多媒體程式設計
簡介

Instructor: 馬豪尚

認識多媒體資料

- 文字
  - 字數、長度、詞性、語意...
- 影像
  - 相片（picture）、圖像（image）、影片（video）、影片擷取出的畫面（frame）
  - 大小、解析度、色彩度...
- 音訊
  - 聲音(audio)、語音(speech)
  - 大小、頻率...

文字資訊處理

- 全文檢索
  - 將全部的文字訊息儲存起來
  - 使用者必須詳細的規劃自己的查詢
- 關鍵字查詢
  - 字詞切割
  - 關鍵字定義與比對
- 相似度比對
  - 向量化表示
  - 機率模型
- 自然語言處理

自然語言處理

- 詞嵌入向量 (Word2vec)
- 語法分析/剖析（Syntactic analysis/Parsing）
- 詞性標註（Part-of-speech tagging）
- 語意分析（Semantic analysis）
- 文字情感分析（Sentiment analysis）

什麼是一個字?

- 英文:
  - Lemma: cat = cats
  - Wordform: cat != cats
- 中文:
  - 葡萄、蜻蜓、蚯蚓、車門

什麼是一個字?

- 字符(character):
  - 獨立的字符就是字符表裡能找到的字
  - 例如:蜻、蜓
- 字(word):
  - 句子裡具有獨立意義的單位
  - 例如:蜻蜓、葡萄
- 詞(phrase):
  - 字+詞綴(構詞/句法)
  - 例如:很高興、很討厭
- 符記(token):
  - 文字經過切分之後的單位結果
  - 例如:我的\興趣\是看\電影\和讀\小說

文字資料前處理

- 斷句
  - 去除標點符號
- 斷詞/分詞
  - 找出有意義的詞彙
- 去除語意意涵較低的詞彙

斷詞/分詞

- 「斷詞」，指的是能夠讓電腦把詞彙以「意義」為單位切割出來
- 例如以下句子:「我的興趣是看電影和讀書」
  - 「我的興\趣是看電\影和讀\書」像這樣分組是不符合現實世界的意義
  - 「我\的\興趣\是\看\電影\和\讀書」透過斷詞技術就可以取得這樣的詞彙。

斷詞/分詞

- **基於設定好的單位切分**：將整個字符串以固定單位來切分，例如N-Gram
- **基於詞典的分詞法**：將待匹配的字符串和一個已建立好的詞典中的詞進行匹配，通常會採用雙向匹配的方法，但這方法的能力有限，例如像是新發明的詞就無法進行匹配
- **統計的機器學習算法**：如HMM，CRF (Conditional Random Field)，常見中文斷詞Jieba套件，對於不存在於字典的字詞就是用統計的方法來處理的
- **深度學習的算法**：例如使用LSTM模型，深度學習的方法應該算是比較新的

N-Gram 斷詞法

- **我的興趣是看電影和讀書**
- Uni-Gram
  - 「我\的\興\趣\是\看\電\影\和\讀\書」
- Bi-Gram
  - 「我的\的興\興趣\趣是\是看\看電\電影\影和\和讀\讀書」

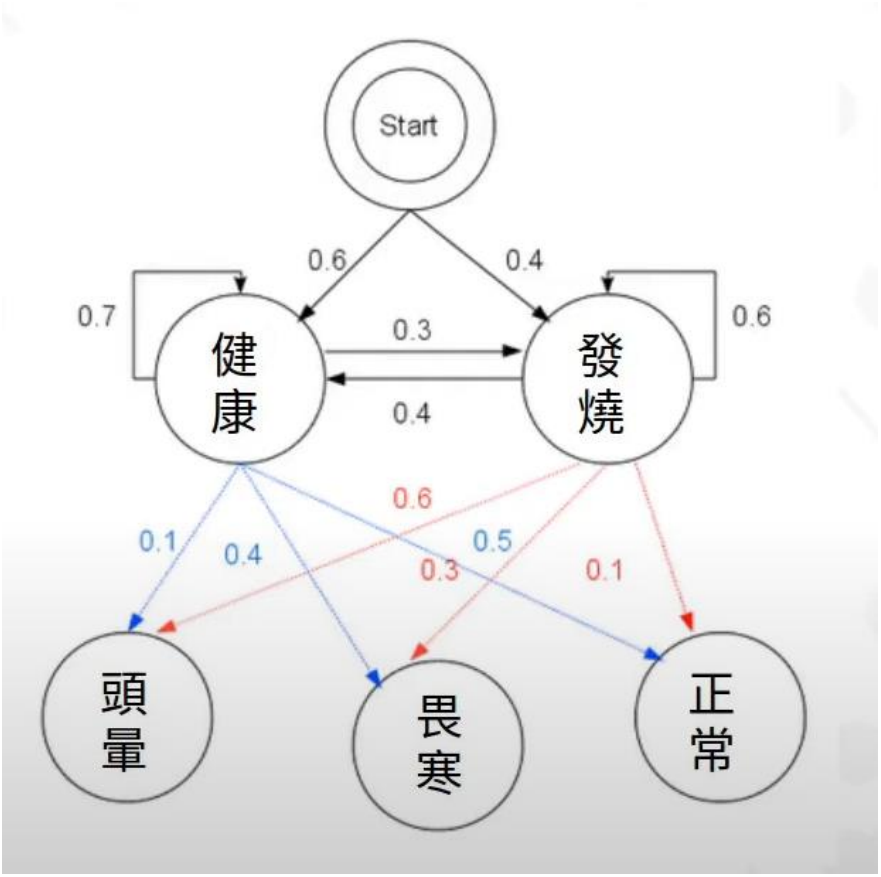
基於詞典的分詞法

- **我的興趣是看電影和讀書**
  - 「我\的\興趣\是\看電影\和\讀\書」

詞典
我
興趣
電影
讀
看
書
看電影

統計的機器學習算法

- HMM(隱馬可夫模型Hidden Markov Model)
  - 運用大量文本去統計"詞與詞"之間的關聯性來模擬機率



去除語意意涵較低的詞彙

- Stop Word (停用詞)大致分為兩類
  - 人類語言中包含的功能詞，這些功能詞極其普遍，與其他詞相比，功能詞沒有什麼實際含義，比如'the'、'is'、'at'、'which'、'on'等。
  - 詞彙詞，比如'want'等，這些詞應用十分廣泛，但是對這樣的詞搜索引擎無法保證能夠給出真正相關的搜索結果，難以幫助縮小搜索範圍。

TF-IDF（Term Frequency - Inverse Document Frequency）

- TF-IDF 是一種用於文字資訊檢索與探勘的常用加權技術，為一種統計方法，用來評估單詞對於文件的集合或詞庫中一份文件的重要程度
- TF（Term Frequency）
  - 這個單詞出現在該文件的次數/該文件的總字數
- IDF(Inverse Document Frequency)
  - 總文件數/這個單詞有出現的文件數

TF-IDF

TF=5/24=0.208，IDF=log(3/1)=0.477

TF=7/24=0.292，IDF=log(3/1)=0.477

TF=19/24=0.792，IDF=log(3/3)=0

TF=7/24=0.292，IDF=log(3/1)=0.477

詞嵌入向量

- One-hot Encoding(獨熱編碼)
  - 為了改良數字大小沒有意義的問題，將不同的類別分別獨立為一欄
  - 缺點是需要較大的記憶空間與計算時間，且類別數量越多時越嚴重

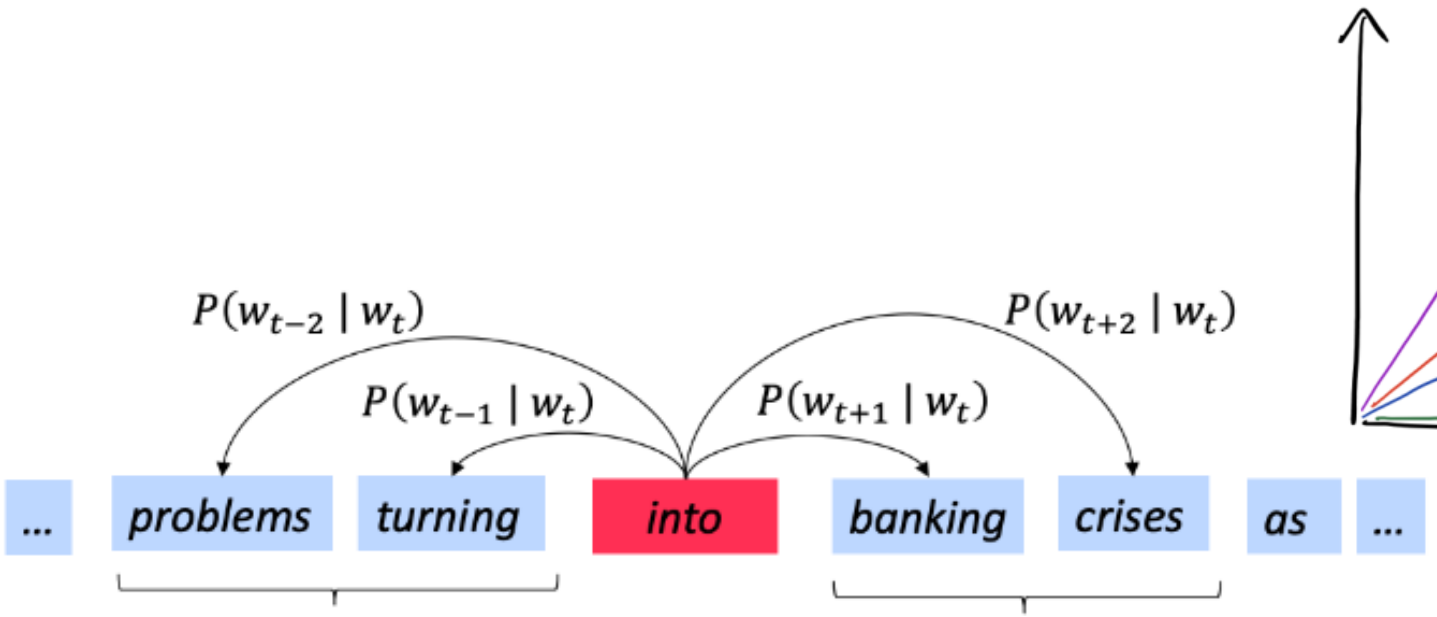
- I =>[1,0,0,0,0]
- like =>[0,1,0,0,0]
- apple =>[0,0,1,0,0]
- and =>[0,0,0,1,0]
- Mango=>[0,0,0,0,1]
- I like apple
  - [[1,0,0,0,0],[0,1,0,0,0],[0,0,1,0,0]]

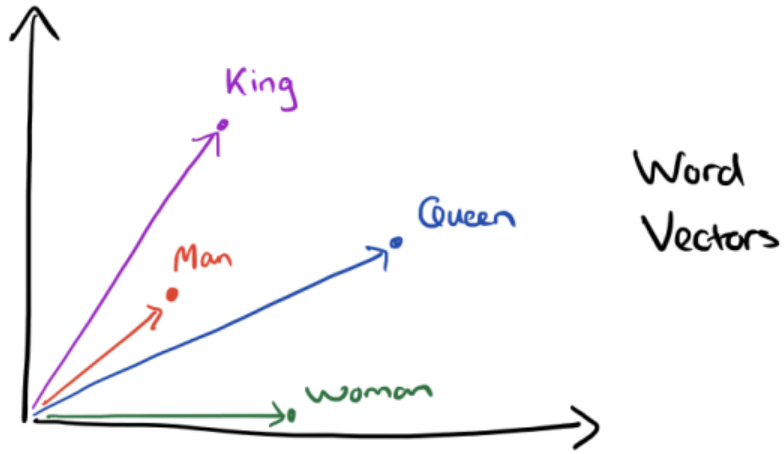
詞典

I
like
apple
and
mongo

詞嵌入向量

- Word2Vec
  - 將文字資訊嵌入在更短的向量中
  - 計算出向量空間上的相似度(Cosine)，來表示文本語義上的相似度
  - 用文章裡的每個字當 input，預測他周圍的字





認識影像

- 相片（picture）
- 圖像（image）
- 影片（video）
- 影片擷取出的畫面（frame）

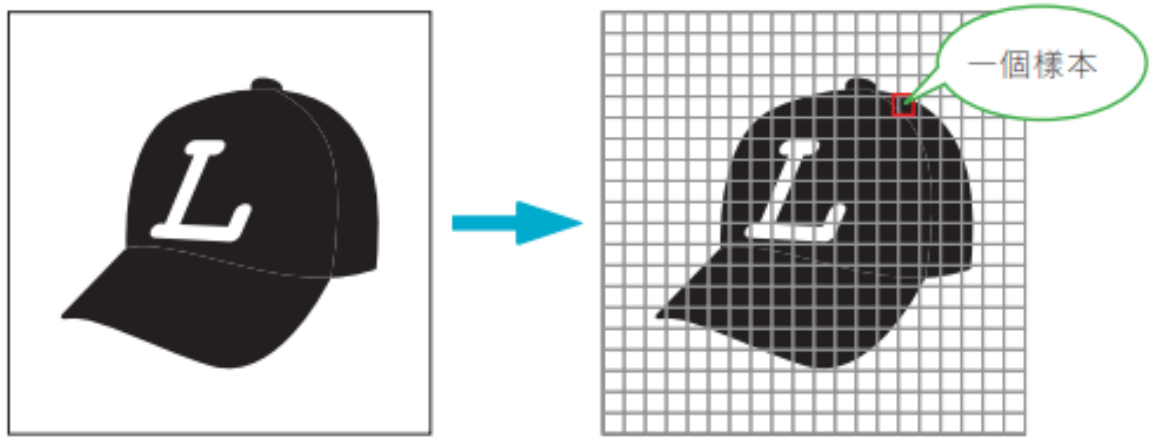
到底有什麼不同?

相片

- 相片(picture)：連續色彩變化的圖畫(也許有顆粒，但裸眼看不出)，因此從數學的觀點來描述，相片是一個連續二維空間的亮度函數

圖像

› 圖像(image)：將相片分割成一個個整齊排列的顆粒，再給予每一顆粒一個數值表示該顆粒的亮度；這樣的空間分割及亮度數值指定合稱為數位化(digitalization) 或離散化(discretization)。數位化後的相片就稱為圖像



原黑白影像圖

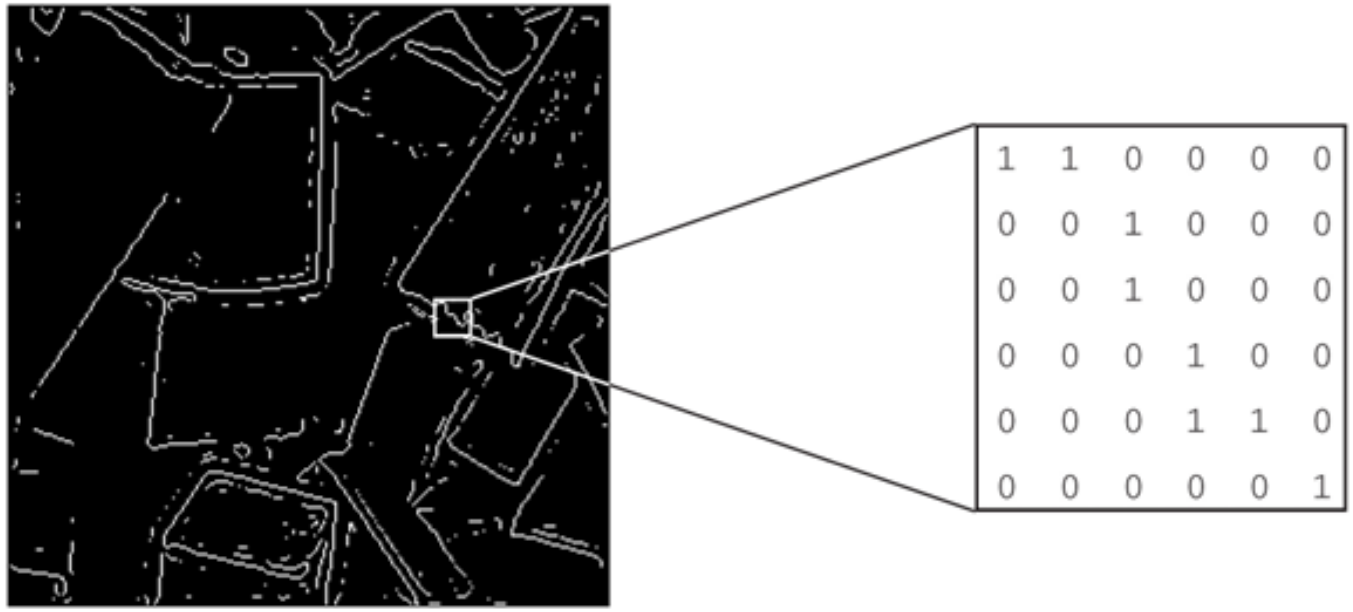
經過取樣後，黑白影像被分割成400（20×20）個樣本

圖像

- 二元圖像：圖像中每個像素的亮度值(Intensity)僅可以取自0或1的圖像，因此也稱為1-bit圖像。
- 灰度圖像：也稱為灰階圖像: 圖像中每個像素可以由0(黑)到255(白)的亮度值(Intensity)表示。0-255之間表示不同的灰度級。
- 彩色圖像：RGB的彩色圖像是由三種不同顏色成分組合而成，一個為紅色，一個為綠色，另一個為藍色。

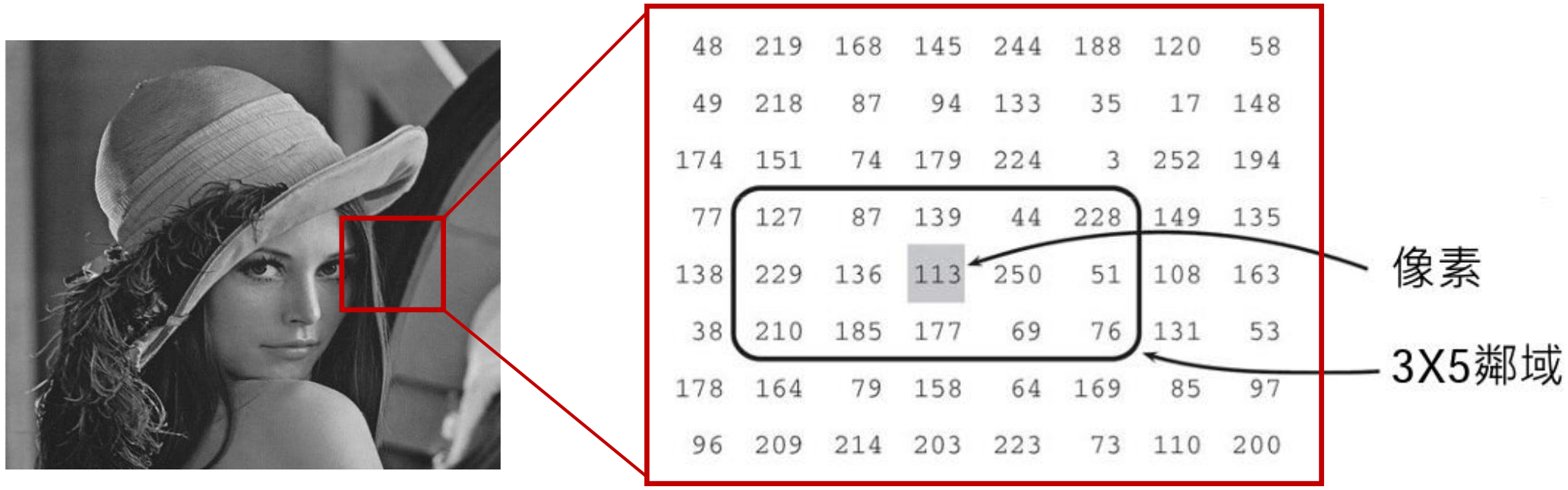
二元圖像

- 圖像中每個像素的亮度值(Intensity)僅可以取自0或1的圖像，因此也稱為1-bit圖像。



灰度圖像

› 圖像中每個像素可以由0(黑)到255(白)的亮度值(Intensity)表示。



彩色圖像

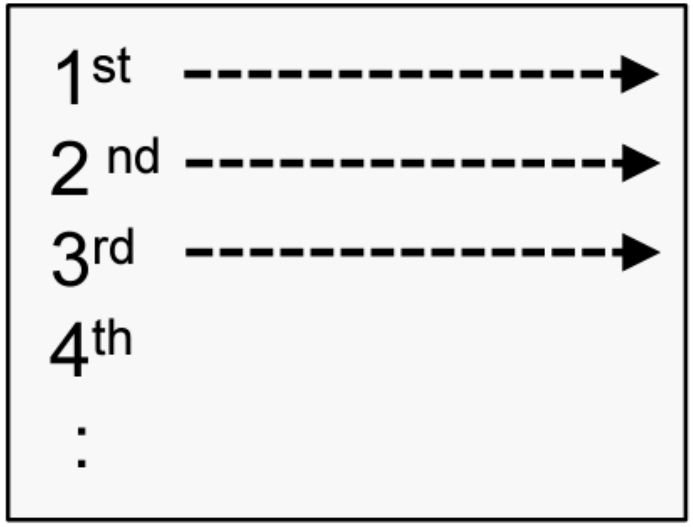
› 由Red、Green、Blue三種顏色組成，每一種顏色的圖都是用0-255的數值來表示，最後合併成為一張彩色的圖像

像素(Pixel) 和 解析度(Resolution)

- 像素為組成數位影像的最小單位
- 解析度為數位影像中的像素數量(例如:1920*1080 或 2,073,6000像素)

影像檔案格式(image file formats)

› 一列一列的紀錄影像灰階或色彩值，又稱為以列為主(row major) 的紀錄方式



Example (a 5×6 image)

1	3	4	6	8	6
5	7	5	..	..	..
:					
:					

至少要記錄 [5; 6; 1, 3, 4, 6, 8, 6, 5, 7, 5, …]

影像處理技術

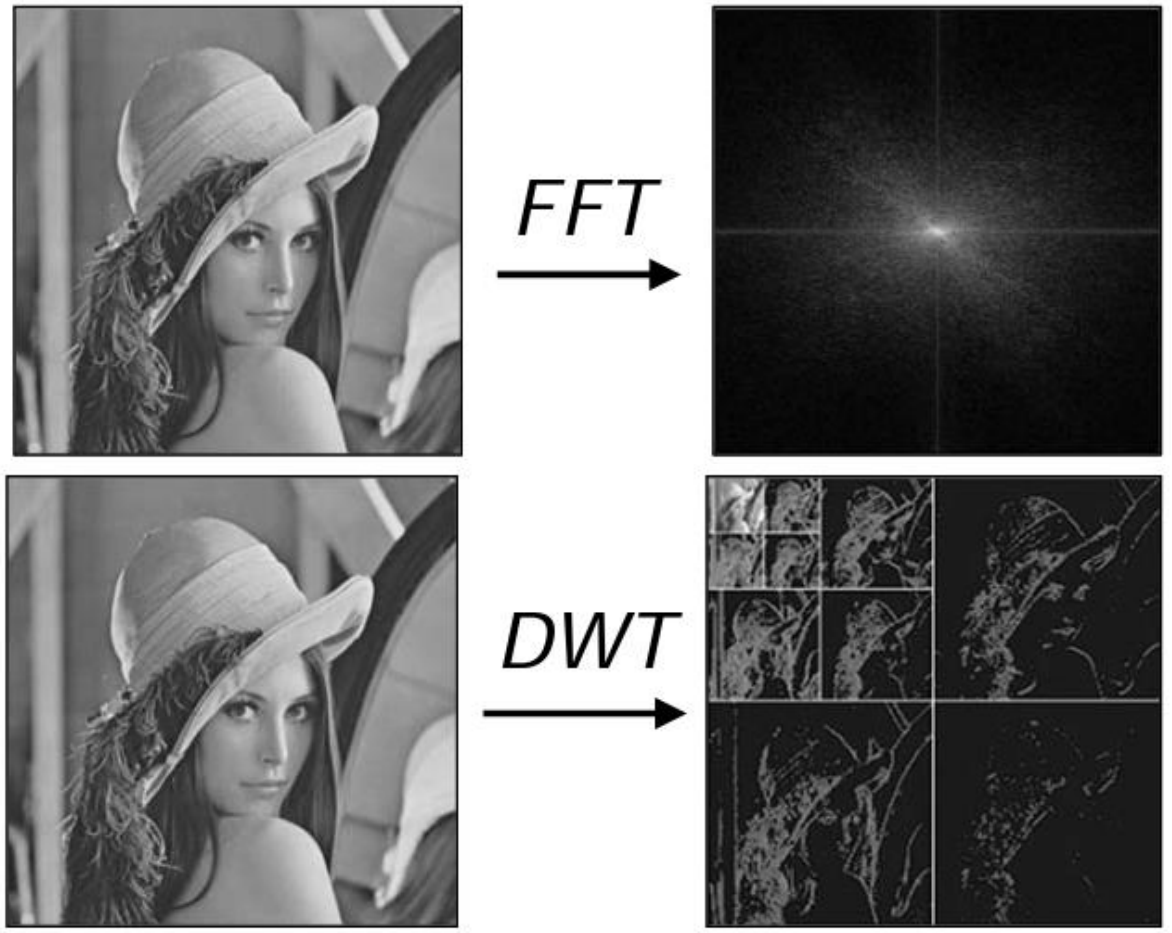
- 常用的影像處理運算(operations)：影像轉換、色彩轉換與分析、影像強化、特徵擷取、影像分割、影像表示與描述、影像壓縮、影像重建、.. 等。
- 與應用相關的技術(techniques)：影像浮水印技術(watermarking)、圖像分類(image classification)、圖形識別(pattern recognition)、三維電腦視覺(3D computer vision)、動態分析與追蹤(motion analysis & tracking)、.. 等。

影像轉換

› 幾何轉換(Geometric Transformations) 是常用的數位影像處理技術，目的是改變數位影像中的空間位置，但不改變其灰階或色彩值。例如: 縮放、旋轉、翻轉等。

影像轉換

› 空間轉換(Transformation)，將圖像空間域的資訊轉換到頻率域上，例如傅立葉轉換和小波轉換



為什麼要空間轉換?

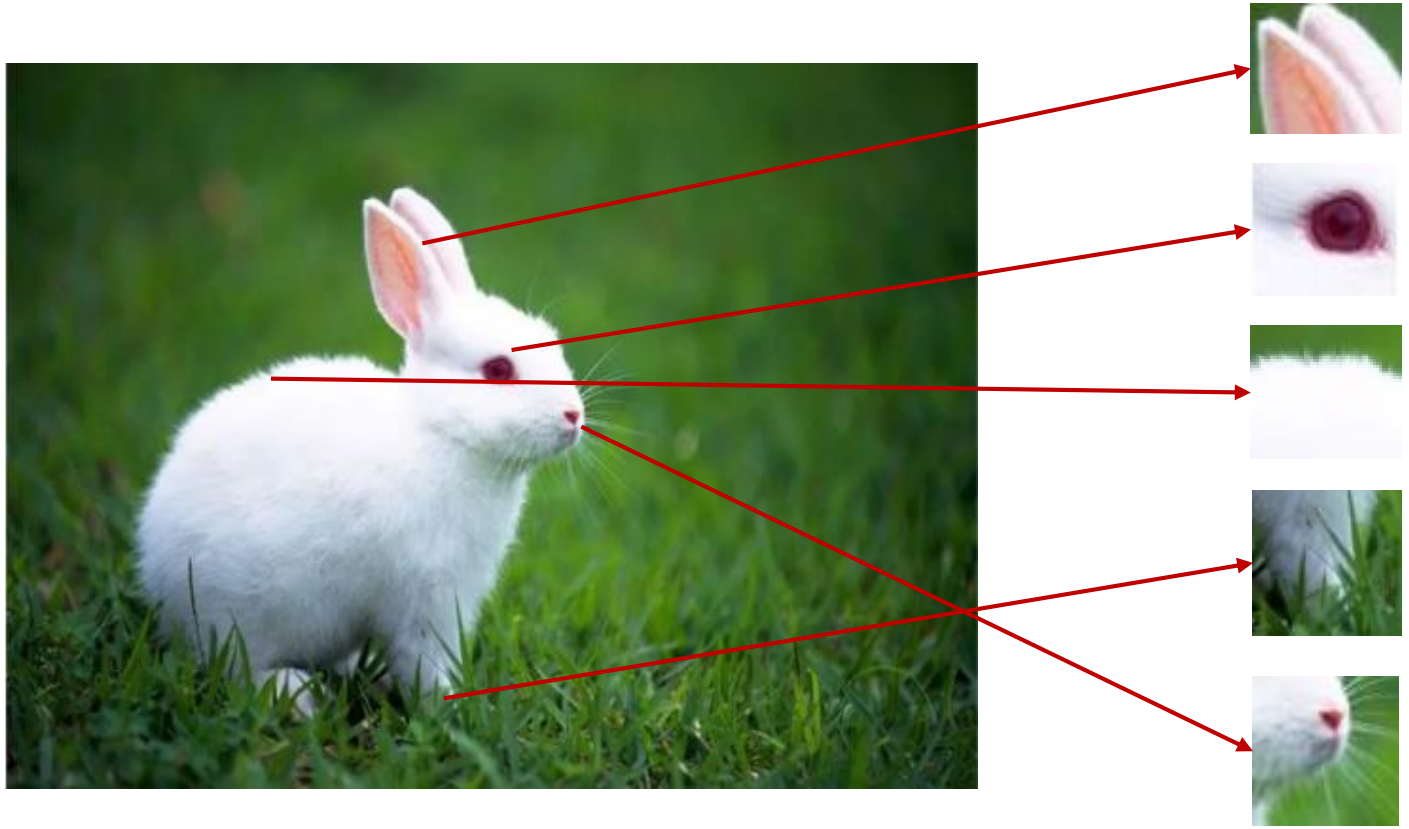
- **為了要模糊、邊緣處理!**
  - 透過一維的傅立葉轉換，將特定頻率的值設為0，逆轉換後可實現濾波功能，達到模糊化處理的效果
  - 透過二維圖像做傅立葉轉換也可以濾波，因為圖像的高頻訊號就視覺上來看，會讓界線較為明顯，如果可以去除低頻訊號，保留高頻訊號，就有機會保留較多的圖像邊緣

影像強化技術

- 強化對比: 擴大影像中灰階(gray level) 或色彩的對比。
- 雜訊去除(noise removing): 去除影像中因不良傳輸或干擾所造成的雜訊。
- 平滑化(smoothing): 去除影像中因不良取像或量化所造成的雜訊，同時會使得影像變模糊。
- 銳利化(sharpening): 強化影像中物體或景觀的邊緣效果。

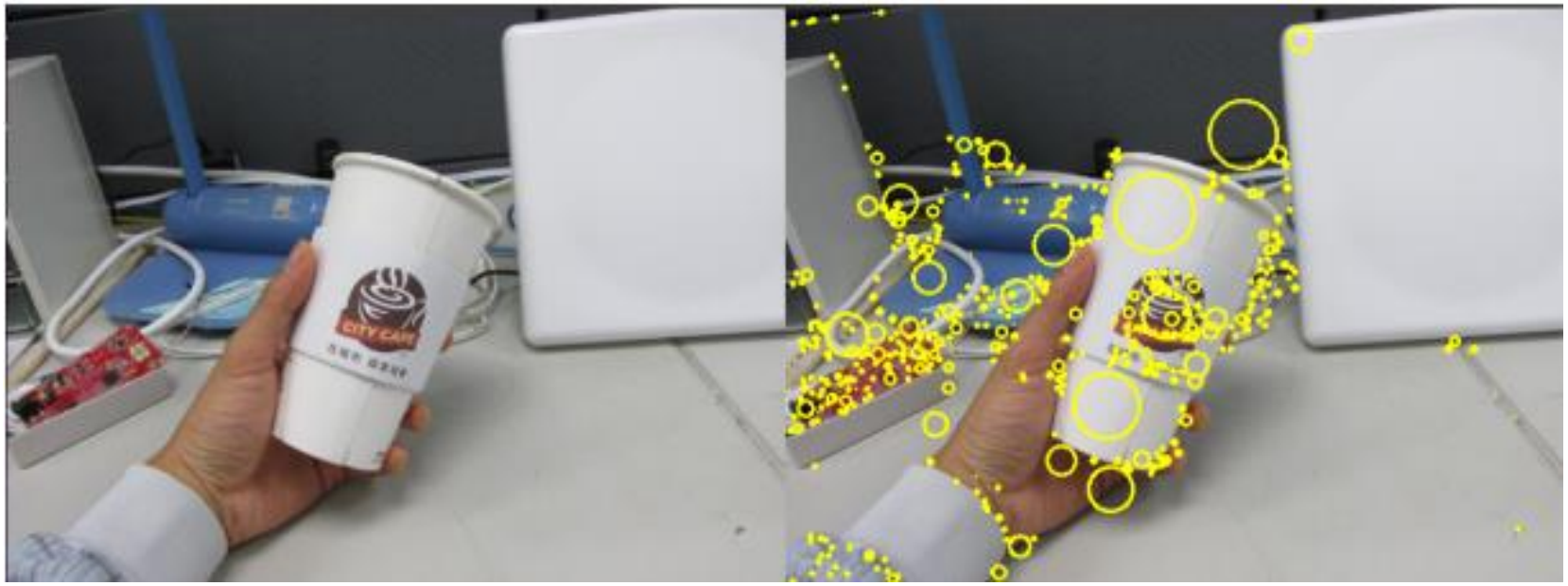
特徵擷取

› 要辨識某物體的條件就是先掌握其特徵！由於我們要辨識的是某個物件而非整張相片，因此需要提取所謂稱為「Local features」的特徵



特徵點偵測

› 在圖片中取得感興趣的關鍵點（可能為edges、corners或blobs



物件偵測

- 在影像辨識中我們會遇到的幾個問題
  - 圖片中有幾個要辨識的物件 (影像切割，Image Segmentation)
  - 他們的位置在哪裡 (物件定位，Object Localization)
  - 要如何辨識(影像分類，Image Classification)
- 而物件偵測 (Object detection) 的技術，就算是物件定位與影像分類的完整解決方案。

影片

- 影片(video):泛指將動態影像以電訊號方式加以捕捉、紀錄、處理、儲存、傳送與重現的各種技術，背後的原理為一連串的靜態畫面(Frame)所組成
  - FPS(Frames Per Second)

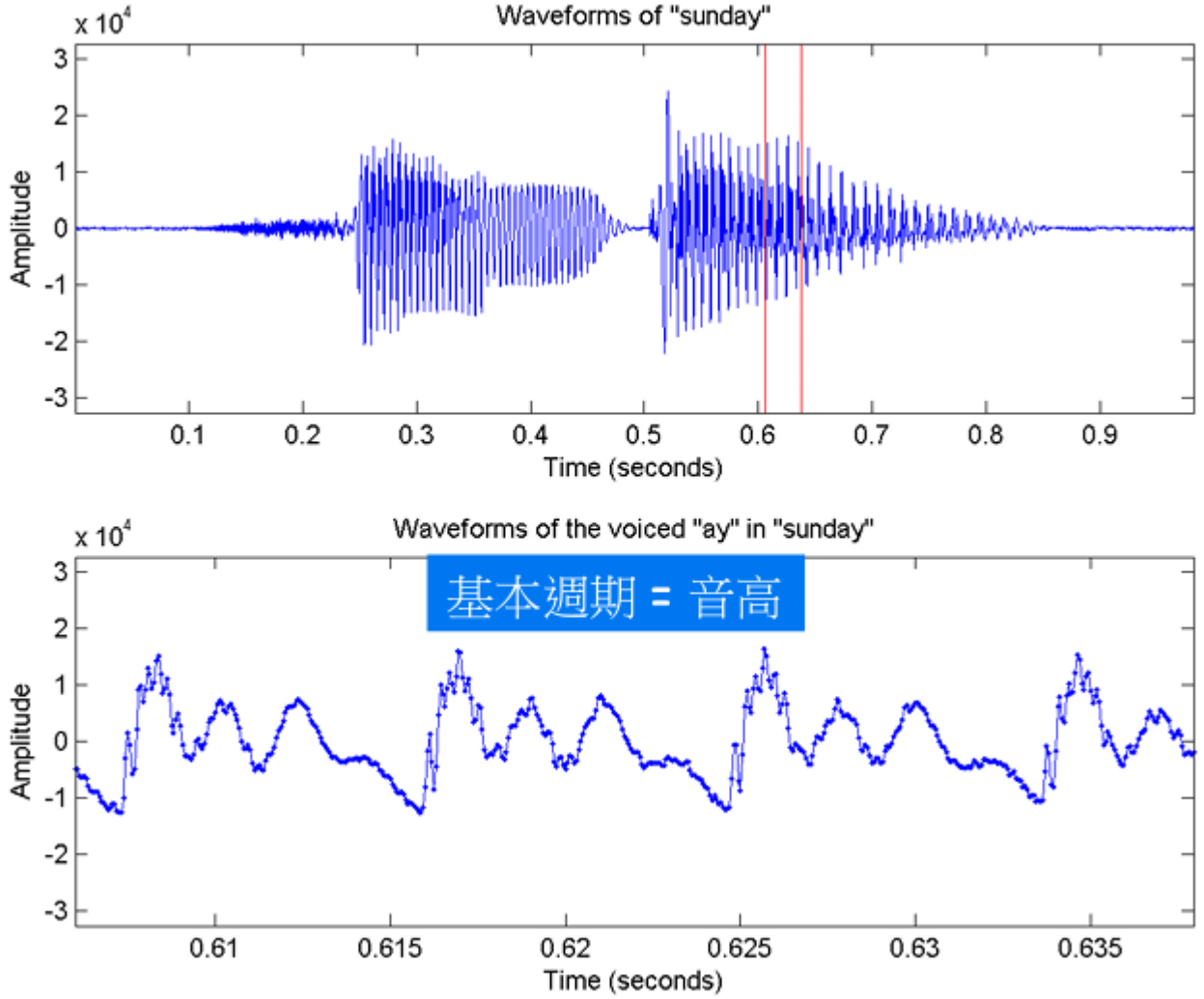
認識音訊

- 音訊可以有很多不同的分類方式，例如，若以發音的來源，可以大概分類如下：
  - 生物音: 人聲 (語音, human voice)、狗聲、貓聲等。
  - 非生物音：引擎聲、關門聲、打雷聲、樂器聲等

訊號的規律性

- 若以訊號的規律性，又可以分類如下：
  - 準週期音：波形具有規律性，可以看出週期的重複性，人耳可以感覺其穩定音高的存在，例如單音絃樂器、人聲清唱等。
  - 非週期音：波形不具規律性，看不出明顯的週期，人耳無法感覺出穩定音高的存在，例如打雷聲、拍手聲、敲鑼打鼓聲、人聲中的氣音等。

音訊資料

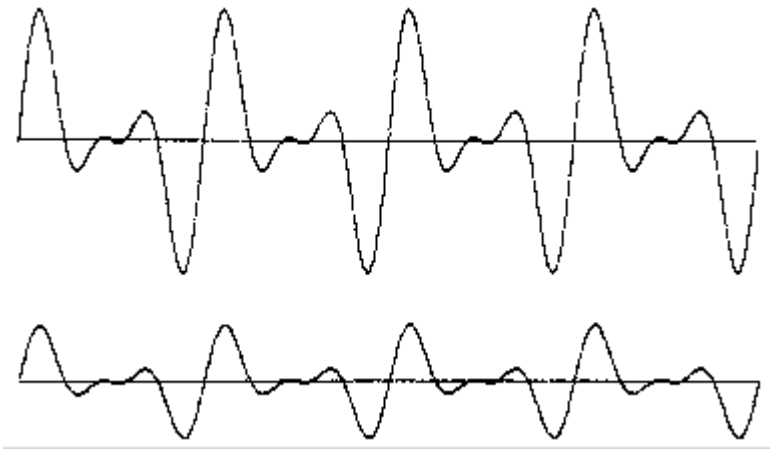

Waveforms of "sunday"
x 10^4
Amplitude
Time (seconds)
Waveforms of the voiced "ay" in "sunday"
基本週期 = 音高

音訊特徵

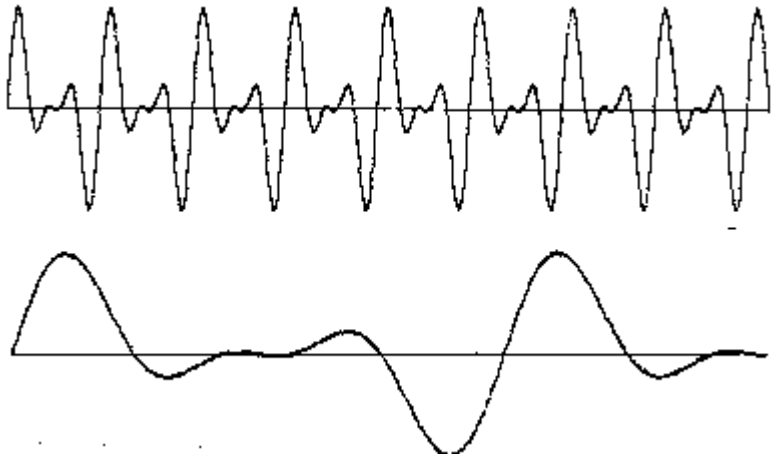
- 在一個特定音框內，我們可以觀察到的三個主要聲音特徵可說明如下：
  - 音量（Volume）：代表聲音的大小，可由聲音訊號的震幅來類比，又稱為能量（Energy）或強度（Intensity）等。
  - 音高（Pitch）：代表聲音的高低，可由基本頻率（Fundamental Frequency）來類比，這是基本週期（Fundamental Period）的倒數。
  - 音色（Timbre）：代表聲音的內容（例如英文的母音），可由每一個波形在一個基本週期的變化來類比。

音訊特徵

› 音量大小

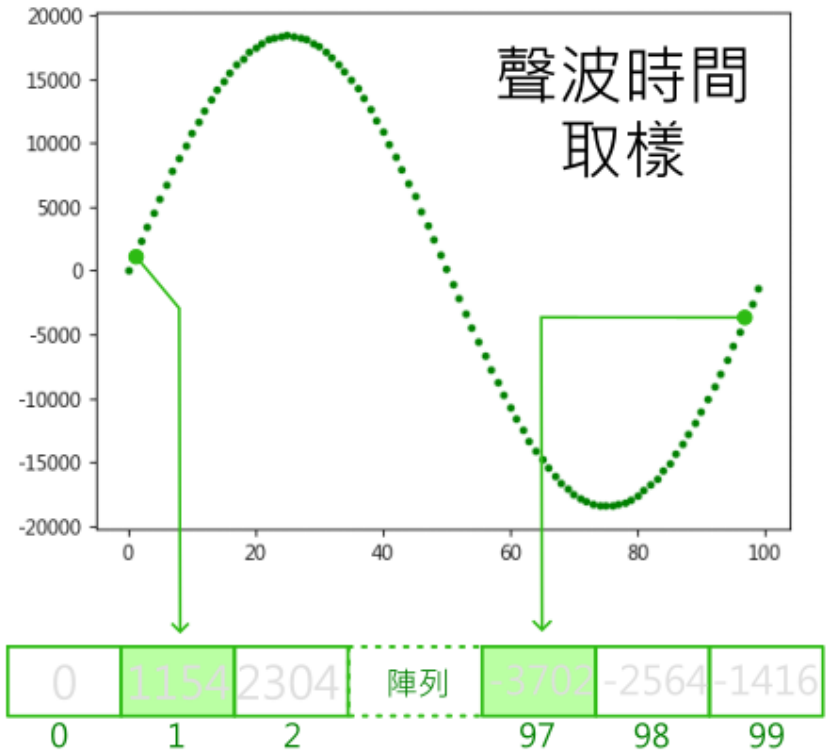


› 音高高低



音訊數位化

- 真實世界中的聲波是連續的類比訊號，如果要將聲波數位化，變成一個個離散的數位訊號，就必須對聲音訊號做「取樣」的動作，取樣的資料因為具有相同型態，多以陣列的資料結構存放



取樣率 sample rate

1秒內取多少點

取樣週期 sample period

隔多久取1點

CD音質取樣率44100

每1秒取44100點

每1/44100秒取1點
(約0.00002秒)

音訊資料處理

- 音訊編碼: 將聲音用數位化的方式編碼並儲存
- 音量調整: 調整音訊的音量
- 音訊拼接: 將音訊做剪接的操作
- 音訊反轉: 將音訊做反轉
- 音訊合成: 將多個音訊疊加並合成為新的音訊