Lecture 8: Multimedia Information Retrieval (I)

A/Prof. Jian Zhang

NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2009
jzhang@cse.unsw.edu.au

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

8.0 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.

8.0 Introduction

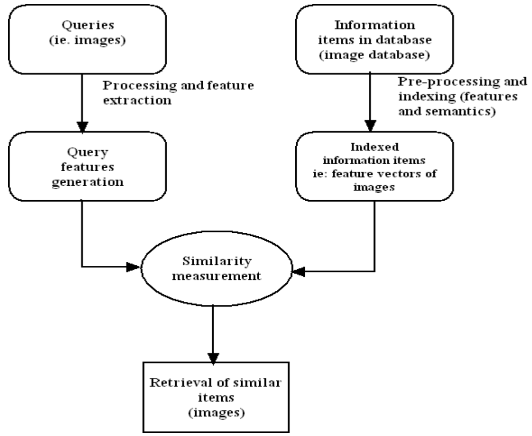
- The fundamental of Multimedia Database (Content) Management research covers:
  - Feature extraction from these multiple media types to support the information retrieval.
  - Feature dimension reduction – High dimensional features
  - Indexing and retrieval techniques for the feature space
    - Similarity 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

8.1 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.

8.1 Multimedia Information Retrieval Systems (MIRS)

- An overview of MIRS operation



8.1 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

8.2 Introduction to Image Indexing and Retrieval

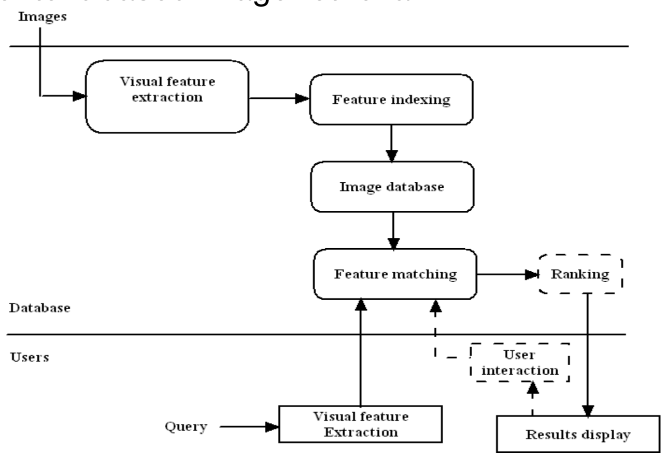
- Four main approaches to image indexing and retrieval
  - Low level features -- Content based Image Retrieval (CBIR)
  - Structured attributes – Traditional database mgt. system
  - Object-recognition – Automatic object recognition
  - Text – Manual annotation (Google search)

8.2 Introduction to Image Indexing and Retrieval

- Four main steps to approaches the 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

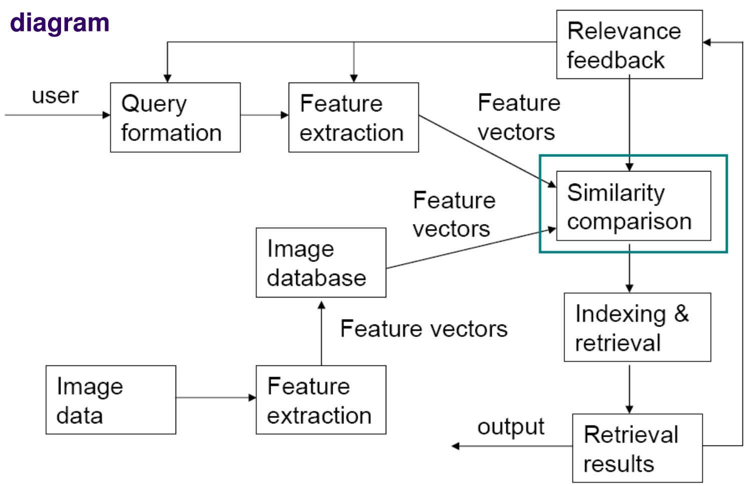
8.2 Introduction to Image Indexing and Retrieval

- Content based image retrieval



8.2 Introduction to Image Indexing and Retrieval

diagram

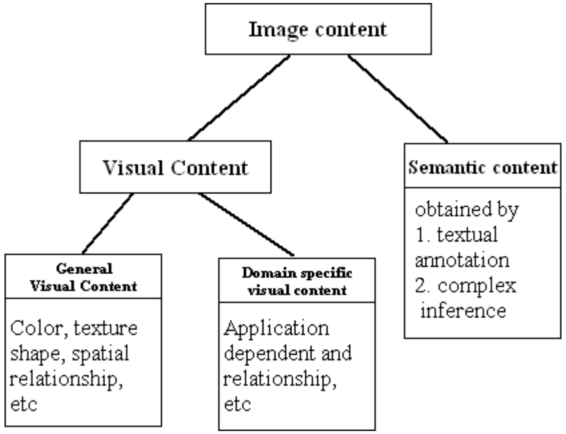


8.2 Introduction to Image Indexing and Retrieval

- 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

8.2 Introduction to Image Indexing and Retrieval

- Image representation



8.3 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

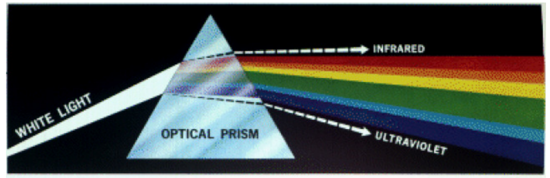


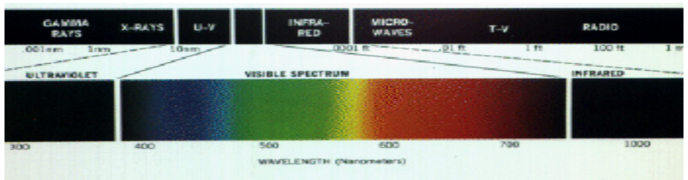
Plate I. Color spectrum seen by passing white light through a prism. (Courtesy of General Electric Co., Lamp Business Division.)

Ref: Gonzalez and Woods, digital image processing

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

8.3 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

8.3 Low level Feature Extraction -- Color Representation

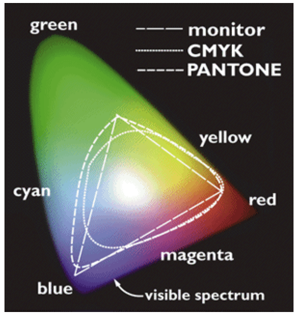
- Color space
  - RGB (Red, Green and Blue) space
    - The RGB color space is the most important means of representing colors used in multimedia.
    - A color can be represented in a form (r-value,g-value,b-value). The value in here is defined as the percentage of the pure light of each primary.

      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.

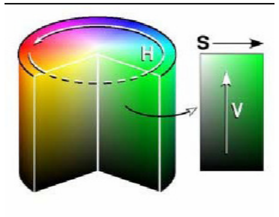
8.3 Low level Feature Extraction -- Color Representation

- 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



8.3 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

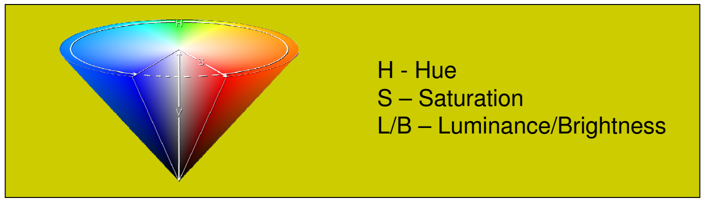


HSV color space as a cylindrical object

  - 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

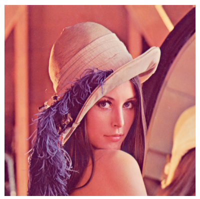
8.3 Low level Feature Extraction -- Color Representation

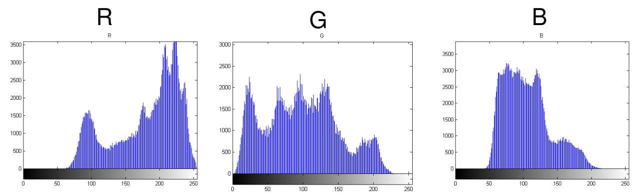
- HSV space
  - Saturation is the colorfulness of an area judged in proportion to its brightness. For example, a pure color has a saturation 100%, while a white color has a saturation 0%.
  - Luminance/Brightness is the attribute of a visual sensation to which an area appears to emit more or less light.



8.3 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 256, 64, 32, 16, 8, 4, 1 (8-bit byte)





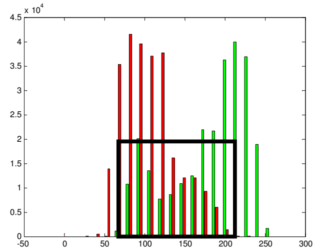
8.3 Low level Feature Extraction -- Color Representation

- 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.

8.3 Low level Feature Extraction -- Color Representation

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

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



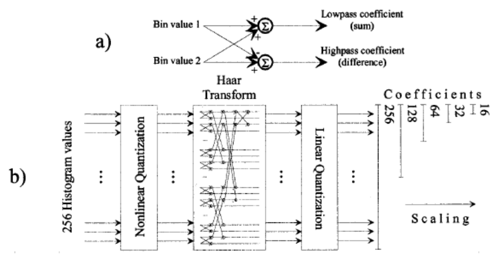
Colors that are not present in the query image do not contribute to the intersection distance

8.3 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.

8.3 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

8.3 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.

8.3 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

8.3 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.

8.3 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.

8.3 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
    - $CCV = (\alpha_1, \beta_1), (\alpha_2, \beta_2) \ldots (\alpha_m, \beta_m)$ where α and β are the number of coherent pixels and incoherent pixels of the color respectively.

8.3 Low level Feature Extraction -- Color Representation

- How to compute CCV

τ= 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

8.3 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

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

8.4 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 $(h_1, h_2, h_3 \ldots h_n)$, Where each h_i represents the number of pixels in image M falling into bin i

M3 = 16x16x16=4096 bins in total

8.4 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|

8.4 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

8.4 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

8.4 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 !

8.4 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.

8.4 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

8.4 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

8.4 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

8.4 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]

8.5 Image Retrieval based on Texture

- Texture
  - Introduction to texture feature
    - The concept of texture is intuitively obvious but has no precise definition
    - Texture can be described by its tone and structure
      - Tone – based on pixel intensity properties
      - Structure – describes spatial relationships of primitives

8.5 Image Retrieval based on Texture

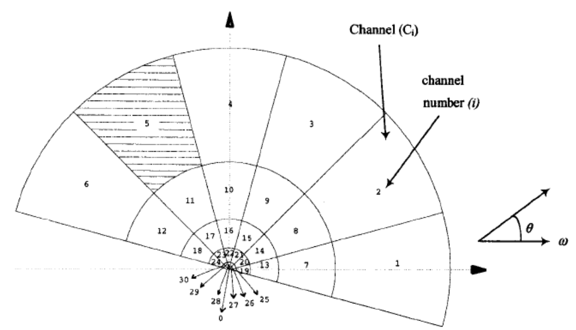
- Texture
  - MPEG-7 standard
    - The homogeneous texture descriptor (HTD). Two components of the HTD will be performed in the whole extraction procedure
      - Mean energy
      - Energy deviation
    - The 2-D frequency plane is partitioned into 30 frequency channels

      The syntax of HTD = [fDC, fSD, e1,e2,…e30,d1,d2,…,d30] .

      where fDC and fSD are the mean and standard deviation of the image respectively,
      Where ei and di are the mean energy and energy deviation that nonlinearly scaled and quantized of the ith channel

8.5 Image Retrieval based on Texture

- Texture



  - The frequency plane partitioning is uniform along the angular direction but not uniform along the radial direction.

8.5 Image Retrieval based on Texture

- Texture
  - Each channel is modeled using Gabor function:
    - If a channel indexed by (*s,r*) where *s* is the radial and *r* is the angular index. Then the (*s,r*)-channel in the freq. domain

$$G_{s,r}(\omega,\theta) = exp\left[\frac{-(\omega-\omega_s)^2}{2\sigma_s^2}\right]\cdot exp\left[\frac{-(\theta-\theta_r)^2}{2\tau_r^2}\right]$$

    - Where σ_s and τ_s are the standard deviation of the Gaussian in the radial direction and the angular direction, respectively

8.5 Image Retrieval based on Texture

- Texture
  - The energy of each channel is defined as the log-scaled sum of the square of the Gabor-filtered Fourier transform coefficients of an image

$$e_i = \log_{10}[1 + p_i]$$

where

$$p_i = \sum_{\omega=0^+}^{1}\sum_{\theta=(0°)^+}^{360°}\left[G_{s,r}(\omega,\theta)|\omega|P(\omega,\theta)\right]^2$$

8.5 Image Retrieval based on Texture

- Texture
  - $P(\omega,\theta)$ the Fourier transform of an image represent in the polar freq. domain $P(\omega,\theta) = F(\omega\cos\theta,\omega\sin\theta)$ where $F(u,v)$ is the Fourier transform in the Cartesian coordinate system
  - The energy deviation of each feature channel is defined as the log-scaled standard deviation of the square of the Gabor-filtered Fourier transform coefficients of an image

$$d_i = log_{10}[1+q_i] \quad \text{where} \quad q_i = \sqrt{\sum_{\omega=0^+}^{1}\sum_{\theta=(0°)^+}^{360°}\left\{\left[G_{s,r}(\omega,\theta)|\omega|P(\omega,\theta)\right] - p_i\right\}^2}$$

  - The HTD consists of the mean and standard deviation of the image intensity, the energy e_i and energy deviation d_i for each feature channel

 where

8.5 Image Retrieval based on Texture

- Texture
  - Demonstrations of using homogeneous texture descriptor for image search
  - Reference – Introduction to MPEG-7

8.5 Image Retrieval based on Texture

- Texture
  - Texture [4] can also be defined as a function of the spatial variation in pixel intensities.
    - One example is to use statistical properties of the spatial distribution of gray-levels of an image. Two types of statistical properties can be used, i.e. (1) first-order statistics and (2) second-order statistics.
      - The first-order statistics measures only depend on the individual pixel gray-levels.
        - *Define* L *-- the number of distinct grey levels*
        - *Define* z *– the random variable denoting the grey-level*
        - *Define* $p(z_i)$ *-- the probability of a grey level occurring in the image*

8.5 Image Retrieval based on Texture

- Texture
  - The first-order statistics measures only depend on the individual pixel gray-levels.
    - *Define* L *-- the number of distinct grey levels*
    - *Define* z *– the random variable denoting the grey-level*
    - *Define* $p(z_i)$ *-- the probability of a grey level occurring in the image*
    - Overall mean: $m = \sum_{i=0}^{L-1} z_i p(z_i)$
    - Overall standard deviation: $\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)}$
    - Skewness: $\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i).$
    - R-Inverse variance: $R = 1 - \frac{1}{1+\sigma^2(z)}$
    - Overall Uniformity: $U = \sum_{i=0}^{L-1} p^2(z_i)$
    - Overall Entropy: $e = -\sum_{i=0}^{L-1} p(z_i)\log_{10} p(z_i)$

8.5 Image Retrieval based on Texture

- The second-order statistics take into account the relationship between the pixel and its neighbors
  - The Grey-level Co-occurrence Matrix (GLCM) is used to calculate the second-order statistics.
  - Suppose the following 4x4 pixel image with 3 distinct grey-levels:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

  - And d = (dx, dy) = (1,0) means that compute the co-occurrences of the pixels to the left of the current one.

8.5 Image Retrieval based on Texture

- The 3x3 co-occurrence matrix is defined as follows. From the table, the element [0,0] in the GLCM matrix is 4. That is the number of counts of pixels with grey-level 0 that have a unit with a gray-level of 0 in the left

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix} \rightarrow$$

Gray-Levels at left	Gray-Levels at current pel: 0	1	2
0	4	0	2
1	2	2	0
2	0	0	2

8.5 Image Retrieval based on Texture

- The Symmetrical GLCM can be computed by adding it to its transpose such as with the position operator (-1,0).
- A GLCM will be then normalized by dividing each individual element by the total count in the matrix giving the co-occurrence probabilities.
- Computing the GLCM over the full 256 gray-level is very expensive and it will also not achieve a good statistical approximation due to a lot of cells with zero values
- A 16 linearly scaled grey-levels is commonly used in CBIR application. The position operation in a CBIR system can be: (1,0), (0,1), (1,1) and (-1,0).

8.5 Image Retrieval based on Texture

- Based on GLCM, the second-order statistics are then computed as follows:
  - ***Angular Second Moment (Energy)*** A measures the homogeneity of the image

$$A = \sum_i \sum_j c_{ij}^2$$

  - ***Entropy*** has the same meaning with one of the first-order statistics but using GLCM instead:

$$\delta = -\sum_i \sum_j c_{ij} \log_2 c_{ij}$$

  - ***Inverse Difference Moment (Homogeneity)*** I is another measure of homogeneity which is sometimes called local homogeneity

$$I = \sum_i \sum_j \frac{c_{ij}}{1+(i-j)^2}$$

8.5 Image Retrieval based on Texture

- Contrast (Inertia) measures how inhomogeneous the image is

$$C = \sum_i \sum_j (i-j)^2 c_{ij}$$

- Correlation cor measures the linear dependency on the pairs of pixels:

$$cor = \frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y)c_{ij}}{\sigma_x \sigma_y}$$

Where $\mu_x = \sum_i [i \sum_j c_{ij}]$ $\quad \mu_y = \sum_j [j \sum_i c_{ij}]$

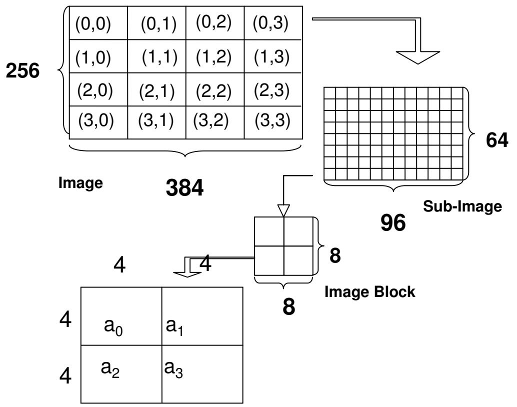
$$\sigma_x = \sum_i [(i-\mu_x)^2 \sum_j c_{ij}] \quad \sigma_y = \sum_j [(j-\mu_y)^2 \sum_i c_{ij}]$$

8.5 Image Retrieval based on Texture

- Local Edge Histograms
  - The edge histogram descriptor (EHD) defined in MPEG-7 represents local edge distribution in the image
    - Specifically, the image is first divided into sub-images.
    - The local-edge distribution for each sub-image can be represented by a histogram.
    - To generate the histogram, edges in the sub-images are categorized into five types:
      - vertical, horizontal, 45 degree diagonal, 135 degree diagonal, non-directional edges and then computed for each sub-images
      - Since there are 16 sub-images, totally 5x16=80 histogram bins are required

8.5 Image Retrieval based on Texture

- Local Edge Histograms



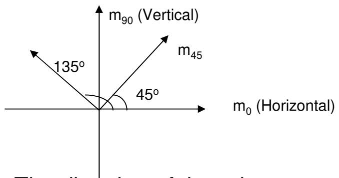
An example for dividing an image into sub-images and 8x8 image blocks

8.5 Image Retrieval based on Texture

- Local Edge Histograms
  - EHD extraction:
    - Each sub-image is first converted to grey-scale levels. The EHD calculation is based on image blocks such as 8x8 pixels.
    - For a 384x256 size of image, 16 sub-images is divided and each sub-image is further divided into 8x8 blocks, the average intensities in the image block are defined as a0, a1, a2 and a3 respectively.
    - The edge direction of a block is determined by calculating the edge magnitudes.

8.5 Image Retrieval based on Texture

- EHD extraction
  - The largest edge magnitude is chosen as the edge direction if the magnitude is larger than the threshold
  - If the magnitude is smaller than the threshold, the block will be decided as containing no-edge and its counts are discarded and not used in computing histograms.
  - The direction of the edge is shown below



The direction of the edge

8.5 Image Retrieval based on Texture

- EHD extraction
  - The edge magnitude can be calculated (digital filtering) as follows

$$m_{90} = a_0 - a_1 + a_2 - a_3 \qquad m_0 = a_0 + a_1 - a_2 - a_3$$

$$m_{45} = \sqrt{2}a_0 - \sqrt{2}a_3 \qquad m_{135} = \sqrt{2}a_1 - \sqrt{2}a_2$$

$$m_{non-directional} = 2a_0 - 2a_1 - 2a_2 + 2a_3$$

  - After calculating the edge magnitude for each image block, 5 histogram columns for this sub-image will be calculated