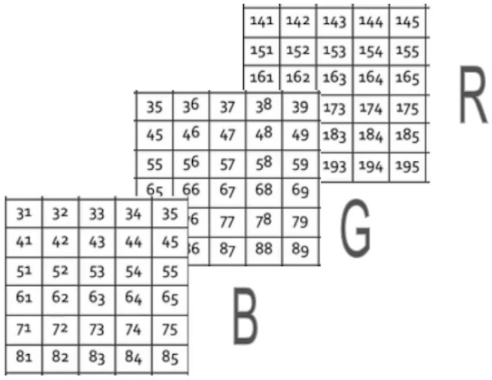
Foundations of Data Science

Report

- V Sushant (2019A7PS0045P)
Shruti Gangwar (2019B2A10920P)

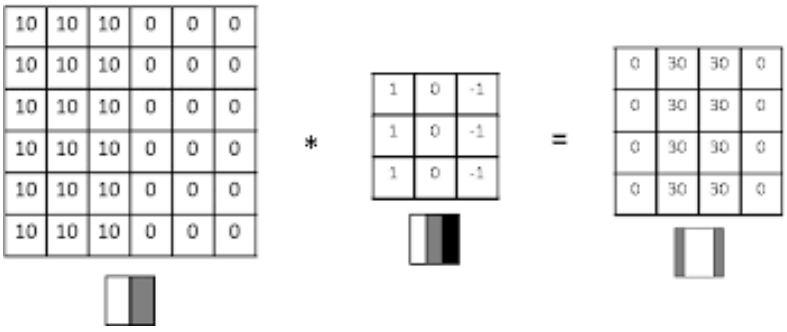
1.Image Data

Images are made up of many different colours, but almost all of them can be made up of the three main colours: red, green, and blue. We can say that each colored image is composed of these three colors or 3 channels- Red, Green, and Blue.
This means that in a colored image the number of matrices or the number of channels will be more. In this particular example, we have 3 matrices- 1 matrix for red known as Red channel, 1 for blue and 1 channel for green.



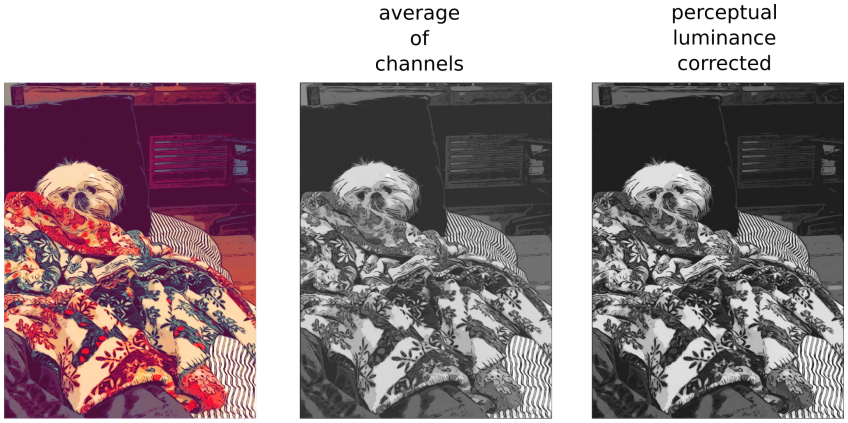
Edge Detection in Images

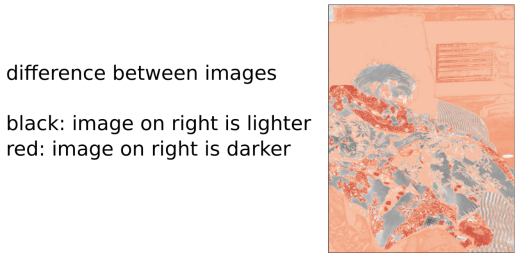
Convolution is simply the process of multiplying and adding corresponding indices of two matrices together (the kernel and the image matrix) after flipping both the rows and the columns of the kernel. Edges represent the object boundaries. As a result, edge detection is a crucial stage in any object detection or recognition process. Simple edge detection kernels are based on gradient image approximation.



RGB to GrayScale Conversion

An intuitive way to convert a color image 3D array to a grayscale 2D array is, for each pixel, take the average of the red, green, and blue pixel values to get the grayscale value. This combines the lightness or luminance contributed by each color band into a reasonable gray approximation.



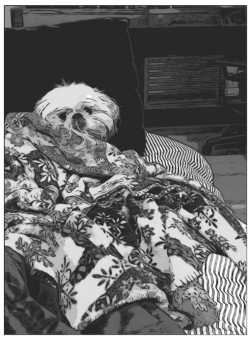


The gamma decompression and re-compression rack up quite a large computation cost, compared to the weighted averages we were working with before. Sometimes speed is more desirable than accurate-as-possible luminance calculations. For situations like these, there is a linear approximation:

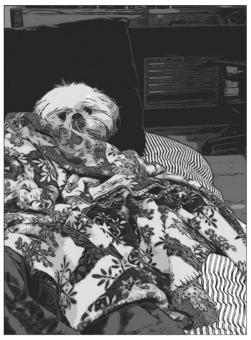
$$Y' = 0.299R' + 0.587G' + 0.114B'$$

This lets you get a result that's a little closer to the gamma-compression-corrected version, but without the extra computation time.

gamma and perceptual luminance corrected

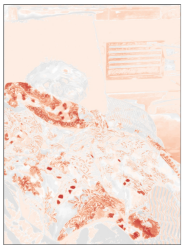


linear approximation of gamma and perceptual luminance corrected



difference between images

black: image on right is lighter
red: image on right is darker



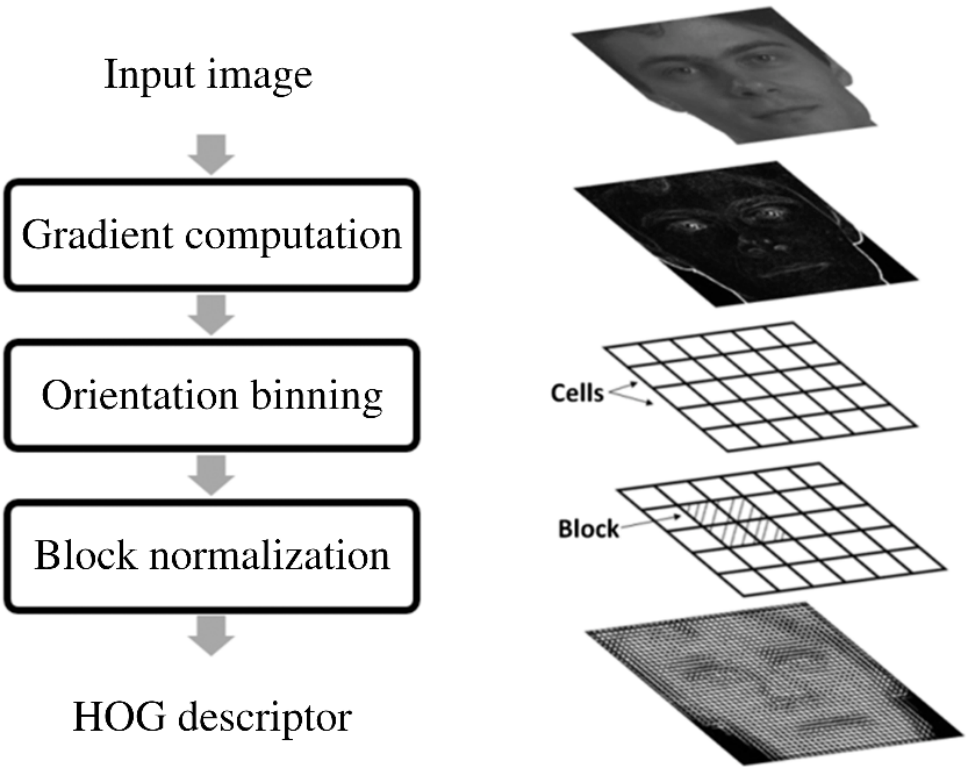
Histogram of Oriented Gradients (HOGs)

A feature descriptor is a representation of an image or an image patch that simplifies the image by extracting useful information and throwing away extraneous information. Typically, a feature descriptor converts an image of size width x height x 3 (channels) to a feature vector / array of length n. In the case of the HOG feature descriptor, the input image is of size 64 x 128 x 3 and the output feature vector is of length 3780.

Input image

Histogram of Oriented Gradients

A Histogram of Oriented Gradients (HOG) of image can be computed by Global image normalisation (optional), followed by computing the gradient image in x and y, then computing gradient histograms, then normalising across blocks and flattening into a feature vector.



2. Video

Frame-by-frame extraction of Videos

A video is a sequence of images (called frames) captured and eventually displayed at a given frequency. However, by stopping at a specific frame of the sequence, a single video frame, i.e. an image, is obtained.
For each video frame k = 1 to N,
Read frame Vk and Vk+1
Obtain the gray level image for Vk and Vk+1
Gk = Gray image of Vk
Gk +1 = Gray image of Vk +1
Find the edge difference between Gk and Gk + 1 using the Canny edge detector.
Let diff(k) be their difference.

dif(k) = summation of i,j (Gk - Gk + 1) where i, j are row and column index.
Compute the mean and standard deviation:
Compute the threshold value
Threshold = M + a*S , where a = constant
Find the key frames for k = 1 to (N-1)
If diff(k) > Threshold then,
Write frame Vk +1 as the output key-frame.

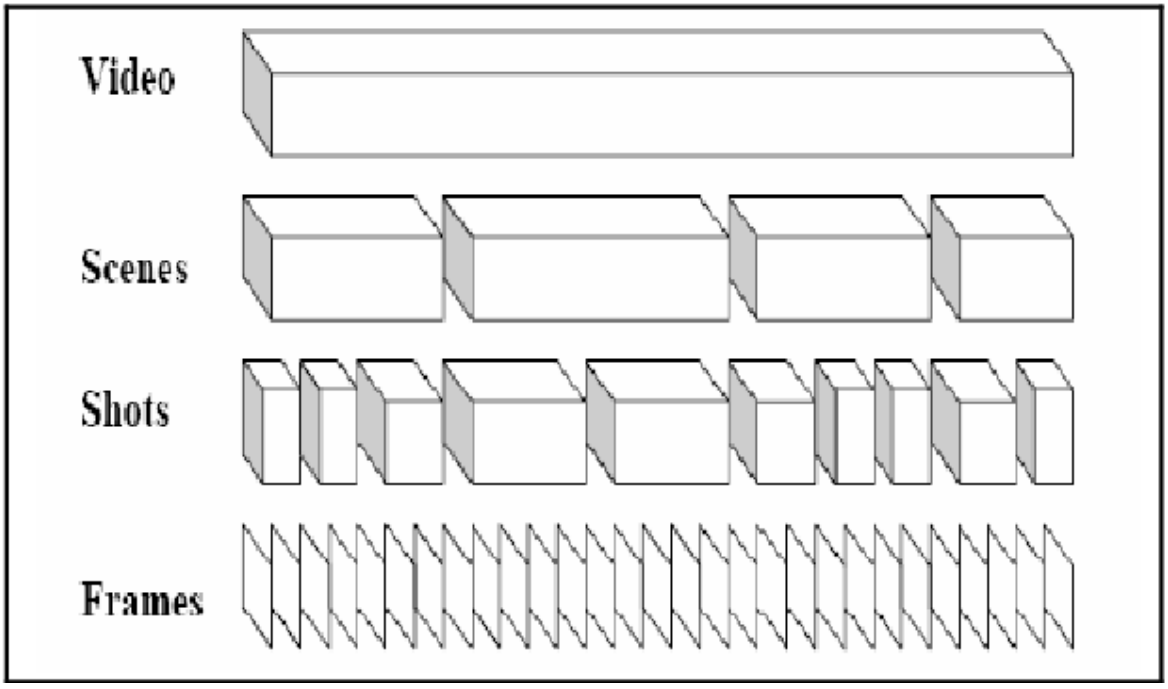


Image Feature Extraction in Video Frames

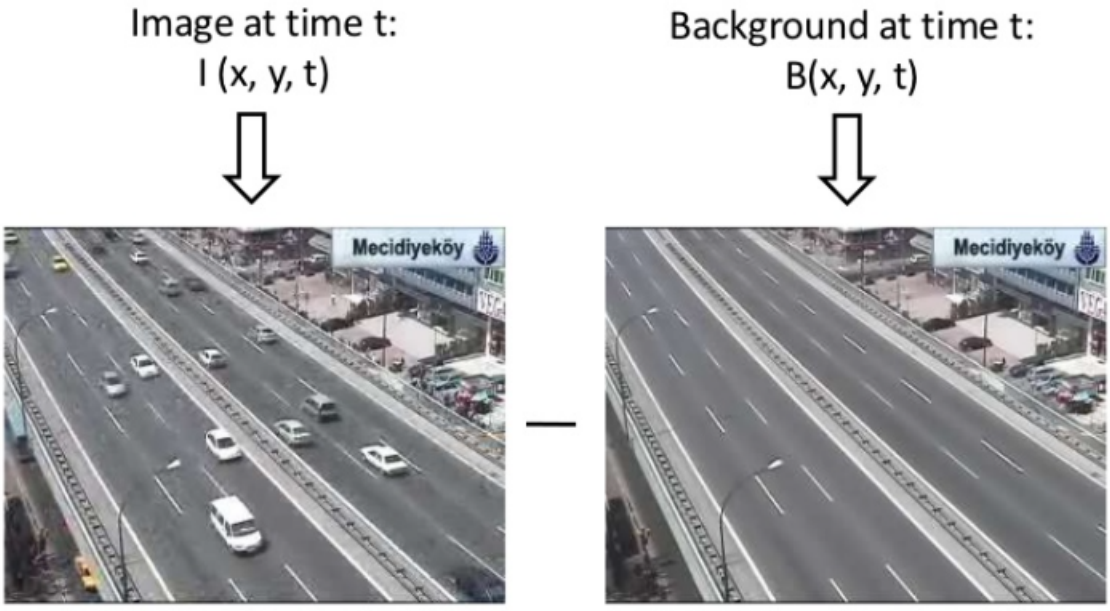
Since the frames are analogous to images, techniques like edge detection, gray scale conversion, changing illumination, RGBtoBGR, etc can be done in video frames.

Background Subtraction

Background subtraction is the process of separating the background and foreground from a sequence of image/video frames. It is generally used for detecting or removing moving objects from the videos of static cameras.
The concept of running average is to detect active objects and remove them i.e. differentiate between pixels that seem to change over time and remove them. Here

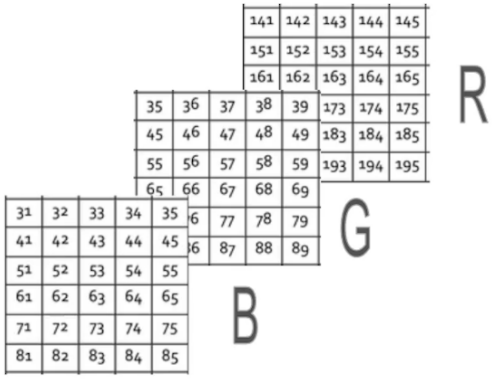
“Running” signifies the fact that the average is being computed over previous and current frames again and again until the frames are exhausted. Simply, the background is the mean of ‘n’ previous frames in the video.

$$B(x, y, t) = \frac{1}{n} \sum_{i=0}^{n-1} I(x, y, t - i)$$



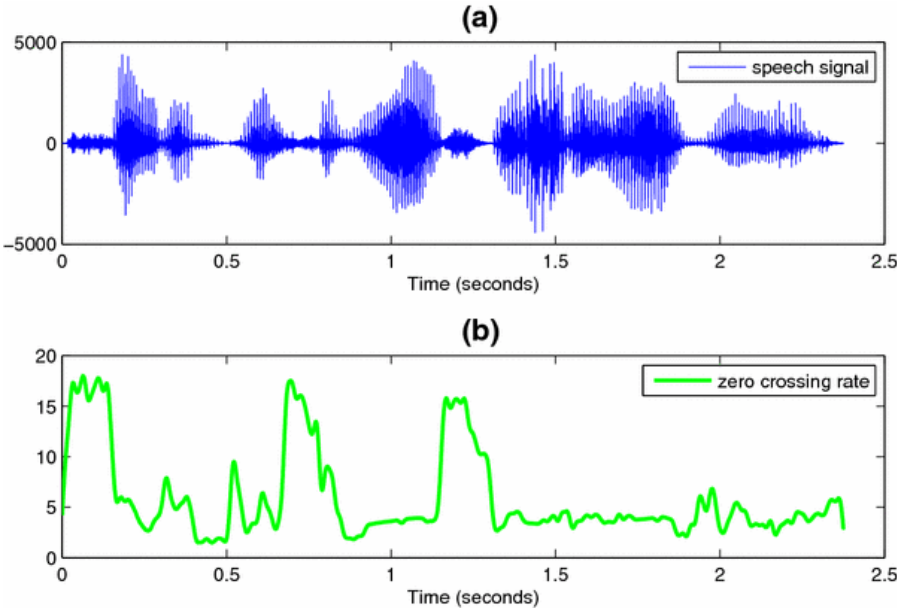
3. Audio

Images are made up of many different colours, but almost all of them can be made up of the three main colours: red, green, and blue. We can say that each colored image is composed of these three colors or 3 channels- Red, Green, and Blue.
This means that in a colored image the number of matrices or the number of channels will be more. In this particular example, we have 3 matrices- 1 matrix for red known as Red channel, 1 for blue and 1 channel for green.



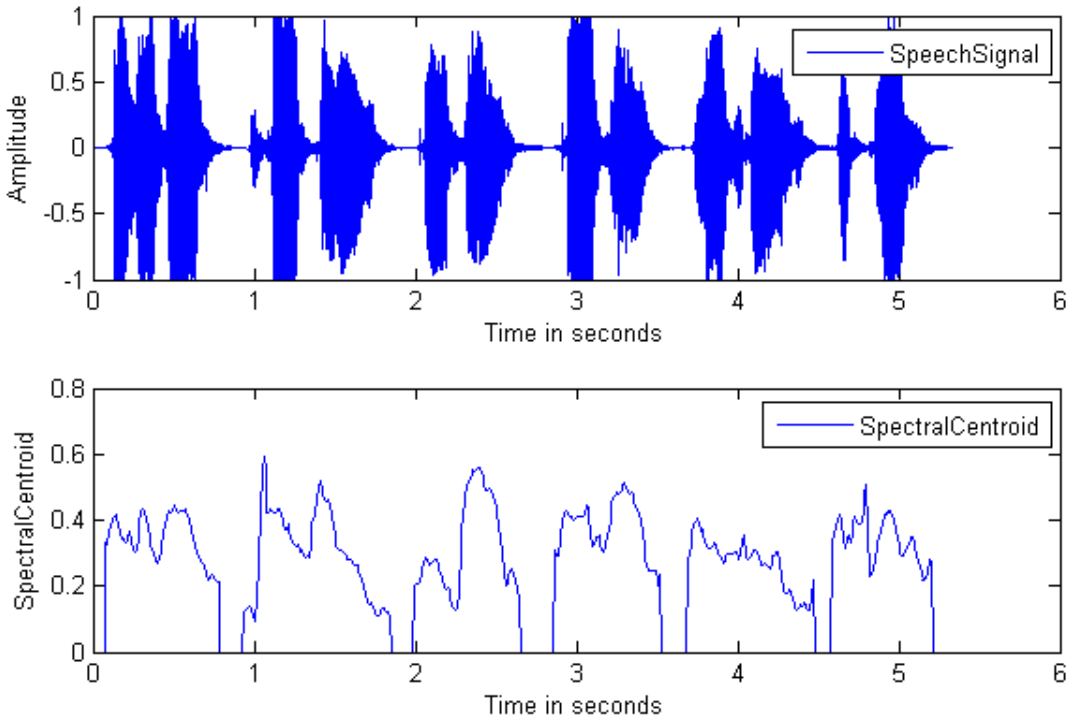
Zero Crossing Rate

The zero-crossing rate (ZCR) is the rate at which a signal changes from positive to zero to negative or from negative to zero to positive. Its value has been widely used in both speech recognition and music information retrieval, being a key feature to classify percussive sounds.



Spectral Centroid

Spectral centroid (SC) measures the shape of the spectrum of EEG signals. A higher value of SC corresponds to more energy of the signal being concentrated within higher frequencies. Basically, it measures the spectral shape and position of the spectrum.



It is computed as follows:

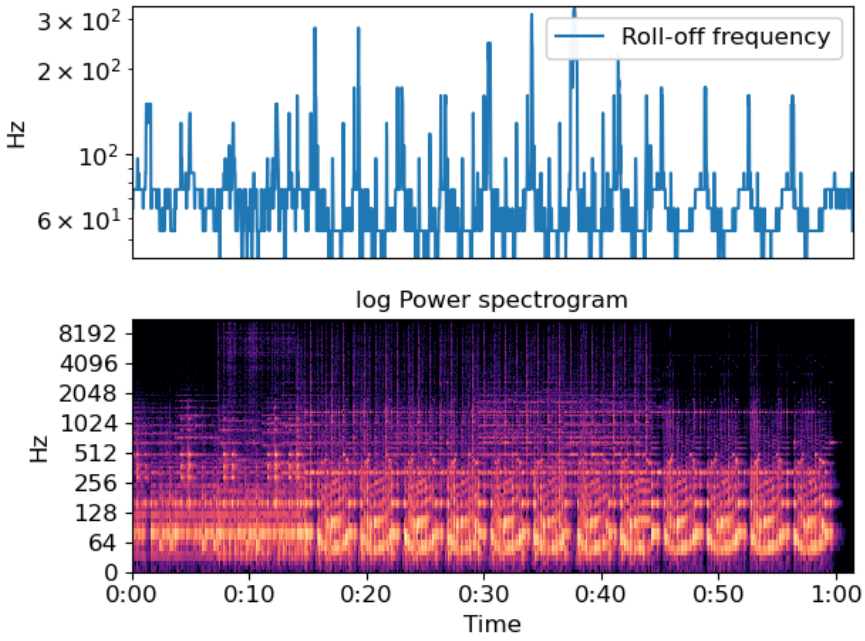
1. Let xi(n), n = 0,1,…N–1 be the sample of the ith frame, with Xi(k), k = 0,1,...N–1 as the discrete Fourier transform (DFT) coefficients of the sequence.
2. Compute the SC of the each frame as:

$$C(i) = \frac{\sum_{k=0}^{N-1} k|Xi(k)|}{\sum_{k=0}^{N-1} |Xi(k)|}$$

3. The mean value of the SC across all the frames can be used as the SC feature for each epoch of the frame of the EEG signal.

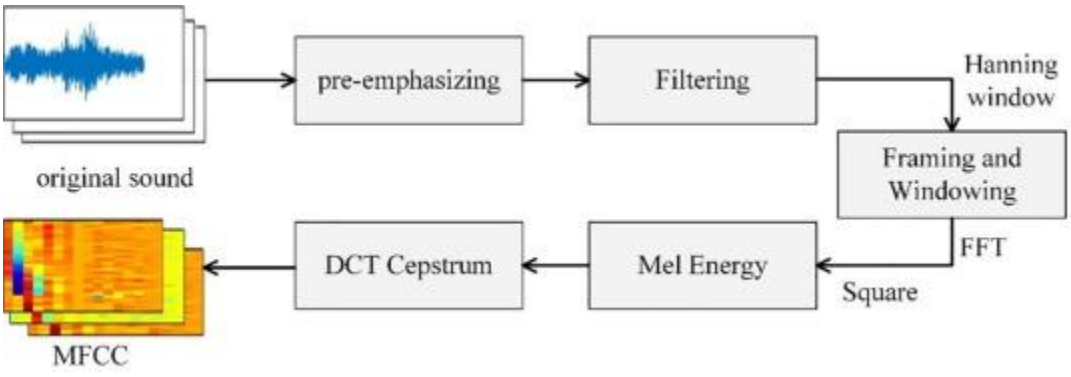
Spectral Rolloff

Spectral rolloff is the frequency below which a specified percentage of the total spectral energy, e.g. 85%, lies. It also gives results for each frame. . spectral_rolloff is used to calculate rolloff for a given frame.



MFCC

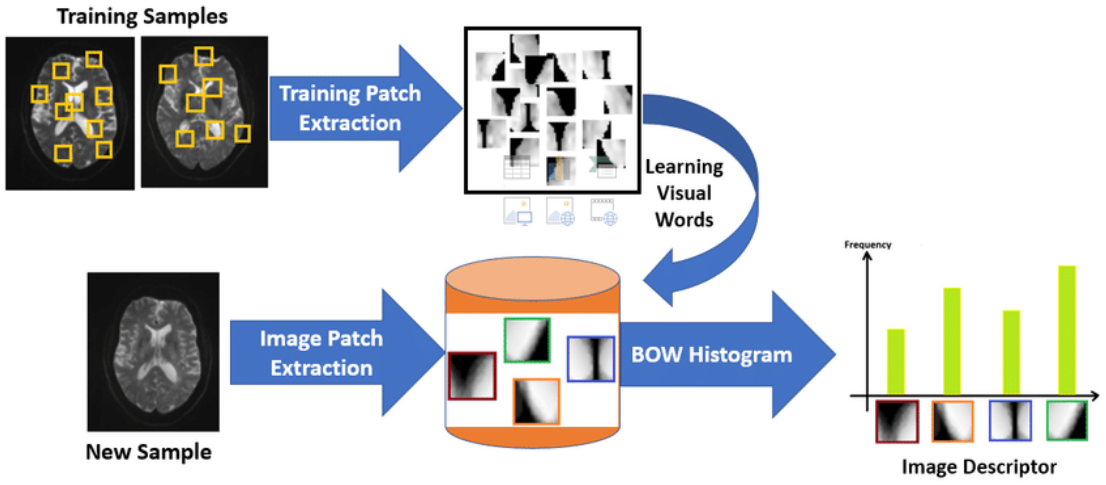
The Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method is a leading approach for speech feature extraction and current research aims to identify performance enhancements. One of the recent MFCC implementations is the Delta-Delta MFCC, which improves speaker verification.



4. Text

Bag of Words

The bag-of-words model is a simplifying representation used in natural language processing and information retrieval. In this model, a text is represented as the bag of its words, disregarding grammar and even word order but keeping multiplicity. The bag-of-words model has also been used for computer vision.



Term Frequency – Inverse Document Frequency (TF – IDF)

It is the most popular method to perform feature extraction. To understand better let's understand TF and IDF separately.

Term Frequency: Simply finds out the frequency of a word in document.

Inverse Document Frequency: Assigns a lower weight to the words which appear most frequently. It basically depicts the rarity of the word in all documents.

$$TFIDF\ score\ for\ term\ i\ in\ document\ j = TF(i,j) * IDF(i)$$

where

$IDF = Inverse\ Document\ Frequency$

$TF = Term\ Frequency$

$$TF(i,j) = \frac{\text{Term i frequency in document j}}{\text{Total words in document j}}$$

$$IDF(i) = \log_2 \left(\frac{\text{Total documents}}{\text{documents with term i}} \right)$$

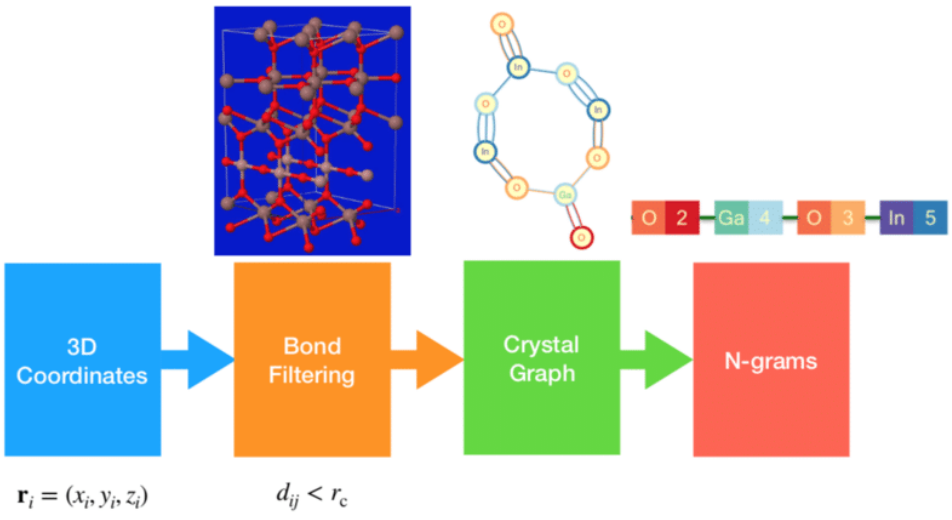
and

$t = Term$

$j = Document$

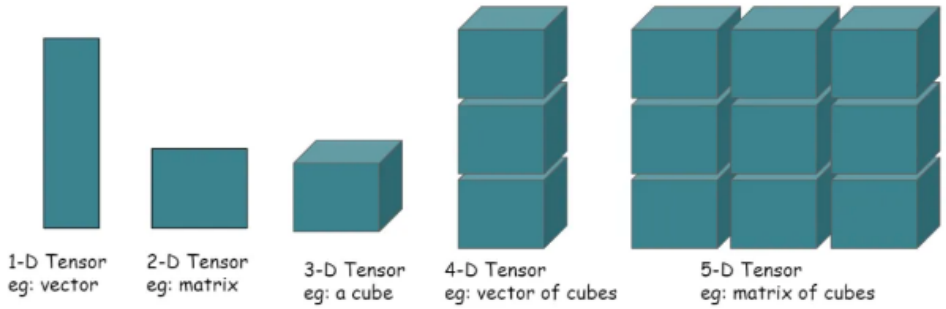
N gram

N-grams are the basic features commonly used in sequence-based malicious code detection methods in computer virology research. The empirical results from previous works suggest that, while short length n-grams are easier to extract, the characteristics of the underlying executables are better represented in lengthier n-grams. However, by increasing the length of an n-gram, the feature space grows in an exponential manner and much space and computational resources are demanded.



5. Tensor Representation of Data

Tensors are the underlying data structure used by all modern machine-learning structures. Tensors are fundamental to the field — so fundamental that Google's TensorFlow was named after them. Even the text data or image data are converted to Numerical features for processing.
At its core, a tensor is a container for data — almost always numerical data. So, it's a container for numbers. You may be already familiar with matrices, which are 2D tensors: tensors are a generalization of matrices to an arbitrary number of dimensions (note that in the context of tensors, a dimension is often called an axis).



Tensors with different dimensions

Real-world examples of data tensors

Vector data

This is the most common case. In such a dataset, each single data point can be encoded as a vector, and thus a batch of data will be encoded as a 2D tensor (that is, an array of vectors), where the first axis is the samples axis and the second axis is the features axis

Time Series data or sequence data

Whenever time matters in your data (or the notion of sequence order), it makes sense to store it in a 3D tensor with an explicit time axis. Each sample can be encoded as a sequence of vectors (a 2D tensor), and thus a batch of data will be encoded as a 3D tensor.

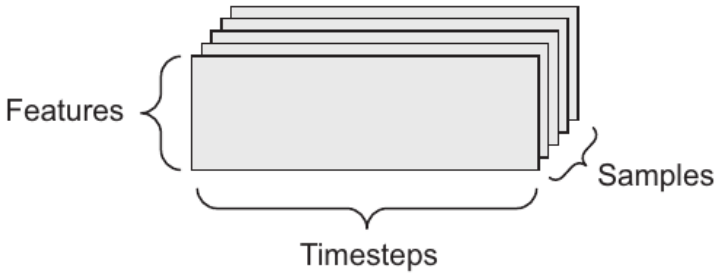
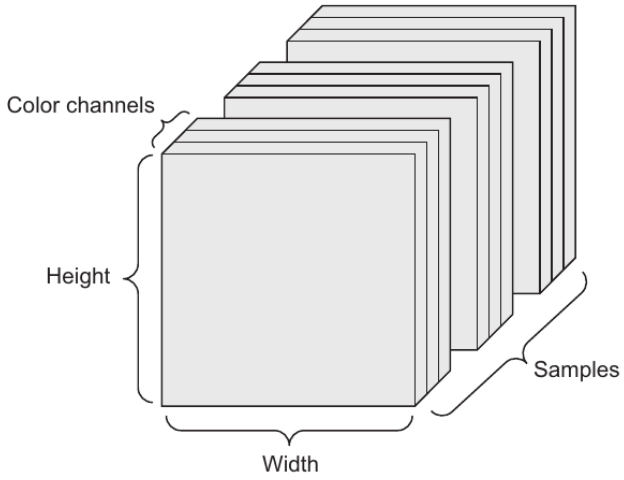


Image data

By convention image tensors are always 3D, with a one-dimensional color channel for grayscale images. A batch of 128 grayscale images of size 256 × 256 could thus be stored in a tensor of shape (128, 256, 256, 1), andbatch of 128 color images could be stored in a tensor of shape (128, 256, 256, 3). Therefore, 4D image data tensors exist.



Video data

A video can be understood as a sequence of frames, each frame being a color image. Because each frame can be stored in a 3D tensor (height, width, color_depth), a sequence of frames can be stored in a 4D tensor (frames, height, width, color_depth), and thus a batch of different videos can be stored in a 5D tensor of shape (samples, frames, height, width, color_depth).