by an file types.
b. Some file types require the image be compressed while some do not.
4 important things for bitmap filetype:
1. Color model 2. Bit depth 3. Compression type, if any 4. Operating system, browser, application software that support it

Digital Video and Audio

Video
Audio: speech | Music

Video

Video: technology of capturing, recording, process, storing, transmitting, and reconstructing a sequence of still images representing scenes in motion

Frame rate: the number of still pictures per unit of time of video

Analogue and Digital video
Analogue: recording methods that stores continuous waves of red, green and blue intensities
Digital: recording methods that works by using digital video signals

Refresh rate and Frame rate
Refresh rate: the number of times in a second that the display hardware draws the data (Repeating drawing of identical frames)
Frame rate: the number of still pictures per unit time (in a second) of video (Measures how often a video source can feed an entire frame of new data to display)
Note: if frame rate > refresh rate, the display will not be able to display all of the frames

- Typical rates: 24 frames per second (framerate) and 48 or 72 Hz (refresh rate)

Interlaced and Progressive
Interlaced scanning displays alternating sets of lines, happens too quickly, give the illusion of whole image

- your eyes can't detect what's happening at least you don't perceive the partially drawn images

Progressive video displays the entire image

Audio

Sound
Sound: produced by the vibration of matter, pressure variations —> wave-like motion

Characteristic of Sound waveforms
Frequency: determines pitch

- Infra-sound: from 0 to 20 Hz
- Human hearing frequency range: 20 Hz – 20 kHz
- Ultrasound: from 20 kHz to 1 GHz

Amplitude: determines volume or intensity (subjective: loudness)

Computer representation of sound (Sampling and quantization)
Sampling

Sampling rate: the rate at which a waveform is sample
e.g. the CD standard sampling rate of 44100 Hz means that the waveform is sampled 44100 times / second

Quantization
Quantisatio depends on the number of bits used in measuring the height of the waveform
the CD standard 16 bits quantisation
8-bit — speech | 16-bit — music

Aliasing (sampling error)
The reason a too-low sampling rate results in aliasing is that there aren't enough sample points from which to accurately interpolate the sinusoidal form of the original wave

Measuring Sound Amplitude in Decibels

dB_SPL: decibels-sound-pressure-level

$$dB\_SPL = 20 \log_{10}\left(\frac{E}{E_0}\right)$$

E0 = 0.0002 Pa, air pressure amplitude for the threshold of hearing

- if you increase the amplitude of an audio recording by 10 dB, it will sound about twice as loud
- 3 dB change in amplitude is the smallest perceptible change

Signal to Quantisation Noise Ratio (SQNR)

$$SQNR = 20 \log_{10}\left(\frac{\max(quantization\ value)}{\max(quantization\ error)}\right)$$

Then the signal-to-quantisation noise ratio **SQNR (or dynamic range)** is:

$$SQNR = 20 \log10(2n) = 20n \log10(2) \sim \mathbf{6n}$$

Quantisation Error

Two ways to deal with quantisation error: **Audio dithering + Noise shaping**
Audio dithering: add small random values to samples in order to mask quantisation error
Noise shaping: it redistributes the quantisation error so that the noise is concentrated in the higher frequencies, where human hearing is less sensitive

Speech Signals

Types of Speech Sounds

Voiced sounds : the vocal chords are vibrated, which can be felt in the throat. All vowels are voiced.
Fricatives (unvoiced sounds) : a consonant, such as f or s in English, produced by the forcing of air through a constricted passage.
Plosives (also unvoiced sounds) : a speech sound produced by complete closure of the oral passage and subsequent release accompanied by a burst of air, as in the sound (d) in dog.

Properties of three types of sound
Voiced:
1. Periodic behaviours: during certan intervals, periodic behavious
2. Formants: characteristic maxima in the spectrum of speech signals, caused by resonance of the vocal tract (声道)

Fricatives: looks like noise in the signal
Plosives: clearly starting (with some silence)

Temporal and Frequency Domains
Sound can be represented in both time and frequency domain. They both fully capture the waveform. Store the information about the waveform differently.
A complex waveform is equal to an infinite sum of simple sinusoidal waves, beginning with a fundamental frequency and going through frequencies that are integer multiples of the fundamental frequency (harmonic frequencies)
Fundamental frequency
Harmonic frequnecies: integer multiples of the fundamental frequency

Power Spectrum and Spectrogram
Power Spectrum: the distribution of power into frequency components
Spectrogram: the spectrum of frequencies of a signal as it varies with time

Audio Histogram
Audio Histogram: how many samples there are at each amplitude level in the audio
A statistical analysis of audio files in time domain

Speech Processing Applications
Who: Verification + Identification
- Verification: Prerecorded signal (voice print), matching
- Identification: Database of prerecorded signal

What: Recogintion + Understanding
- Recognition: Recognize the speeches (Convert to sentences)
- Understanding: semantic, progmatic knowlege

How

Music
Digital audio (Two ways to store): MIDI, sampled & quantised digital audio

MIDI
MIDI stores "sounds events" or "human performances of sound"

MIDI data format
Contains message: note on, note off, velocity, aftertouch
Each midi message communicate one musical event
MIDI standards identify 128 instruments

MIDI hardware and software
Controller: generate MIDI message
Synthesizer: read MIDI message, turn into audio signals
Sequencer: recieve, store, and edit MIDI data

MIDI is small
Digital audio files contain thousands of samples of sound, MIDI files only stores the "sounds" events, which are just strings of data. Hence, MIDI files are much smaller

MIDI advantage and disadvantage
Advantages:
1. Small size (much more compact)
2. Completely editable (can change the particular instrument)

Disadvantage:
1. MIDI playback will be accurate only if the MIDI playback device is identical to the device used for production
2. MIDI cannot easily be used to play back spoken dialogue
3. MIDI data is device dependent
4. it can sound more artificial or mechanical than sampled digital audio

Musical Acoustics and Notation
Tones: pitch, timbre, and loudness
note: With the addition of onset and duration, a musical sound is called a note
pitch: how high or low it sounds to the human ear
timbre: tone color

Fundamental frequency and harmonics
The lowest frequency of a given sound produced by a particular instrument is its **fundamental frequency**. Then there are other frequencies combined in the sound, which are integer multiples of the fundamental frequency, referred to as **harmonics**.

$$h = 2^n \cdot g,$$

h is the frequency of a musical tone n octaves higher than g

Lossless Compression

Compression can be lossless or lossy
Lossless techniques:

- RLE
- LZW
- Huffman

Compression can be lossy or lossless

- Lossless encoding
- No information is lost
- Redundant data is not encoded
- Motivation: some data appear more often than other and some data normally appear together

Lossy enocoding

- details not perceived by users are discarded; e.g. small change in lightness (luminance) noticed more than change of colour details
- approximation of the real data (sacrifice some information)

Lossless compression — applied many times, Lossy compression — applied only once

Compression rate
Compression rate: ratio of the original file size a to the size of the compressed file b, expressed as a:b

Types of Compression Algorithms
Dictionary-based, entropy, perceptual …

Dictionary-based methods (e.g. LZW compression): use a look-up table of fixed-length codes, where one code word may correspond to a string of symbol

Entropy compression: use a statistical analysis of the frequency of symbols and achieves compression by encoding more frequently-occurring symbols with shorter code words

Run-Length Encoding (RLE)

Exploits **spatial redundancy**
Physically reduce any type of repeating sequence (run), once the sequence reach a predefined number of occurrence (length)
Idea: store number pairs (c, n), where *c* indicates the pixel value and *n* is how many consecutive pixels have that value. (Reduce repeating sequence)

RLE: Calculating the length (Method 1)
First, determine the size of the largest run of colors, r (pre-processing)

RLE: Calculating the length (Method 2)
Choose a bit depth d in advance, if more than 2^d consecutive pixels of same colours, the runs are divided into blocks

RLE suitable image examples (Country flags)
Long identical pixels

LZW (Lempel-Ziv-Welch)

LZW
Exploits **spatial redundancy**
Commonly applied to GIF and TIFF image files
Motivation: Sequences of color in an image are often repeated (sequence of characters in a text file) (Some data normally appears together)

Dictionary-based encoding:
encode variable-length strings of symbols as single tokens (indexes).
Tokens form an index into a phrase dictionary
If tokens are smaller than phreases they replace, compression occurs

LZW for text
The dictionary initialised with 256 entries representing ASCII table
9 bits per code: first 256 — ASCII characters, Second 256 — strings of length 2 or more
12 bits per code: 4000 entries

LZW for image
The dictionary initialised with individual colours that exist in the image file

LZW suitable image case
Large distribution of similar colour distribution blocks

Huffman encoding

Exploits **statistical redundancy**
Ideal: Encode frequent bit patterns with short code, infrequent bit patterns with longer codes
Fixed-length inputs become variable-length outputs

Shannon's entropy equation

$$H(S) = \eta = \sum_i p_i \log_2\left(\frac{1}{p_i}\right)$$

Entropy and bit depth
the minimum value for the average number of bits

Huffman encoding
(1) determining the codes for the colours (To determine the codes, a tree data structure is built from the colour frequency table, according to the tree create the corresponding Huffman table, which stores the matching between colour to code)
(2) compressing the image file by replacing each colour with its code

Huffman suitable image case
Probabilities vary widely

JPEG

JPEG is lossy
Transform encoding
JPEG compression process:

- Pre-processing: colour conversion, chroma-subsampling, 8x8 blocks
- DCT
- Quantisation
- Lossless encoding in JPEG

JPEG

JPEG = Joint Photographic Experts Group
JPEG: standard for compressing and decompressing still images

- 适用于彩色和灰度图
- 适用于任何长宽比例 (aspect ratio)
- 适用于任何图片（复杂程度）， 任何spatial and statistical性质
- 用户确定compression ratio

JPEG is lossy
issue: artifacting — parts of the image become noticeably blocky

Transform Encoding (Lossless)

Main idea: changing the representation of data (spatial (time) domain - frequency domain) can sometimes extract details that can be removed because they are beyond the acuity of human perception
Two useful transform (digital media): DCT, DFT

High frequency components
Quick fluctuations of colour in a short space — changes that can't capture by human eyes (not easy to see)
Once the high frequency components have been seperated out, remove them

Frequency in Digital Images
Frequency: the rate at which colour values change

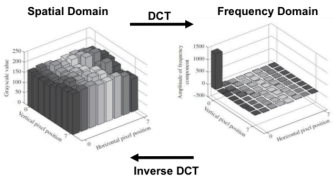
The Discrete Cosine Transform
DCT transform an image from the spatial domain (i.e. pixel values), to the frequency domain (i.e. coefficients of prequency component)
DCT is generally applied to **8x8 pixel blocks**

Base Frequencies

$$\cos\left(\frac{(2r+1)u\pi}{16}\right)\cos\left(\frac{(2s+1)v\pi}{16}\right)$$

DCT
input: a bitmap image in the form of a matrix of colour values
output: the frequency components in the form of a matrix of coefficients ($F(u, v)$)
A negative coefficient amounts to adding the inverted waveform

DC component: The first element $F(0,0)$ — the DC component is a scaled average value of the waveform
AC component: All the other components



The discrete cosine transform is invertible
The transform encoding process is lossless

A negative coefficient amounts to adding the inverted waveform

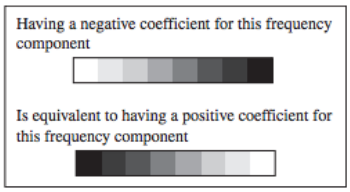


Figure 2.23 A frequency component with a negative coefficient

There are 8 basis function in 1-dimensional DCT. If the coefficient in basis function F(1) is negative, then we can invert the F(1) and then apply the corresponding positive coefficient.

Steps of JPEG compression

algorithm jpeg
/*Input: A bitmap image in RGB mode. Output: The same image, compressed.*/
{
(Convert image to Y'CbCr and do chroma sub-sampling)
Divide image into 8 x 8 pixel blocks
Use Discrete Cosine Transform (DCT) to transform the pixel data from the spatial domain to the frequency domain
Quantise frequency values
DPCM and zigzag order
Do run-length encoding
Do entropy encoding (e.g., Huffman)
}

Summary of the JPEG structure (4 steps):

1. Preprocessing
2. DCT transform
3. Quantising
4. Lossles encoding

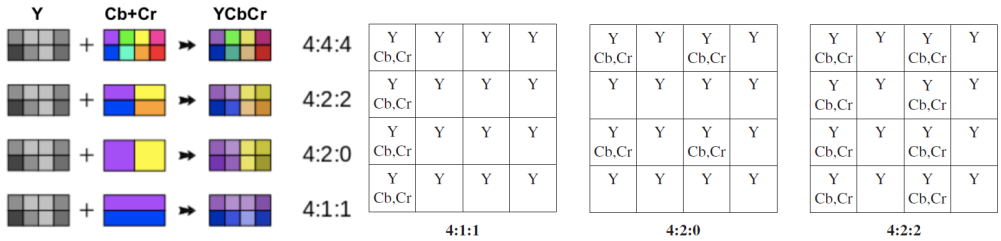
Preprocessing

Three steps:

- Colour conversion: RGB to YCbCr
- Chroma Sub-Sampling
- Divde into 8 x 8 pixel blocks

Colour Conversion: RGB to YCbCr
Using the relationship between these two colour models

Chroma Sub-sampling (lossy)
动机: The human eye is more sensitive to changes in light (luminance) than in colour (chrominance)
Chroma sub-sampling: a process of throwing away some of the bits used to represent colour information
具体过程: save only one Cb value and one Cr value but four Y values for every four pixel values



In JPEG, we choose 4:2:0 form for Chroma sub-sampling
Compression rate: 2 : 1 — colour images achieve greater compression rate

Divided into 8 x 8 pixel blocks
With YCbCr colour model, divide the image into 16 x 16 pixel macroblocks
Then for each 16 x 16 pixel macroblocks, with 4:2:0 chroma sub-sampling, we get four blocks of 8x8 Y data for every one block of 8x8 Cb and one block of 8x8 Cr data.

DCT Transform

Shifting Pixel Values (-128)
On an intuitive level, shifting the values by −128 is like looking at the image function as a waveform that cycles through positive and negative values
具体过程: shifting pixel values, DCT Transform, obtain a 8 x 8 matrix of frequency component coefficient matrix

Quantise frequency values

DCT coefficient matrix

Values are arranged from lowest frequencies to highest frequencies

Lowest frequencies: represent scaled average value for the block

Highest frequencies: represent fine detail (can be dropped)

Quantisation (lossy)

动机: 1. Eye unable to perceive brightness levels above or below thresholds 2. Gental gradation of brightness of colour are more sensitive to the eye than abrupt changes

具体过程: Quantisation involves dividing each frequency coefficient by an integer and rounding off. Rouding off during quantisation makes JPEG lossy. The coefficient for high frequency components are small, so they will be rounded down to 0, which means they are removed or thrown away.

$$F^{Q}(u,v) = Integer\ Round\left(\frac{F(u,v)}{Q(u,v)}\right)$$

Quantisation table

Qunatisation table: store the quantisation integer in the matrix form. It will be stored with the compressed image.

In the top left area, quantisation values are small, since coefficent of that area should not be lost as they contain fundamental information about image

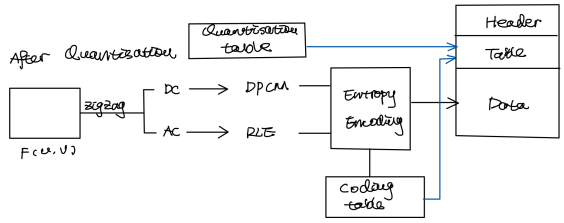
In the bottom right area, quantisation values are large, since coefficient of that area can be lost as they contain fine details that human eyes are hard to captured with.

Quantisation factor

Can be used to increase or decrease the quantisation value.

Higher q value means greater compression and less quality, lower q value means less compression and greater quality.

Lossless encoding



DPCM (differential pulse code modulation)

DPCM: recording the difference between consecutive values rather than the actual values

何时有效: consecutive values don't change very much — fewer bits for recording the change

In JPEG, the upper leftmost value in an block (the DC component) is stored as the difference from the DC component in the previous block

Zigzag ordering and RLE

The zigzag reordering sorts the values from low-frequency to high frequency components

High freuqency coefficients are mostly 0, If many of them round to zero after quantisation, run-length encoding is even more effective

JPEG encoding

After all these steps, the compressed file is put into a standardised format that can be recognised by decompresser

header: contains global information: type of file, width and height, quantisation tables, Huffman code tables, etc

JPEG 2000: high compression rate and good quality (use wavelets instead of DCT)

MPEG

MPEG
Video Compression: Spatial redundancy | Temporal redundancy
Three types of fram:
- Intra Coded Frames (I)
- Predictive Coded Frames (P)
- Bi-directional Predictive Coded Frames (B)

MPEG 4: media object
Other codecs

MPEG

MPEG = Moving Pictures Experts Group

MPEG-1,2,4: compression
MPEG-7: media content description
MPEG-21: rights management and protection

MPEG-1/mp3: (for CD)
- allow moving pictures and sound to be encoded into the bitrate of a Compact Disc (CD).
- Low bit requirement, Limited to 1.5 Mbps, frame rate of 24-30HZ
- Support only progressive pictures
- Include MPEG-1 MP3 audio compression format

MPEG-2: (DVD, digital TV)
- Standards for broadcast-quality television
- Support progressive and interlacing
- Support high resolution

Video Compression

— Usually lossy
Video: a sequence of pictures (frames)

Inter-frame and intra-frame coding
Inter-frame compression: uses one or more earlier or later frames in a sequence to compress the current frame
Intra-frame compression: uses ony the current frame, which is effectively image compression

JPEG algorithm used for intra-frame coding
moving JPEG (MJPEG) exploits only intra-frame coding (spatial redundancy)

Motion estimation and motion compensation: inter-frame coding
High correlation between succesive frames (temporal redundancy)

Use a combination of actual frame contents and predicted frame contents

Temporal redundancy
Temporal redundancy: Consecutive images contain similariteis
Only the changes are encoded (a subtractive image)

Spatial compression (intra-frame coding)
- Allow fast random access
- Applied to key frames
- I-frames

Temporal compression (inter-frame coding)
- Allow high compression
- Exploit temporal redundancy between frames
- Applied to frames occuring between key frames — based on motion detection
- P-frames & B-frames

Group of pictures (GOP)
Picture —> Slice —> macroblock (16 x 16) —> block (8 x 8)
Typically have one I, several P and Bs

MPEG video encoding

- input frames are preprocessed
  - color space conversion
  - spatial resolution adjustment (chroma sub-sampling)
- frame types (Intra or Inter) are decided for each frame/picture
- each picture is divided into macroblocks of 16 X 16 pixels
- macroblocks
  - are intracoded for I frames
  - are predictive coded or intracoded for P and B frames
  - are divided into six blocks of 8 X 8 pixels
    - 4 luminance and 2 chrominance
    - DCT is applied to each block → transform coefficients
      - » quantized
      - » zig-zag scanned
      - » variable-length coded

Preprocessed —> Decide frame types —> devide into macroblocks (16 x 16) —> intra-coded for I, predictive or intra-coded for P and B —> divide into block (8 x 8) —> DCT quantised —> lossless compression

Three types of frame

Intra Coded Frames (I)
Predictive Coded Frames (P)
Bi-directional Predictive Coded Frames (B)

I-frames (Intra-coded frames)

- Treated as still image and use JPEG
- Compression rate is lowest
- No reference to others (self-contained)
- random access

Motion estimation

具体过程: For a given block in the target frame, finding a matching macroblock within the search windows of the refrence frame (succesive frames).
Measuring the difference between matching macroblocks by two things:

- Motion vectors: position difference within the frame
- Error term: Luminance and colour difference

Matching macroblocks: only if a "close" match can be found

- Evaluate with MSE or etc
- If no suitable match, encode the macroblock as an I-block

Not too many P or B frames

Reason: 1. error will keep propagating until next I frame 2. Delay in decoding (先传I,P再传B)

Three situation

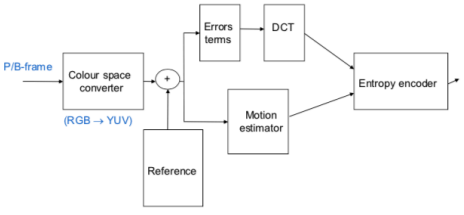
1. One identical macroblock is found: no need to encode anything
2. No matching macroblock is found: use intra-frame encoding
3. One similar macroblock is found: encode error terms and a motion vetor between two macroblocks

P-frames (Predictive coded frames)

- Use a reference frame for motion estimation
- Hold only the changes in the image for a previous frame
- Compression rate is higher
- P-frame can be encoded as either an I-macroblock or P-macroblock (Three situation)

B-frames (Bi-directional Predictive coded frames)

- Serach for matching MBs in both past and future frame
- Compression rate is highest



P- and B- frames
P-frame contains 1/3 data of I-frame
– motion vectors
– error macroblock
B-frame contains 1/2 to 1/5 data of P-frame
– least data but most computational
– **delay issue:** need to transfer future I or P frames before any dependent B-frames can be processed (Transmission order)
Compression rate: B>P>I

Group of pictures (GOP):
- I frames do not rely on other frames
- P frames rely on previous I and P frames
- B frames rely on previous and future I and P frames
- No frames rely on B frames

Transmission order: need to transfer future I or P frames before any dependent B-frames can be processed

Bitrate Allocation
CBR — Constant bit rate
- Streaming media
- Easier to implement

VBR — Variable bit rate
- DVD
- 2-pass coding
- Allocate more bits for complex scenes

MPEG 1&2 summary
Images coded INTRA (I):
– Random access
– Error resiliance (没有太多的错误，不会传播错误)

Images coded INTER(P):
– Prediction from previous decoded image (I, P)

Images coded BI-INTER (B):
– Prediction from previous and / or future decoded image (I, P)
– allow effective prediction of **uncovered background** (areas of the current picture that were not visible in the past and visible in the future)

MPEG 4: encode media objects
MPEG 4: standardised way of representing content as discrete "media objects"

– Objects can be coded and decoded independently
– Object can be flexibly composed to create different scenes
– Natural and synthetic object are treated in the same way
– Excellent error resilience

Other codecs

Codec: a method for encoding and decoding data, a protocol for compressing data
Container: the thing that holds the grouping of compressed video defined by the codecs. Taking care of packaging, transport and presentation

Part 10 and H.264 AVC:
- half the bit rate of MPEG2/MPEG4 Part 2
- 16 reference frames (B-frame: 2)
- block size from 16 x 16 to 4 x 4
- Wide range of bit rates and resolutions

QuickTime
- .mov
- Apple Computer

AVI
- Microsoft
- Windows multimedia container format

WMV (Windows Media Video)
- Adopted as format for Blu-ray dics

Flash Video
- Container file format used to deliver video over the Internet

HTML5 Video
- for the purpose of playing videos or movies

Perceptual encoding

Audio compression
A-law encoding
Psychoacoustics
Perceptual encoding
Mp3 and AAC compression methods

Audio Compression

Lossless Audio Compression: (not strictly)
– Detect silence = samples falling below a threshold
– Treat them as zero and compress using RLE (Run Length Encoding)
Compression rate = 50 - 60%

FLAC: one of lossless audio compression formats:
- Linear prediction
- predictor and error coding
- residual : difference between predictor and smaple data

A-law companding

Companding = compressing + expanding
作用: reducing the bit depth

Linear method for reducing the bit depth
Rounding down has a greater impact on low amplitude samples than on high amplitude ones
Example: 16-bit samples (-32768 ~ 32767) —> 8-bit samples (-128~127)
32767 —> error: 255/32768 < 1%, 0~255 —> error: 100%

A-law companding (non-linear)
Samples are encoded into different number of bits
动机: human ear is more sensitive to quantisation noise in small signals than large signals

The human auditory system is believed to be a logarithmic process in which high amplitude sounds do not require the same resolution as low amplitude sounds: the human ear is more sensitive to quantisation noise in small signals than large signals

Logarithmic quantisation function: smaller signals are represented with greater precision — more data bits — than larger signals

Advantage of A-Law: presume some of the dynamic range, that would be lost if the lower method of reducing the bit depth

Psychoacoustics

- the study of how the human ears and brain perceive sound
- Human hearing is non-linear

Threshold of Hearing

Threshold of hearing = minimum level at which sound can be heard

- Human hear best in: 1000 to 5000 Hz
- Changes with age

Critical Bands

The inner ear is divided into 24 critical bands in human hearing range

动机: Human ability to distinguish frequencies decreases nonlinearly from low to high frequencies

- Critical bands for low frequencies are narrower than those for high ones (原因: Human ear can distinguish better in lower frequencies)
- If two tones are in the same critical band, they are not easily distinguishable as separate, distinct tones

Critical bands example

4 Hz apart, single tone, low frequency modulation or beating

70 Hz apart, a rapid modulation or beating

350 Hz apart, two tones are in different critical bands, can be distinguished

Frequency Masking

Frequency Masking: A loud tone may mask a softer tone of similar or higher frequency

Masking tone: the loud frequency

Masked tone: the quite frequency

Masking threshold

Masking causes the threshold of hearing to be raised within a critical band in the presence of a masking tone. The new threshold of hearing is called masking threshold

Temporal Masking: After a loud sound stops, there is a small delay before we can hear a softer tone

The duration of masking depends on: duration of masker, its amplitude and its frequency

Perceptual encoding

The reason for applying psychoacoustics to compression

Determine the components of sounds that human ears don't perceive very well, those components can be discarded in order to decrease the amount of data that must be stored in digitised sound.

Preceptual Audio Coder (Perceptual encoding)
1. Use convolution filters to divide the audio signal into frequency subbands —> subband filtering.
2. Determine amount of masking for each band caused by nearby band using a psychoacoustic model
3. If the power in a band is below the masking threshold, don't encode it.
4. Otherwise, determine number of bits needed to represent the coefficient.
5. Format bitstream

MP3 and AAC compression methods

MPEG Audio
MP3: MPEG-1 (and MPEG-2) Audio Layer III
- Included in MPEG-1
- 16 bits
- Sampling rate: 32, 44.1, or 48 kHz

AAC: Advanced Audio Coding
- Part of the MPEG-2 and MPEG-4 specification
- More sample frequencies (8 kHz ~ 64 kHz)
- Higher coding efficiency and simpler filter bank
- 96 kbps AAC sounds better than 128 kbps MP3

MP3
Compression rate: 75 - 95% reduction in size compared to CD-quality audio (i.e. 2 channel signed 16-bit sampled at 44,100 Hz)

Quality depends on:
1. Quality of encoder algorithm
2. the complexity of signal

MP3 procedure
1. FFT
2. Psychoacoustic analyser
3. Filter bank
4. MDCT
5. Scaling and Quantising
6. Side information
7. Huffman encoding
8. Bit stream

1. FFT
Divide the audio signal in frames of 1152 samples, use the Fourier transform to transform the time domain data to frequency domain, sending the results to psychoacoustical analyser

2. Psychoacoustic analyser
The psychoacoustic analyser identifies masking tones and masked tones in a local neighborhood of frequencies over a small window of time
outputs: a set of signal-to-mask ratios (SMRs)

SMR: the ratio between the amplitude of a masking tone and the amplitude of the minimum masked frequency in the chosen vicinity.
The SMR at a given frequency = the difference between the masker and the masking threshold at that frequency

3. Filter bank
Divide each frame into 32 frequency bands between 0 and 22.5 kHz, using filter banks (bandpass filters)

4. MDCT
Use MDCT (modified DCT) to divide each of 32 frequency bands into 18 subbands for a total of 576 frequency subbands

5. Scaling and Quantising
Sort the subbands into 22 groups — scale factor bands
Use nonuniform quantisation, combined with scaling factors: bands that have lower SMR are multiplied by larger scaling factors because the quantisation error for these bands has less impact, falling below the masking threshold

An appropriate psychoacoustical analysis provides scaling factors that increase the quantisation error where it doesn't matter, in the presence of masking tones
Scale factor bands effectively allow less precision (i.e., fewer bits) to store values if the resulting **quantisation error** falls below the audible level

lower-SMR —> large masking threshold —> scale factor bands need less precision (fewer bits) (quantisation noise fall below the audible level | masking threshold) —> larger scaling factors

6. Side information
side information is the information needed to decode the rest of the data, including where the main data begins, where scale factors and Huffman encodings begin, the Huffman table to use, the quantisation step, and so forth

7. Huffman encoding
Use Huffman encoding on the resulting 576 quantised MDCT coefficients.

8. Bit stream
Put the encoded data into a properly formatted frame in the bit stream

MP3 bitrates

- MP3 was designed to encode data at 320 kbit/s or less
- A bit rate of 128 kbit/s is commonly used (11:1 compression rate)
- Variable Bit Rate (VBR): specify a given quality, adjust bit rate accordingly

ABR (Average Bit Rate)

A type of VBR where the bitrate is allowed to vary for more consistent quality, but is controlled to remain near an average value chosen by the user, for predictable file size

AAC

- More sampling rate, more channels, arbitrary bit rates
- Simpler filter banks — solely with MDCT, improved frequency resolution
- Frequencies over 16kHz are better preserved
- Better sound quality than MP3 for files compressed at the same bit rate (same file size)

Digital Broadcasting

What is broadcasting
Digital broadcasting system
Major standards for digital broadcasting
Digital Video Braodcasting - Satellite (DVB-S)

What is broadcasting

Broadcasting: a method of transferring a message to all recipients simultaneously

特点 (compared to uni-cast)

- Point-to-multipoint communication
- Simpler transmission scheme
- Higher transmission power

Why Digital broadcasting

- Digital signals are more robust
- Better quality
- More reliable
- More flexible
- Less expensive
- Time-vs-Frequency-domain multiplexing (transmit over single datalink)
- Additional devices requiring digital data
- Commercial reasons

Benefits of Digital Switchover
Potential benefits to consumers: 1. More services 2. Interative features & more info. Programmes 3. Easier tuning(调整) & new functions 4. Less interference
Potential benefits to company: 1. Less cost 2. Less spectrum 3. Diversifying
Potential benefits to government/regulatory body: 1. Wider coverage 2. Freeing up spectrum 3. Better management

Building blocks of a digital broadcasting system

Essential stages

- Channel codding: error protection of bits
- Modulation: transmitting signal onto carrier

Other stages

- Source coding: data compression
- Multiplexing: combining into single data stream
- Signal processing

Fundamental components of DB
Audio, Visual, Data

Transmitter
- Compression (source coding): e.g. MPEG-2
- Multiplexing: multiplexing information to single Transport Stream (TS)
- Channel coding (Forwarding error correction) e.g. Reed-solomon
- Modulation: e.g. OFDM (Orthogonal Frequency-Division Multiplexing)
- Transmission: e.g. antenna or optical fibre

Receiver (Reverse Process)

Transport Stream (TS)
Transport stream: specifies a container format encapsulating packetised Elementary Streams (ES)

Elementary Stream (ES)
Elementary Stream (ES): sequence of TS packets with same PID (packet identifier) value in header
- One set of elementary streams for global signalization
- One set of elementary streams per service

Channel Coding
Channel Coding (forwarding error control coding): a process of detecting and correcting bit errors in digital communication systems

Modulation
Digital Modulation: uses discrete signals to modulate a carrier wave
Three main types of digital modulation:
- Amplitude Shift Keying (ASK)
- Frequency Shift Keying (FSK)
- Phase Shift Keying (PSK)

Major standards for digital broadcasting
Why do we need standards?
Safety and reliability
Support of government policies and legislation
Interoperability (互用性) : The ability of devices to work together

Major standards for digital braodcasting
- Digital Video Broadcasting (DVB)
- Advanced Televison System Committee (ATSC)
- Integrated Services Digital Broadcasting (ISDB)
- Digital Terrestrial Multimedia Broadcasting (DTMB)

DVB
standards that defines digital broadcasting using existing satellite, cable, and terrestrial infrastructures
Focus of digital television development
Based on MPEG2 source coding
- DVB-S (1993), -C (1994), -T (1995), -SH, ...

DVB family of standards
Every DVB standard defines the channel coding and modulation
The system input and output signals are MPEG-2 Transport Streams

ATSC
depends on numerous interwoven standards
specification for HDTV (High Definition Television)
- Uses Dolby, not MPEG for audio

ISDB
digital television (DTV) and digital radio
Main differences compared to DVB (modulation):
– ISDB-S: 8-PSK and Trellis coding , ISDB-T: split into subchannels for adaptive use
– DVB-S: QPSK

DTMB
CMMB: Chinese Mobile Multimedia Broadcasting

Digital Video Broadcasting - Satellite (DVB-S)
DVB-S: standard for Direct-to-home Broadcasting via Satellite (DBS)

DVB-S Encoding
After the data has been coded following MPEG-2 standard, it needs to go to next several steps before transmitted to satellite:
- Multiplexing and randomisation for energy dispersion
- Reed-Solomon Encoder (Error Protection)
- Convolutional Interleaving
- QPSK modulation

Energy dispersal
Energy dispersal: at the encoding end, scrambling with a pseudo random sequence
目的: in order to achieve a power-density spectrum of the modulated signal that is as even as possible (调制信号的PDS—能量谱密度均匀分布)

Error Protection
Error Protection scheme permitting various code rate

Reed-Solomon coding
RS(204, 188, t=8) is used, where 16 parity bytes are introduced in each transport packet. With this the decoder is able to correct up to 8 error bytes in each packet of 204 received bytes.

Reed-Solomon coding principle
Ideal: The Reed-Solomon encoder takes a bloke of digital data and adds extra "redundant" bits.
工作原理: encoder takes k data (symbols of s bits) add parity symbol to make an n symbol codeword. A Reed-Solomon decoder can correct up to t symbols that contain errors in a codeword, where $2t = n-k$.

Interleaving
目的: in order to avoid errors in consecutive packets

Modulation (QPSK — Quadrature Phase Shift Keying)
With four phases, QPSK can encode two bits per symbol

Data Rate Calculation
1. QPSK offers 2 bits/Symbol, QPSK-modulated signal must first be provided — gross_data_rate
2. Error protection: Reed-Solomon code with rate (204, 188) — net_data_rate_Reed-Solomon
3. Further error protection: Code rate = $\frac{Input\ data\ rate}{Output\ data\ rate}$, in DVB-S, code rate can be selected with the range of : 1/2, 3/4, 2/3, …. 7/8 — net_data_rate

Observation on data rate
Code rate = 1/2: error protection — maximum, net data rate — minimum
Code rate = 7/8: error protection — minimum, net data rate — maximum
The code rate can then be used to control the error protection and thus, as a reciprocal of this, also the net data rate

DVB-S2
- Improved version of DVB-S standard
- Broadcast Services for standard definition TV and HDTV
- Improvements in channel coding
- Improvements in channel modulation
- offers 30% data rate increase under the same condition compared to DVB-S