Lecture 5: Multimedia Information Retrieval
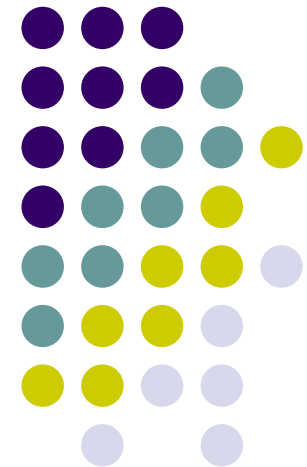
DR. JIAN ZHANG

NICTA & CSE UNSW

COMP9314 Advanced Database

S1 2007

jzhang@cse.unsw.edu.au

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL ICT AUSTRALIA

Course Objectives & Outlines

- Objectives:
  - On successful completion of this subject, students will:
    - understand fundamental concepts, theory and techniques:
      - multimedia content description
      - multimedia database indexing, browsing and retrieval
    - be familiar with applications of multimedia systems and their implementations;
    - gain skills and knowledge beneficial to future work and post-graduate study in multimedia area

- Outlines:
  - Basic concepts for multimedia application and research
  - Multimedia data types and formats
  - Multimedia indexing and retrieval

Reference Books

- [1] Multimedia database management systems --Guojin **Lu.**
  - Publication Details   Boston, MA : Artech House, 1999.

- [2] **Introduction to MPEG-7** : multimedia content description interface -- edited by B.S. Manjunath, Phillipe Salembier, Thomas Sikora.
  - Publication Details Chichester ; Milton (Qld.): Wiley, 2002

- [3] Multimedia information retrieval and management
  : technological fundamentals and applications / **David** Dagan **Feng**, Wan-Chi Siu, Hong-Jiang Zhang (eds.).
  - Publication Details   Berlin ; New York : Springer, 2003.

- [4] Digital Image Processing -- Rafeal Gonzalez

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.0 Introduction

- The needs to develop multimedia database management
  - Efficient and effective storage and retrieval of multimedia information become very critical

  - Traditional DBMS is not capable of effectively handling multimedia data due to its dealing with alphanumeric data

  - Characteristics and requirements of alphanumeric data and multimedia data are different

  - A key issue in multimedia data is its multiple types such as text, audio, video, graphics etc.

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA
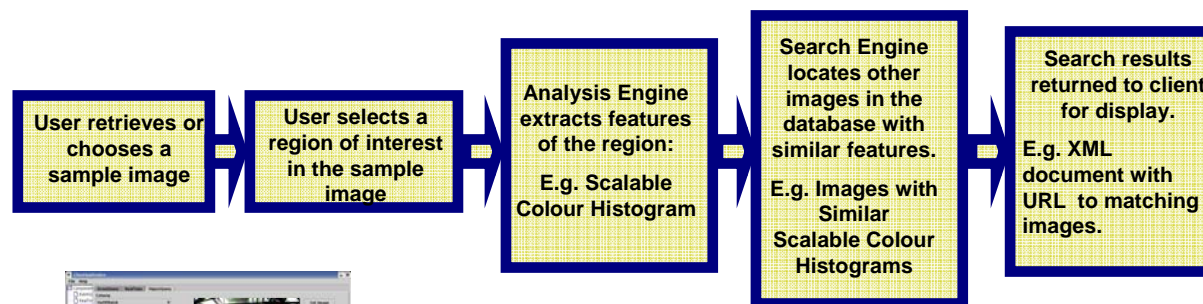
NATIONAL ICT AUSTRALIA

5.0 Introduction

- The fundamental of Multimedia Database (Content) Management research covers:

  - Feature extraction from these multiple media types to support the information retrieval.

  - Feature dimension reduction – High dimensional features

  - Indexing and retrieval techniques for the feature space
    - Similarity measurement on query features

  - How to integrate various indexing and retrieval techniques for effective retrieval of multimedia documents.

  - Same as DBMS, efficient search is the main performance concern

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL ICT AUSTRALIA

5.1 Multimedia Applications-- Content Management Demonstration Platform

- Client / Server platform demonstrating content based search using MPEG-7 visual descriptors
  - Content can be searched using methods "query by specification" or "query by example".
  - For "query by example", the Analysis Engine at the server extracts visual features from images
  - the Search Engine searches for archived images that have similar features to those of the example region.

| User retrieves or chooses a sample image | User selects a region of interest in the sample image | Analysis Engine extracts features of the region:

E.g. Scalable Colour Histogram | Search Engine locates other images in the database with similar features.

E.g. Images with Similar Scalable Colour Histograms | Search results returned to client for display.

E.g. XML document with URL to matching images. |

Query by Example

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL ICT AUSTRALIA

5.2 Multimedia Content Management

- Why we need multimedia content management ?
  - There is a strong need to effectively manage the vast of amount of resulting multimedia content

    - It is necessary to develop forms of audiovisual information representation that go beyond the simple formats

    - It is necessary to provide a common multimedia content description interface (defined by MPEG-7 standard)

    - It is necessary to develop a rich set of standardized tools to describe multimedia content.
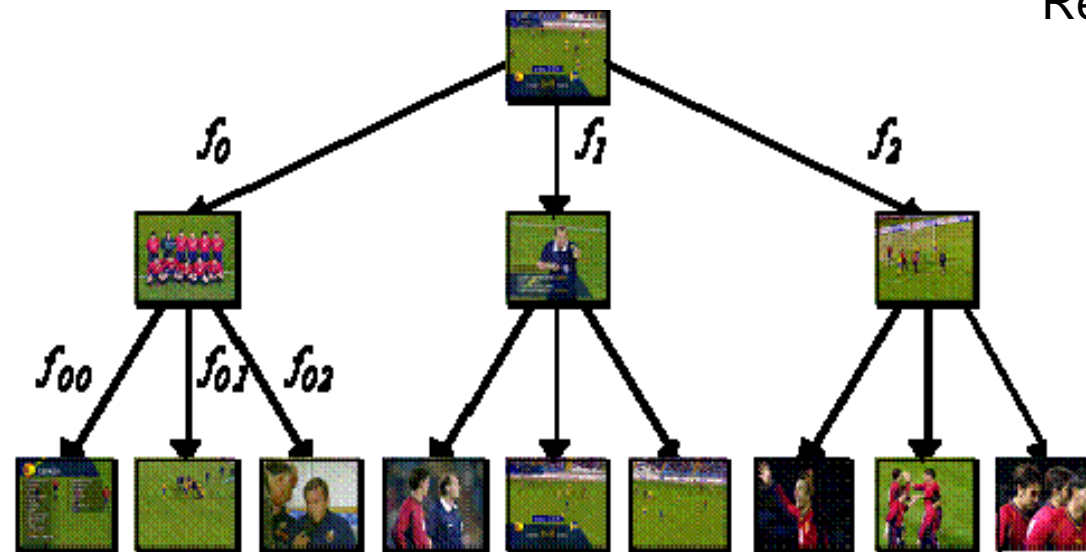
5.2 Multimedia Content Management

- Why we need multimedia content management ?
  - It is necessary to develop a platform to handle the following cases

    - where audiovisual information is created, exchanged, retrieved, and re-used by computational systems

    - where information retrieval is required for quickly and efficiently searching for various types of multimedia documents of interests of the users

    - where a stream of audiovisual content description is needed for users to receive only those multimedia data items which satisfy their preference

5.2 Multimedia Content Management

- Example of a Hierarchical Summary of a video of a soccer game -- a multiple level key-frame hierarchy

Ref: J. Martinez

The Hierarchical Summary denotes the fidelity (i.e., f0, f1) of each key-frame with respect to the video segment referred to by the key-frames at the next lower level.

THE UNIVERSITY OF NEW SOUTH WALES SYDNEY · AUSTRALIA

NATIONAL ICT AUSTRALIA

5.2 Multimedia Content Management

- The Space and Frequency Graph describes the decomposition of an audio or visual signal in space (time) and frequency

Ref: J. Martinez
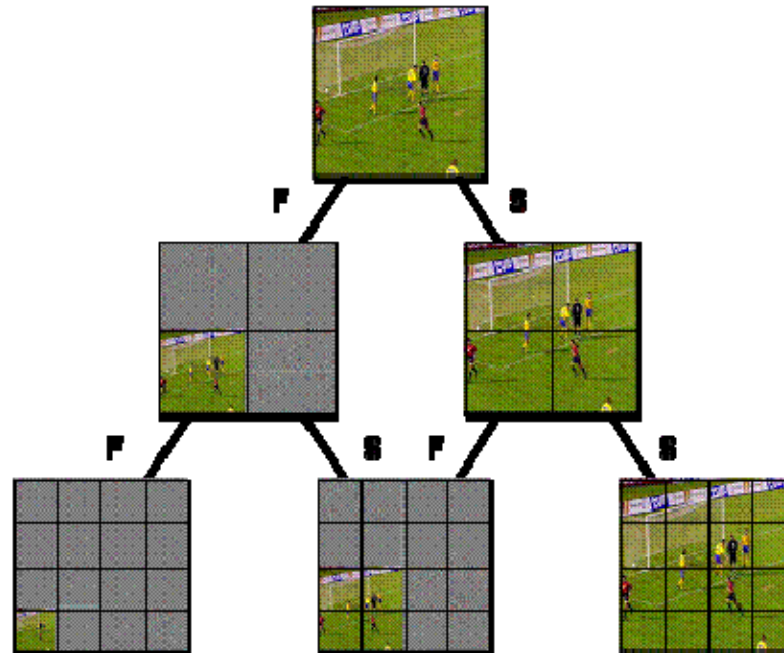
5.2 Multimedia Content Management

- Given that the strong need for multimedia content management, there are some key challenges:

  - Majority of existing techniques for content management are based on low-level features

  - There is a significant gap between:
    - low-level feature extraction and user expectation on high level understanding (semantic level)

  - Video analysis and understanding technologies (tools) serve as the key enabling technology towards semantic content description and representation

  - **This field of research presents significant challenges and enormous opportunities !!!**

5.3 Layered Multimedia Computing Research

- A three-layer of Multimedia Computing Research
  - Image/Video Signal Processing (Demo)

    - Image/video signal processing is the basis for high level of video analysis.

    - Topics include

      - image/video pre-processing

      - pixel based feature extraction

      - statistical feature descriptors for the next level of processing in content analysis.

THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY · AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.3 Layered Multimedia Computing Research

- Video Content Analysis (Demo)

  - This research investigates the provision of content features and object information to enable meaningful video content presentation.

  - Topics include
    - Key frame extraction and video shot segmentation
      - The basic element to build video scene (story) segmentation

    - Object segmentation & tracking (multiple objects and occlusion)

    - Object classification
      - Classify human being, car, trucks, in indoor/outdoor area

THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.3 Layered Multimedia Computing Research

- Video Content Understanding (Demo)

  - Aims to achieve semantic and structural representation (Ontology) of video content to enable meaningful content search and retrieval

  - Topics include

    - Video summarization towards table of video content generation
      - E.g: video shot with semantic description and scene (story) generation

    - Automatic and semi-automatic annotation of image/video
      - E.g: supervised learning to build statistical mode for video sequence annotation – indoor/outdoor, car, sky etc

    - Semantic video representation
      - E.g: different modalities video to key frame plus text, video to synthetic video with audio explanation.

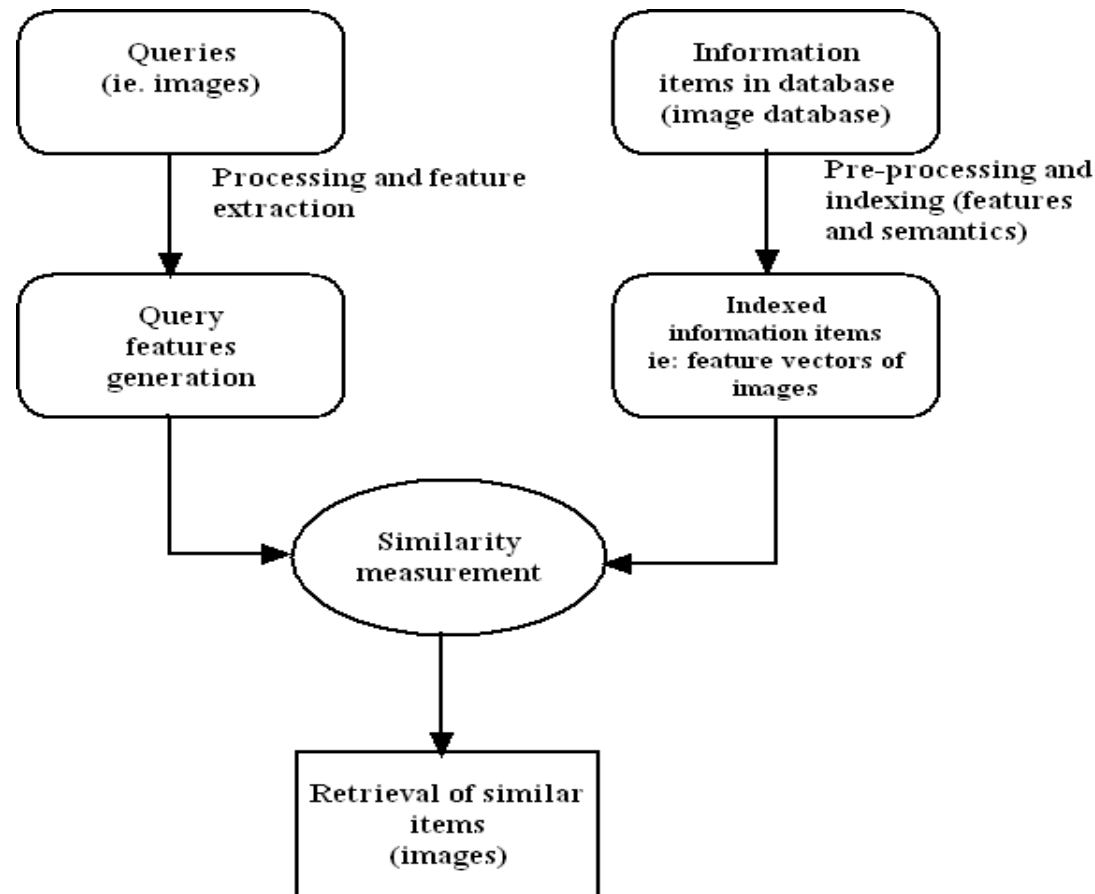# 5.4 Multimedia Information Retrieval Systems (MIRS)

- The needs for MIRS

  - A vast multimedia data – captured and stored

  - The special characteristics and requirements are significantly different from alphanumeric data.

  - Text Document Information Retrieval (Google search) has limited capacity to handle multimedia data effectively.

5.4 Multimedia Information Retrieval Systems (MIRS)

- An overview of MIRS operation

5.4 Multimedia Information Retrieval Systems (MIRS)

- Expected Query types and Applications

  - Metadata-based quires
    - Timestamp of video and authors' name

  - Annotation-based quires (event based quires)
    - Video segment of people picking up or dropping down bags

  - Queries based on data patterns or features
    - Color distribution, texture description and other low level statistical information

  - Query by example
    - Cut a region of picture and try to find those regions from pictures or videos with the same or similar semantic meaning

THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.5 Introduction to Image Indexing and Retrieval

- Four main approaches to image indexing and retrieval

    - Low level features -- Content based Image Retrieval (CBIR)

    - Structured attributes – Traditional database mgt. system

    - Object-recognition – Automatic object recognition

    - Text – Manual annotation (Google search)

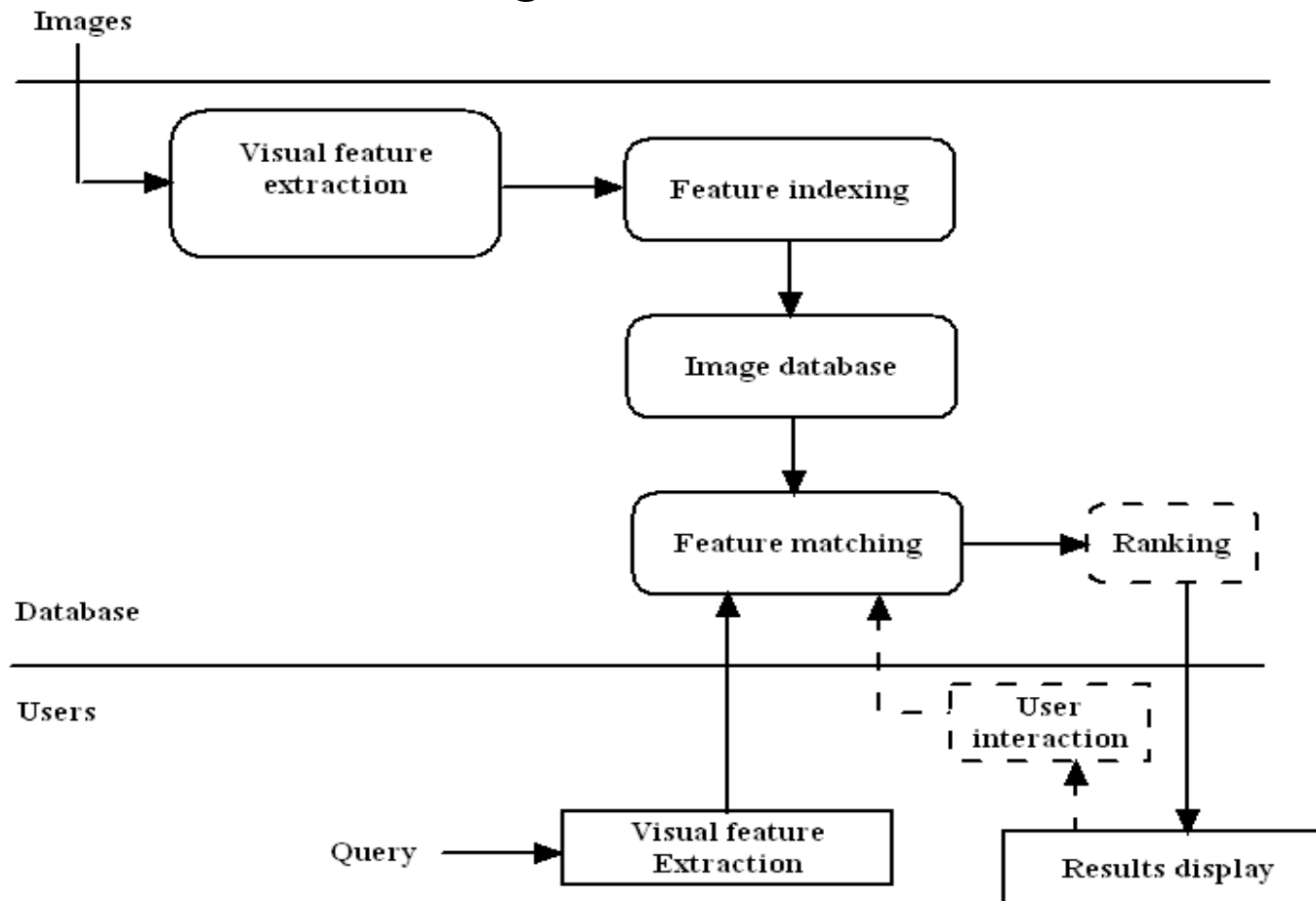# 5.5 Introduction to Image Indexing and Retrieval

- Four main approaches to image indexing and retrieval
  - Content based Image Retrieval (CBIR)– low level features

    - Extract low level image features (color, edge, texture and shape)

    - Expand these image feature towards semantic levels

    - Index on these images based on similar measurement

    - Relevance feedback to refine the candidate images
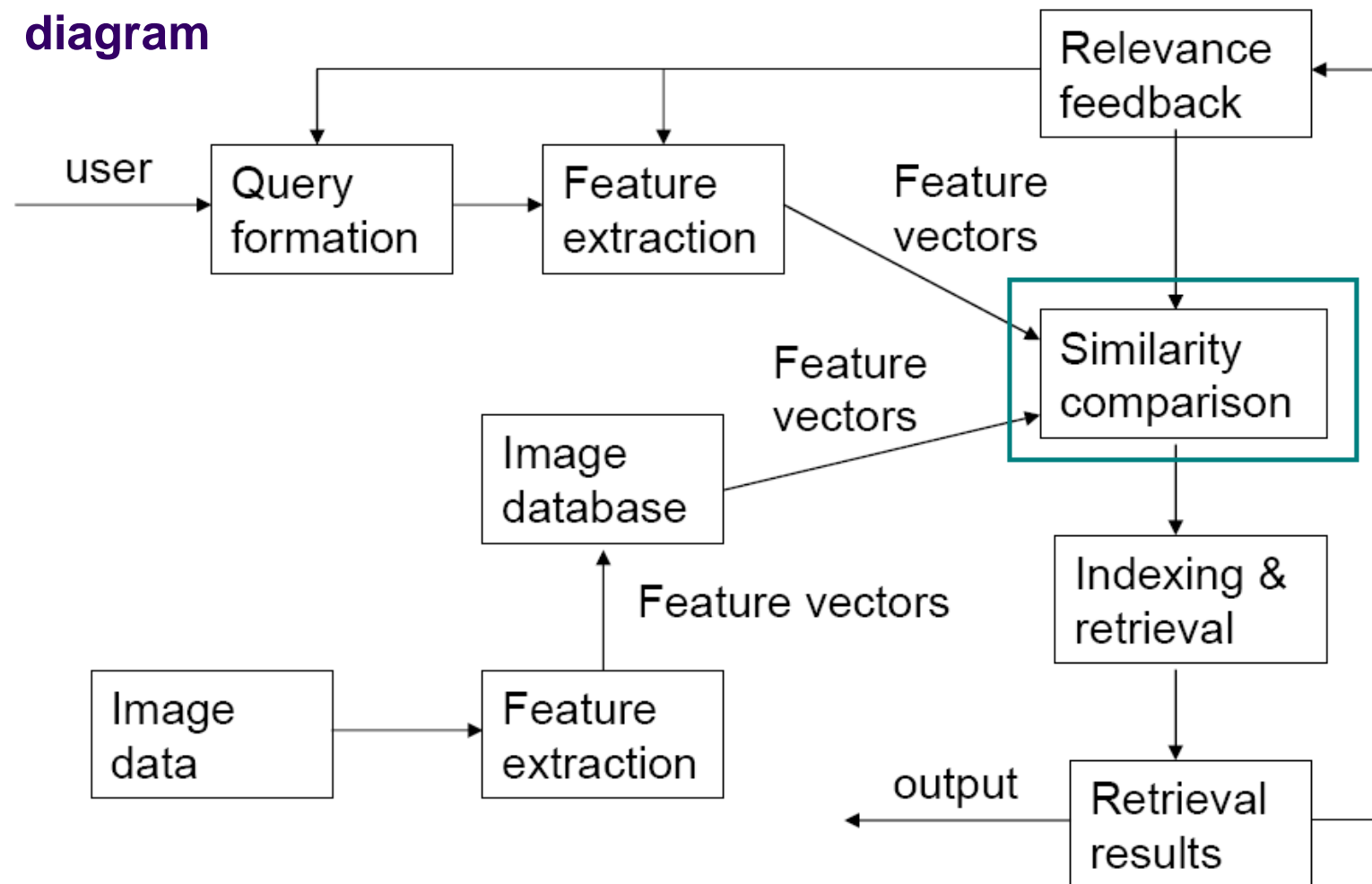
5.5 Introduction to Image Indexing and Retrieval

- Content based image retrieval

Images

```
┌──────────────────┐        ┌──────────────────┐
│  Visual feature  │───────▶│  Feature indexing │
│    extraction    │        └──────────────────┘
└──────────────────┘                 │
                                      ▼
                          ┌──────────────────┐
                          │  Image database  │
                          └──────────────────┘
                                      │
                                      ▼
                    ┌──────────────────┐      ┌ ─ ─ ─ ─ ─ ─ ┐
                    │ Feature matching │─────▶   Ranking
                    └──────────────────┘      └ ─ ─ ─ ─ ─ ─ ┘
```

Database

Users

```
                 ┌ ─ ─ ─ ─ ─ ┐
                    User
                  interaction
                 └ ─ ─ ─ ─ ─ ┘
                      ▲
┌──────────────────┐        ┌──────────────────┐
Query ───▶│  Visual feature  │        │  Results display │
          │    Extraction    │        └──────────────────┘
          └──────────────────┘
```

5.5 Introduction to Image Indexing and Retrieval

diagram

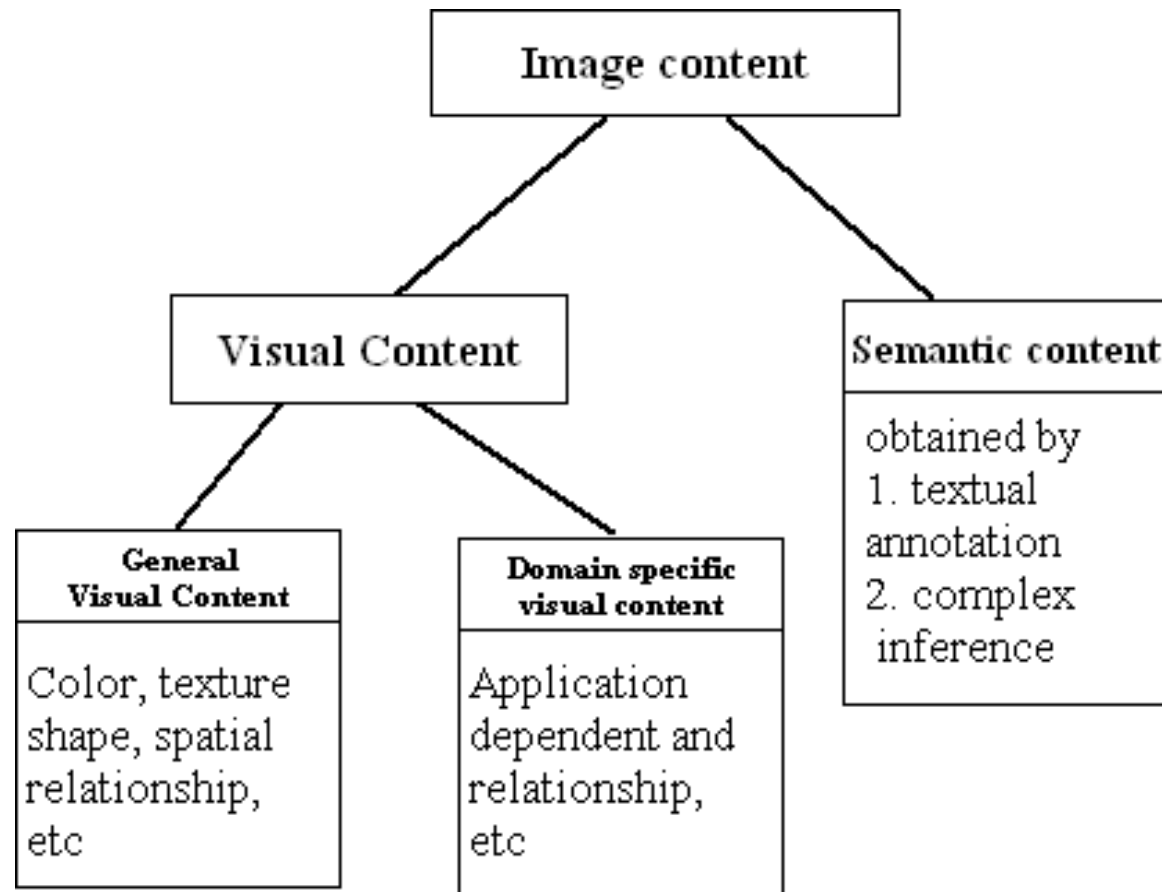# 5.5 Introduction to Image Indexing and Retrieval

- Image representation

  - A visual content descriptor can be either global or local.

  - The global descriptor uses the visual features of the whole image

  - A local descriptor uses the visual features of regions or objects to describe the image content, with the aid of region/object segmentation techniques

Image Representation

5.5 Introduction to Image Indexing and Retrieval

- Image representation


```
                  ┌─────────────────┐
                  │  Image content  │
                  └─────────────────┘
                    /             \
    ┌───────────────┐          ┌──────────────────┐
    │ Visual Content│          │ Semantic content │
    └───────────────┘          ├──────────────────┤
        /        \             │ obtained by      │
┌──────────┐  ┌──────────────┐ │ 1. textual       │
│ General  │  │Domain specific│ │    annotation    │
│ Visual   │  │visual content │ │ 2. complex       │
│ Content  │  ├──────────────┤ │    inference     │
├──────────┤  │ Application   │ └──────────────────┘
│Color,    │  │ dependent and │
│texture   │  │ relationship, │
│shape,    │  │ etc           │
│spatial   │  └──────────────┘
│relationship,│
│etc       │
└──────────┘
```

5.6 Low level Feature Extraction -- Color Representation

- Color
  - Color is very powerful in description and of easy extraction from nature images in its considerable variance changes:
    - Illumination
    - Orientation of the surface
    - Viewing geometry of the camera
- Color fundamentals

Ref: Gonzalez and Woods, digital image processing
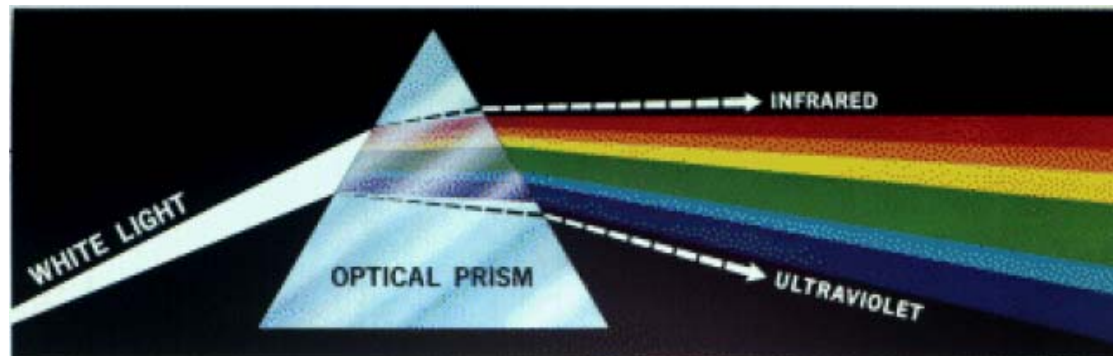
Plate I. Color spectrum seen by passing white light through a prism. (Courtesy of General Electric Co., Lamp Business Division.)

- Light of different wavelengths produces different color sensations such as in different broad regions (violet, blue, green, yellow, orange and red)

5.6 Low level Feature Extraction -- Color Representation

- Color fundamentals
  - The colors that humans perceive in an object are determined by the nature of the light reflected from the object.
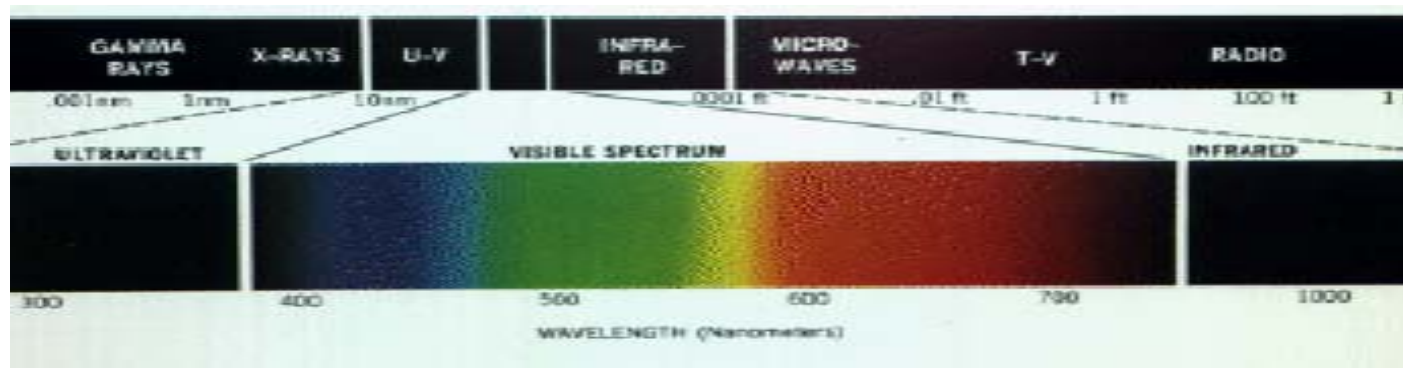
Ref: Gonzalez and Woods, digital image processing

  - Visible light is electromagnetic radiation with a spectrum wavelength ranging approximately from 400 to 780 nm.
  - Red, Green and Blue are the additive primary colors. Any color can be specified by just these three values, giving the weights of these three components

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.6 Low level Feature Extraction -- Color Representation

- ## Color space
  - ### RGB (Red, Green and Blue) space
    - The RGB color space is the most important means of representing colors used in multimedia.
    - A color can be represented in a form (r-value,g-value,b-value). The value in here is defined as the percentage of the pure light of each primary.

      Examples:
   (100%,0%,0%) – pure saturated primary red
      (50%,0%,0%) – a darker red
      (0%,0%,0%) – black
      (100%,100%,100%) -- white

    - A Cartesian Coordinate System is defined to measure each color with a vector.

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL ICT AUSTRALIA

5.6 Low level Feature Extraction -- Color Representation

- RGB (Red, Green and Blue) space

  - The value range for each primitive color is from 0 to 255 which is a 8-bit byte. Thus, a RGB color can be represented by 24 bits, three bytes

  - In a practical system, a RGB color can hold different bits such as 24-bit, 15-bit and 12-bit color depth.
    - 24-bit -- full RGB color space
    - 15-bit – 5-bit for R, 6-bit for G and 5-bit for B
    - 12-bit – 4-bit for R, 4-bit for G and 4-bit for B
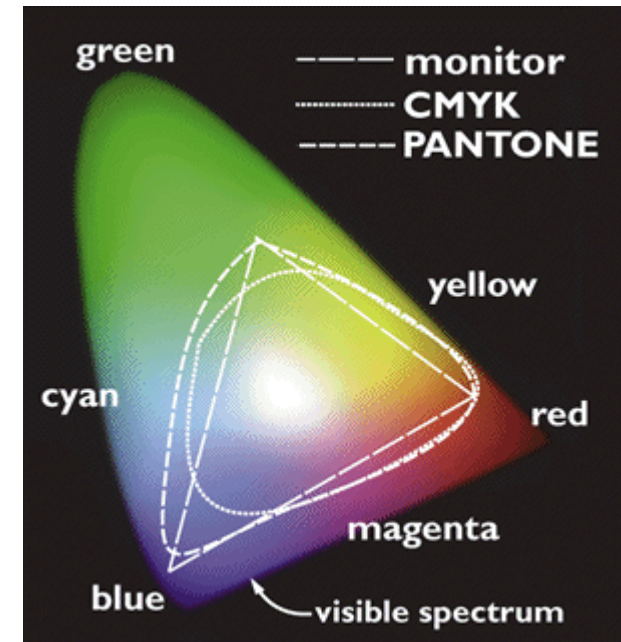
5.6 Low level Feature Extraction -- Color Representation

- HSV space
  - From physical properties of color radiation, three basic components called Hue, Saturation and Value (HSV) of a pixel form another method for representing the colors of an image.
   - The value of a pixel can be either Intensity or Brightness
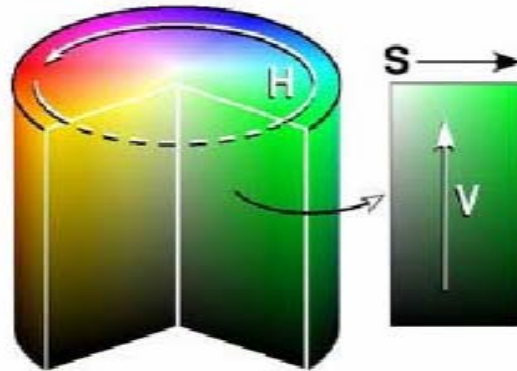
HSV color space as a cylindrical object

  - Hue is the attribute of a visual sensation according to which an area appears to be similar to one of the perceived colors such as red, yellow, green and blue.
    - Hue is usually represented in the range from 0 to 360 degrees. For example, the color located at 90 degree corresponds to yellow and green

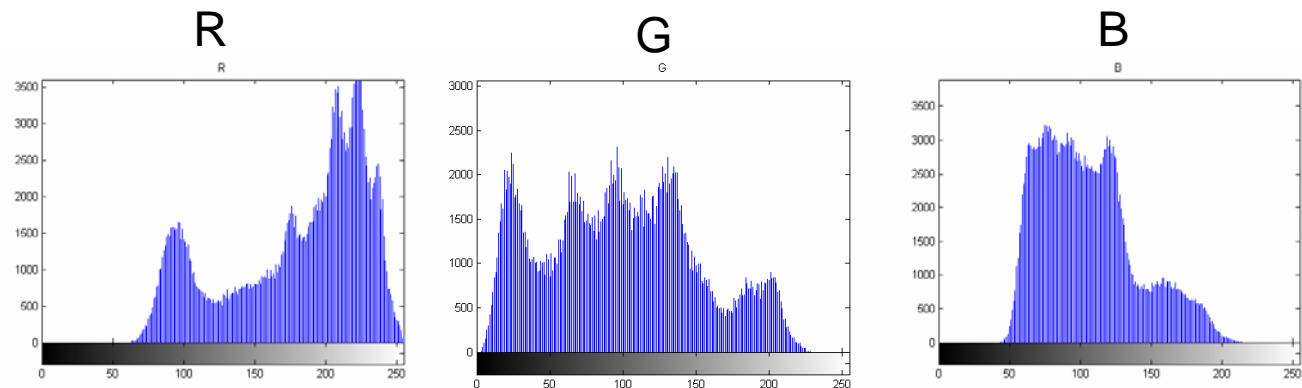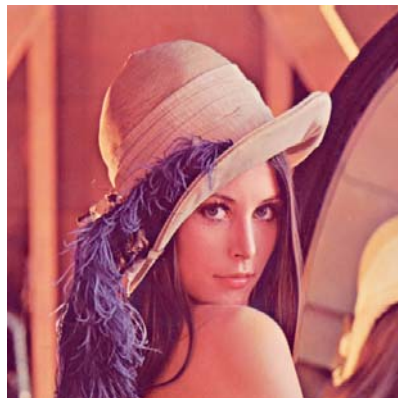# 5.6 Low level Feature Extraction -- Color Representation

- HSV space
  - Saturation is the colorfulness of an area judged in proportion to its brightness. For example, a pure color has a saturation 100%, while a white color has a saturation 0%.

  - Luminance/Brightness is the attribute of a visual sensation to which an area appears to emit more or less light.

H - Hue
S – Saturation
L/B – Luminance/Brightness
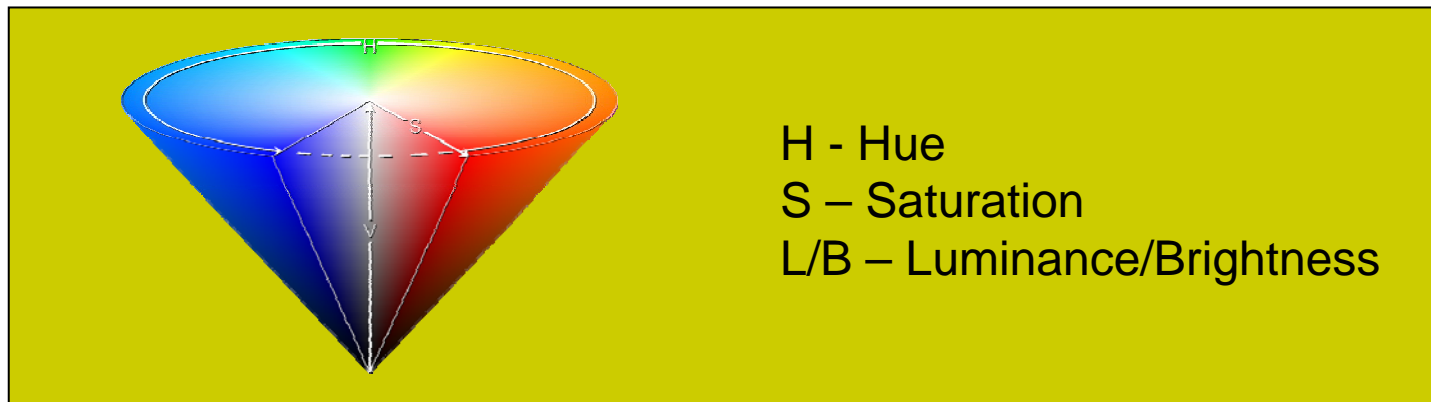
5.6 Low level Feature Extraction -- Color Representation

- ## Color descriptors

  - ### Color histogram

    - It characterizes the distributions of colors in an image both globally and locally

    - Each pixel can be described by three color components.

      - A histogram for one component describes the distribution of the number of pixels for that component color in a quantitative level – a quantized color bin.

      - The levels can be 265, 64, 32, 16, 8, 4, 1 (8-bit byte)

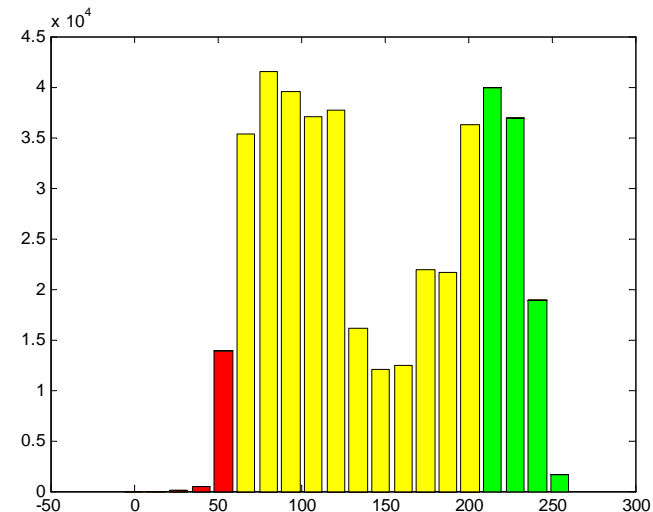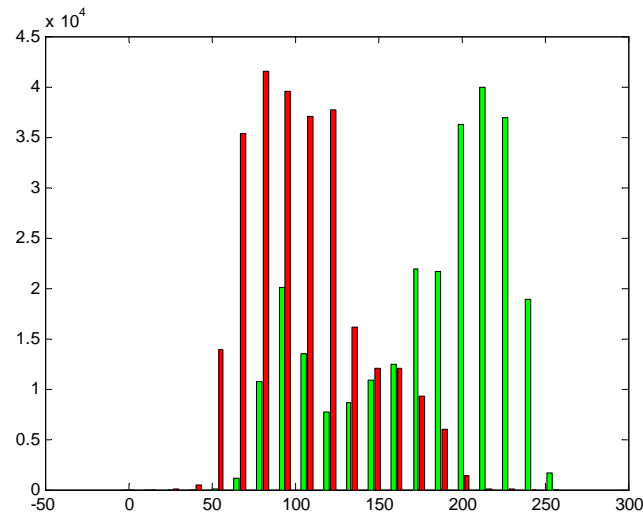# 5.6 Low level Feature Extraction -- Color Representation

- Color histogram

  - In general, if more bins are defined in histogram calculation, it represents the more discrimination power. However,

    - It will increase the computation cost if use a combined color bin histogram systems
      - E.g. R*G*B = 256*256*256 = 16777216 bins!

    - it might generate color indexes for image database inappropriately

    - In some cases, it might not help the image retrieval performance

  - A effective method should be developed to select an adequate color bin numbers for different image retrieval systems.

THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.6 Low level Feature Extraction -- Color Representation

- Color Histogram Intersection

  - Histogram Intersection is employed to measure the similarity between two histograms

$$S(I_p, I_q) = \frac{\sum_{i=1}^{N} min(H_i(I_p), H_i(I_q))}{\sum_{i=1}^{N} H_i(I_q)}$$

Colors that are not present in the query image do not contribute to the intersection distance

# 5.6 Low level Feature Extraction -- Color Representation

- Scalable color descriptor

  - A Haar transform-based encoding scheme

    - It applies across values of a color histogram in the HSV color Space

    - The basic unit of the transform consists of low-pass and high-pass filters.

    - The HSV color space for scalable color descriptor is uniformly quantized into a combined 256 bins – 16 levels in H, 4 levels n S and 4 levels in V.

5.6 Low level Feature Extraction -- Color Representation

- Scalable color descriptor
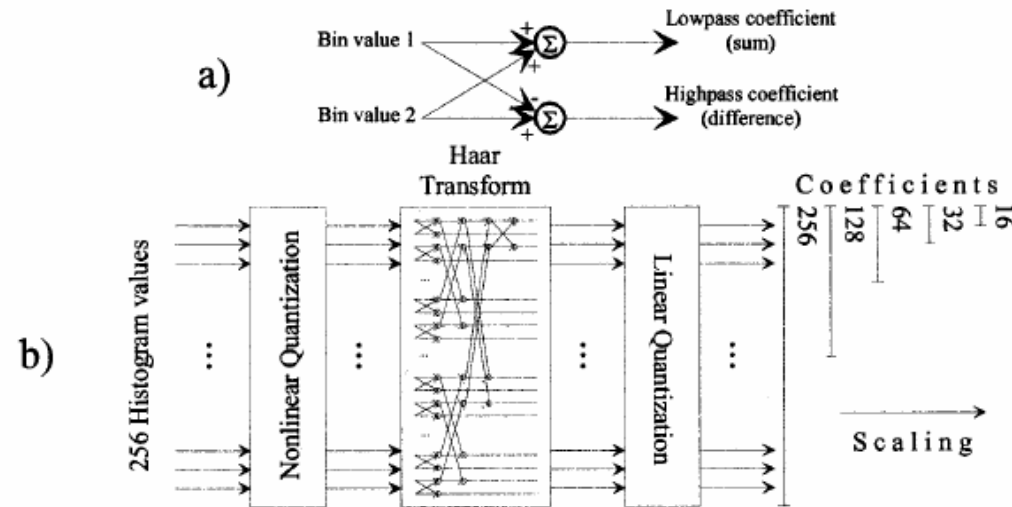
- Since the interoperability between different resolution levels is retained, the matching based on the information from subsets of the coefficients guarantees an approximation of the similarity in full color resolution

5.6 Low level Feature Extraction -- Color Representation

- Color Coherence Vector
  - Motivation

    - Color histogram is sensitive to both compression artifacts and camera auto-gain.

    - Color histogram is suitable for image content representation if the color pattern is unique compared with the rest of the dataset

    - Color histogram does not present spatial information
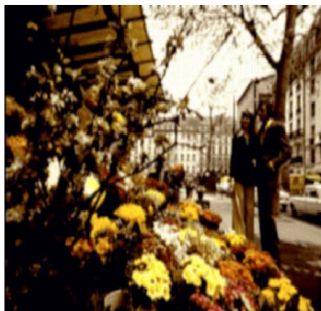
These two images have very similar color histograms, despite their rather different appearances.

5.6 Low level Feature Extraction -- Color Representation

- Color Coherence Vector

  - Can we do something better?

    - The color coherence vector (CCV) is a tool to distinguish images whose color histograms are indistinguishable

    - The CCV is a descriptor that includes relationship between pixels – spatial information

5.6 Low level Feature Extraction -- Color Representation

- Color Coherence Vector (CCV)

  - A color's coherence is defined as the degree to which pixels of that color are members of large similar-color regions.

  - These significant regions are referred as coherent regions which are observed to be of significant importance in characterizing images

  - Coherence measure classifies pixels as either coherent or incoherent

  - A color coherence vector represents this classification for each color in the image.

5.6 Low level Feature Extraction -- Color Representation

- How to compute CCV

  - The initial stage in computing a CCV is similar to the computation of a color histogram. We first blur the image slightly by replacing pixel values with the average value in a small local neighbourhood

  - We then discretize the colour space, such that there are only n distinct colors in the image.

  - To classify the pixels within a given color bucket as either coherent or incoherent. A coherent pixel is part of a large group of pixels of the same color, while an incoherent pixel is not.

  - We determine the pixel groups by computing connected components.

5.6 Low level Feature Extraction -- Color Representation

- How to compute CCV

  - Conduct average filtering on the image
    - To eliminate small variations between neighbor pixels

  - Discretize the image into n distinct colors

  - Classify the pixels within a given color bucket as either coherent or incoherent
    - A pixel is coherent if the size of this connected component exceeds a fixed value τ; otherwise, the pixel is incoherent

  - Obtain CCV by collecting the information of both coherent and incoherent into a vector
    - $CCV = (\alpha_1, \beta_1), (\alpha_2, \beta_2)....(\alpha_m, \beta_m)$ where α and β are the number of coherent pixels and incoherent pixels of the color respectively.

5.6 Low level Feature Extraction -- Color Representation

- How to compute CCV

$\tau = 4$, R=G=B

22	10	21	22	15	16
24	21	13	20	14	17
23	17	38	23	17	16
25	25	22	14	15	14
27	22	12	11	17	18
24	21	10	12	15	19

Blurred Image

1: 10-19
2: 20-29
3: 30-39

2	1	2	2	1	1
2	2	1	2	1	1
2	1	3	2	1	1
2	2	2	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1

Discretized Image

B	C	B	B	A	A
B	B	C	B	A	A
B	C	D	B	A	A
B	B	B	A	A	A
B	B	A	A	A	A
B	B	A	A	A	A

Connected Components

Label	A	B	C	D
Color	1	2	1	3
Size	17	15	3	1

Connected Table

Color	1	2	3
α	17	15	0
β	3	0	1

Color Coherent Vector

Comparison

5.6 Low level Feature Extraction -- Color Representation

- How to compare CCVs
  - Consider two images I and I', together with their CCV's G_I and $G_{I'}$, and let the number of coherent pixels in color bucket i be α_i (for I) and α'_i (for I'). Similarly, let the number of incoherent pixels be β and β'_i. So

  $$G_I = \langle (\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n) \rangle$$

  and

  $$G_{I'} = \langle (\alpha'_1, \beta'_1), \ldots, (\alpha'_n, \beta'_n) \rangle$$

  - Non-normalized

  $$\Delta_G = \sum_{j=1}^{n} \left| (\alpha_j - \alpha'_j) \right| + \left| (\beta_j - \beta'_j) \right|.$$

  - Normalized

  $$Nor\_G = \sum_{i=1}^{n} \left| \frac{\alpha_i - \alpha'_i}{\alpha_i + \alpha'_i + 1} \right| + \left| \frac{\beta_i - \beta'_i}{\beta_i + \beta'_i + 1} \right|$$

5.7 Color-based Image Indexing and Retrieval Techniques

- ## Basic color-based image retrieval

  - ## Color histogram bins

    - For RGB color space, if each color channel M is discretized into 16 levels, the total number of discrete color combinations called histogram bins N.

    - H(M) is a vector $(h_1, h_2, h_3 ... h_n)$, Where each h_i represents the number of pixels in image M falling into bin i

    > M3 = 16x16x16=4096 bins in total

5.7 Color-based Image Indexing and Retrieval Techniques

- Simple histogram distance measure

  - The distance between the histogram of the query image and images in the database are measured

    - Image with a histogram distance smaller than a predefined threshold are retrieved from the database

    - The simplest distance between images I and H is the L-1 metric distance as

$$D(I,H) = sum \ |I-H|$$

5.7 Color-based Image Indexing and Retrieval Techniques

- Example 1
  - Suppose we have three images of 8x8 pixels and each pixel is in one of eight colors C_1 to C_8.
    - Image 1 has 8 pixels in each of the eight colors
    - Image 2 has 7 pixels in each of colors C_1 to C_4 and 9 pixels in each of colors C_5 to C_8
    - Image 3 has 2 pixels in each of colors C1 and C2, and 10 pixels in each of colors C_3 to C_8.

> H_1= (8,8,8,8,8,8,8,8)
> H_2= (7,7,7,7,9,9,9,9)
> H_3= (2,2,10,10,10,10,10,10)
> The distances between these three images
> $D(H_1,H_2)$ =1+1+1+1+1+1+1+1=8
> $D(H_1,H_3)$ = 24
> $D(H_2,H_3)$ = 23

  - Therefore, Images 1 and 2 are most similar

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.7 Color-based Image Indexing and Retrieval Techniques

- Similarity among colors

  - The limitation of using L-1 metric distance is that the similarity between different colors or bins is ignored.

    - If two images with perceptually similar color but with no common color, These two images will have maximum distance according to the simple histogram measure.

    - Users are not only interested in images with exactly same colors as the query, but also in the images with perceptually similar colors. Query on content not on color space !

    - Images may change slightly due to noises and variations on illumination

THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL
ICT AUSTRALIA

5.7 Color-based Image Indexing and Retrieval Techniques

- Similarity among colors
  - The limitation of using L-1 metric distance is that the similarity between different colors or bins is ignored (Cont.).

    - In the simple histogram measure, it might not be able to retrieve perceptually similar images due to these changes

  - Contributions of perceptually similar colors in the similarity calculation

    - Image distance and similarity have an inverse relationship.

    - The similar color measurement is a way to go !

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY·AUSTRALIA

NATIONAL ICT AUSTRALIA

5.7 Color-based Image Indexing and Retrieval Techniques

- Example 2 – Niblack's similarity measurement

X – the query histogram; Y – the histogram of an image in the database
Z – the bin-to-bin similarity histogram

The Similarity between X and Y ➜, $\| Z \| = Z_t A Z$

Where A is a symmetric color similarity matrix with $a(i,j) = 1 - d(c_i,c_j)/d_{max}$

c_i and c_j are the ith and jth color bins in the color histogram

$d(c_i,c_j)$ is the color distance in the mathematical transform to Munsell color space and d_{max} is the maximum distance between any two colors in the color space.

- The similarity matrix A accounts for the perceptual similarity between different pairs of colors.

5.7 Color-based Image Indexing and Retrieval Techniques

- Cumulative histogram distance measure

  - Instead of bin-to-bin distance without considering color similarity, a cumulative histogram of image M is defined in terms of the color histogram H(M):

$$Ch_i = \sum_{j<=i} h_j$$

  The cumulative histogram vector matrix
  CH(M)=(Ch_1,Ch_2….Ch_n)

  - The drawback of this approach is that the cumulative histogram values may not reflect the perceptual color similarity

5.7 Color-based Image Indexing and Retrieval Techniques

- Perceptually weighted histogram (PWH) distance measure

  - Representative colors in the color space are chosen when calculating the PWH.

  - While building a histogram, the 10 perceptually most similar representative colors are found for each pixel.

  - The distance between the pixel and 10 Rep. colors are calculated

5.7 Color-based Image Indexing and Retrieval Techniques

- Other techniques

  - Statistics of color distribution

    - Color regions where pixels are highly populated in the color space are quantized more finely than others.

    - Color coherence vector is one of the types of statistics of color distribution

5.7 Color-based Image Indexing and Retrieval Techniques

- ## Other techniques
  - ### Other color spaces

    - #### RGB color spaces are not perceptually uniform.
      - The calculated distance in a RGB space does not truly reflect perceptual color difference.

    - #### Scalable color descriptor
      - HSV has characteristics to distinguish one color from another

    - #### HMMD (Hue-Max-Min-Diff) histogram
      - The color space is closer to a perceptually uniform color space [2]