Digital Audio

- Acoustic energy flowing outwards from its generation point
- Characteristics: Amplitude, Frequency, Waveform, Speed
- Pyscho-acoustic: the branch of acoustic to deal with human auditory perceptions
- Amplitude masking occurs because an audible sound has tendency to distort the threshold curve and shift upward
- The amount of distortion of the curve is restricted to a small region surrounding the strongest sounds.The entire range of audible frequencies is divided into a number of such regions known as critical bands.
- Temporal masking: after the ear hears the sound , it takes some time for the ears to hear another quitter sound with a frequency within the critical bandwidth of the louder tone which takes some time to decay.
- During the masking period , signals with amp below the masking could not be heard, the masking period may last more then 50ms
- Digital representation of audio signals
- Advantages: Easy : manipulation of data, combination with other data, better protection against transmission errors, enable encryption of secure data, enable compression for efficient storage and transmission.
- Disadvantages: simple data representation results in high data rates e.g. PCM
- dB = 10Log(amp1^2/amp2^2)
- SNR: signal to noise ratio
- SQNR: 6.02N dB
- PCM: it is simply the use of sampling and quantisation to create digital signals from analog Signals
- Differential PCM code difference between the actual and predicted value of the sample
- Delta modulation: special case of DPCM , here difference between predicted and current value is coded with a single bit.
- DM can cause 2 type of distortion , 1. Slope overload distortion when signals changes too rapidly; 2. Granular distortion when signals changes too slowly
- Adaptive DCPM: due to avoid above distortions , in this method number of bits is varied depending on the amplitude of the differential signal
- Zero crossing rate: the rate at which signal changes from +ve to -ve or back.
- Short time energy: sum of energy of all samples in a frame
- Sub-band energy: after FFT, frequency domain singal of a frame is divided into sub bands. The sub band energy can better depict the energy distribution over spectrum. Sub band energy ratio is the relative energy value normalised by the total energy of the frame.
- MFC(Mel-frequency cepstrum) is a representation of short-term power spectrum of a sound. Based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale frequency. MFCCs are coefficients that collectively make up an MFC.
- In Mel frequency bands are equally spaced on the Mel scale , which approximates the human auditory system's response more closely than the linearly spaced frequency bands used in normal cep strum. It allows for. Better representation of sound
- A mel is a unit of measure for the perceived pitch of a tone

Digital Image

- Sampled quantities: continuous values at discrete space or time
- Quantized quantities: discrete values at continous space or time
- Digital quantities: discrete values at discrete space or time
- Non uniform sampling: divide image into blocks and perform different sampling on each block to prevent checkerboard and false contours , fine sampling in region of sharp grey level transitions and coarse sampling in relatively smooth transition
- Color compression: CLUT, indexing b bits to 256 colors
- Median Cut algo; values smaller than median are labeled 0 and greater are 1
- FT: it involved the transformation of a signal into a form which is some sense in more desirable for analysis, it is unique and invertible
- F(u) = integrate[f(x)e^(-j2(Pi)ux)]-inf to +inf
- FFT has nLogn operations
- DFT : F(u, v) = 1/MN Sum(f(x, y)e^(-j2(Pi)(ux/M + vy/N)))

- Properties of FT: invertibility, separablity(means that F(u,v) can be computed be sucsessive application of 1D transform of its inverse), Rotation, Scalablity(f(ax, by) = F(u/a, v/b)/|ab|), Convolution(f(x)*g(x) <-> F(u)G(u))
- DCT: Discrete cosine transform has been very popular in block transformation based image compression for a long time, JPEG/MPEG
- The compression is done in 2 stages: 1 . During quantisation and second during the entopy coding process on generated 64 coefficients by DCT on blocks
- DCT vs DFT
- Similarities: both decompose a finite length discrete time vector into a sum of scaled-and-shifted basis function
- Differences : DFT uses a set of harmonically related complex exponential functions, DCT uses only real valued cosine functions
- DCT representation tend to have more of its energy concentrated in a small number of coefficients when compared to other transforms like DFT
- DCT is used in lossy data compression like JPEG, DFT is used for general spectral analysis application

Digital Image Indexing

- Indexing: Inverted indexing, Tree indexing(effective for low dimensions , 10), Hashing,
- Hashing: hash high dimensional data into low dimensional hamming-space based on a family of has functions, hashes similar samples into same bucket
- Hamming space is the set of all 2^L binary strings on length L
- Hamming distance is the number of changed bits in two equal length binary strings.
- Good hash code
- Compact: required a small number of bits to code the full data
- Effective: maps similar samples to similar binary code words
- Efficient: easily computed for a new image
- Locality Sensitive Hashing (LSH): based on Random projection
- Focuses on approximate nearest neighbour search by hasing similar points together as much as possible using Random projection
- Each data point is mapped to lower dimensional b bit vector called hash key, each function must satisfy LSH property
- Limitations: 1. Need long codes for higher accuracy, need many hash tables for good recall, size of L and M are heuristic
- KD tree: index global dense features like color histo, texture hits
- Inverted: bag of visual words feature
- Hashing index: for global and any combined features
- Google architecture: Main indexing, additional indexing features: link information in form of Page Rank, anchor text, location information, formatting info

Image Content Analysis

- Numeric representation of a 2D image could be vector or rastered image
- Raster images: composed of pixels , contains array of pixels of various colors; JPEG, PNG, GIF, BMP
- Vector images: composed of path/lines, contain points where the path start and end how much the paths curve and the color that either border of fill the paths, CGM, SVG, PDF etc
- Vector graphics uses 2D point locked polygons to represent images, each point has definite position (direction of path) on the work plane, each path maybe assigned various attributes
- YUV, Y for intensity and U and V channels provide color information , it is a affine transformation of RGB color space
- Properties of histogram
- Desirable properties of feature vector f(I)
- |f(I) - f(I')| should be large iff I and I' are very different
- f(.) should have proper of monotonicity, should be fast to compute, f(I) should be small in dimension
- But has no spatial info and not robust to large apperenace changes
- Linear color spaces, Luv, Lab, YUV, YCrCb
- For spatial info in consideration when utilizing colors: Segmented Image representation(more expensive , may have mis-alignment problem), Color cohenrence vectors
- Tamura representation: classifies textures based on : coarseness, contrast, linelikeness, directionality, roughness, regularities
- SIFT properties: invariant to uniform scaling, rotation and partially to changes in illumination, minor changes in view direction, occlusion, clutter

- Harris corner detection: it is rotation invariant but not scale, looking for dominant gradient directions or Eigen vectors
- Laplacian of Gaussian for Scale invariant blob detection
- SIFT use edge oriented representation, take square window around detected features, compare edge orientation, throw out weak edges, , create histogram of rest

NLP

- Normalization: series of related tasks meant to pull all text on a level playing field, converting to same case , removing punctuation, expanding contraction , converting numbers to their word equivalent
- Stemming: process of emilnating affixes from a word in order to obtain a word stem
- Lemmatization: converting a word into its canonical word
- TF helps in recall and idf in precision
- Vector Space models
- Advantages: simplicity, able to handle weighted terms, easy to modify term vectors
- Disadvantages: term independent assumption, first order , synonym and polysemy problems
- Query space RF: relevant documents resemble each other , query is re formulated based on known relevant documents
- Re formulation is based on 2 complementary operation
- Terms occur in relevant document are added or weight are increased of existed
- Terms occur in non relevant document are deleted or weight decreased
- Pseudo RF: it automated manual part of true RF
- Retrieve a ranked list of hits for the user's query
- Assume that the top k documents are relevant
- Do RF
- But can go horribly wrong , problem of query shift can occur
- problem of RF: It benefits only current query session, does not benefit from wisdom of other users
- Document Space RF
- Modify content of relevant and/or irrelevant documents to make documents more easily retrievable in future by other users
- Popular in Image R to add text annotation in images
- DRF:
- Query terms are added in relevant documents with initial weight lambda , or weight of such terms is increased by alpha
- Query term appear in non-rel docs are decremented bt weight sigma and suck terms are removed if its weight is less than nil
- Where lambda, alpha, sigma, are constant
- Problems in DRF
- Document may lose original meaning
- Cannot support personslised search
- Cannot support changing user needs
- Precision = a /(a+b)
- Recall = a / (a+c)
-