Lecture 6: Multimedia Information Retrieval

Dr. Jian Zhang
NICTA & CSE UNSW
COMP9314 Advanced Database
S1 2007
jzhang@cse.unsw.edu.au

Reference Papers and Resources
- Papers:
  - Colour spaces-perceptual, historical and applicational background: An overview of colour spaces used in image processing.
  - Colour indexing: using Histogram Intersection for object identification and Histogram Back-projection for object location.
  - Comparing Images Using Color Coherence Vectors: The original paper for CCV.
  - Using Perceptually Weighted Histograms for Colour-based Image Retrieval: The original paper for PWH.
  - The QBIC Project-Querying Images By Content Using Color, Texture, and Shape: The original paper for IBM QBIC project.
- Useful resources
  - IBM QBIC system homepage: http://www.qbic.almaden.ibm.com/
  - UIUC CBIR system homepage: http://www.cs.uiuc.edu/~qitian/MARS.html

6.1 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
- 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,
      \(e_i\) and \(d_i\) are the mean energy and energy deviation that nonlinearly scaled and quantized of the \(i\)th channel.
6.1 Image Retrieval based on Texture

- Texture

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

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

- Where \( \sigma_s \) and \( \tau_r \) are the standard deviation of the Gaussian in the radial direction and the angular direction, respectively.

6.2 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

  \[ p_i = \sum_{s=1}^{S} \sum_{r=1}^{R} [G_{sr}(\omega_s, \theta_r)]^2 |p(\omega_s, \theta_r)|^2 \]

- 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.

  \[ d_i = \log_{10}[1 + q_i] \]

  where

  \[ q_i = \sqrt{\sum_{s=1}^{S} \sum_{r=1}^{R} [G_{sr}(\omega_s, \theta_r)]^2 |p(\omega_s, \theta_r)|^2 - p_i^2} \]

- 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.
6.2 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
    - Overall mean
    - Overall standard deviation
    - Skewness
    - R-Inverse variance
    - Overall Uniformity
    - Overall Entropy

- 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.

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.
6.2 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).

Based on GLCM, the second-order statistics are then computed as follows:

- **Angular Second Moment (Energy)** $A$ measures the homogeneity of the image
  \[ A = \sum \sum c_{ij} \]

- **Entropy** has the same meaning with one of the first-order statistics but using GLCM instead:
  \[ \delta = -\sum \sum c_{ij} \log c_{ij} \]

- **Inverse Difference Moment (Homogeneity)** $I$ is another measure of homogeneity which is sometimes called local homogeneity
  \[ I = \sum \sum \frac{c_{ij}}{1 + (i-j)^2} \]

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.
6.2 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.

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

**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 edge magnitude can be calculated (digital filtering) as follows:

\[
\begin{align*}
    m_{90} &= a_0 - a_1 + a_2 - a_3 \\
    m_{45} &= \sqrt{2} a_0 - \sqrt{2} a_3 \\
    m_{135} &= \sqrt{2} a_1 - \sqrt{2} a_2 \\
    m_{\text{non-directional}} &= 2a_0 - 2a_1 - 2a_2 + 2a_3
\end{align*}
\]

After calculating the edge magnitude for each image block, 5 histogram columns for this sub-image will be calculated.
6.3 Image Indexing and Retrieval based on Shape

- Shape
  - Basic concept on shape
    - The shape of an object or region reflects to its profile and physical structure.
    - A low-level feature – shape of objects within the images
      - For retrieval based on shapes, image must be segmented into individual objects
      - Due to the difficulty of robust and accurate image segmentation, the use of shape features for image retrieval has been limited to special applications where objects or regions are readily available

- Shape Representation
  - Boundary-based methods
    - Chain Codes, fitting line segmentation, Fourier description…
  - Region-based methods
    - Moments, orientation …
  - Geometry-based methods
    - Perimeter measurement, area attribute …
  - Structure-based methods
    - Medial axis transform (MAT) – Skeleton and thinning algorithm
6.3 Image Indexing and Retrieval based on Shape

- **Boundary-based methods -- Chain Code**

![Chain Code Representation Diagram]

6.3 Image Indexing and Retrieval based on Shape

- **Boundary-based methods -- Fourier Descriptors (FDs)**
  - A shape is first represented by a feature function called a shape signature. A discrete Fourier Transform (in frequency domain) is applied to the signature to obtain FD of the shape.
  
  \[
  F_u = \frac{1}{N} \sum_{i=0}^{N-1} f(i) \cdot \exp \left[ -\frac{j 2\pi u i}{N} \right]
  \]

  For \( u = 0 \) to \( N-1 \), Where \( N \) is the number of samples of \( f(i) \).

  - Three commonly used signature:
    - curvature based
    - radius based
    - boundary coordinator based

- **The distance between shapes is calculated as the Euclidean distance between their feature vectors.**
- **Using FDs is to convert the sensitive radius lengths into the frequency domain where the data is more robust to small changes and noise.**
- **The FDs capture the general features and form of the shape instead of each individual detail**
6.3 Image Indexing and Retrieval based on Shape

- Region-based shape representation and similarity measure
  - The shape similarity measurements based on shape representations, in general, do not conform to human perception.
  - The following similarity measurements do not match well with human similarity judgment. They are:
    - Algebraic
    - Spline curve distance
    - Cumulative turning angle
    - Sign of curvature and,
    - Hausdorff-distance

- Basic idea of region-based shape representation
  - As shown in the figure below, if 1 is assigned to the cell with at least 15% of pixels covered by the shape, and a 0 to each of the other cells. The more grids, the more accurate the shape Rep.
  - A binary sequence is created by scanning from left to right and top to bottom – 11100000,11111000,01111110,01111111.

Generation of binary sequence for a shape

- Rotation normalization
  - Rotate the shape so that its major axis is parallel with the x-axis including two possibilities:
  - Only one of the binary sequences is saved while two orientations are accounted for during retrieval time by representing the query shape using two binary sequences

- Scale normalization
  - All shapes are scaled so that their major axes have the same fixed length.
  - Unique shape representation – shape index
  - After rotation and scale normalization and selection of a grid cell size, a unique binary sequence for each shape based on a unique major axis.
  - This binary sequence is used as an index of the shape
  - When the cell size is decided, the number of grid cells in the x direction is fixed (i.e 8), The number of cells in the y direction depends on the eccentricity of the shape. The cell number for Y can range from 1 to 8.
6.3 Image Indexing and Retrieval based on Shape

- Similarity measure between two shapes based on their indexes
  - Based on the shape eccentricities, there are three cases for similarity measurement
    - Same basic rectangle of two normalized shapes: bitwise compare and distance calculation between the shape point position values, For example:
      - A and B have the same eccentricity of 4
        - A = 11111111 11100000 and B= 11111111 1111100, then the distance value between A and B is 3
    - If two normalized shape have very different basic rectangles, we can assume these two shapes are quite different (i.e. different on Minor Axis)

- If two normalized shapes have slightly different basic rectangles, the perceptual similarity is still possible.
  - Add the 0s at the end of the index of the shape with shorter minor axis to extend the index to the same length as the other shape
  - Example:
    - A = (2, 11111111 11110000) ,and
    - B = (3, 11111111 11111000 11100000), then the shape A binary number is extended to the same length of B. Hence A = (3, 11111111 11110000 00000000). The distance of A and B is 4

6.4 Data Structure for Efficient Multimedia Similarity Search

- Introduction
  - The retrieval is based on the similarity between the query vector and the feature vector
  - If the feature dimensions high and the number of stored objects are huge, it will be too slow to do the linearly search for all features vectors
  - Techniques and data structures are required to re-organize feature vectors and develop fast search method to locate the relevant features quickly
  - The main idea is to divide the high dimension feature vector space into many sub-space and focus on one or a few sub-spaces for effective search

- Three common queries:
  - Point query – users’ query is represented as a vector
    - Feature vectors exactly match
  - Range query – users’ query is represented as a feature vector and distance range
    - The distance metrics – i.e. L1 and L2 (Euclidean distance)
  - The k nearest neighbours query – users’ query is specified by a vector and a integer k.
    - The k objects whose distances from the query are the smallest are retrieved.
6.4 Data Structure for Efficient Multimedia Similarity Search -- Filtering Process

- Query methods based on color-histogram
  - Use histograms with very few bins to select potential retrieval candidates
  - Then use the full histograms to calculate the distance
  - For a special case, calculate the average of RGB value such as \( \vec{x} = (R_{avg}, G_{avg}, B_{avg}) \)
  - Given the average color vectors \( \vec{x} \) and \( \vec{y} \) of two images. The Euclidean distance:
    \[ d_{avg}(\vec{x}, \vec{y}) = \sqrt{\sum_{i=0}^{2} (x_i - y_i)^2} \]

6.5 Data Structure for Efficient Multimedia Similarity Search -- B+ Tree

- To achieve an efficient way for query process
  - The weakness of traditional similarity calculation on feature vectors within search space is sequential
  - A B+ tree is a hierarchical structure with a number of nodes to store the feature vectors

6.6 Similarity Comparison

- Given two feature vectors, I, J, the distance is defined as \( D(I,J) = f(I,J) \)
- Typical similarity metrics
  - \( L_p \) (Minkowski distance)
  - \( \chi^2 \) metric
  - KL (Kullback-Leibler Divergence)
  - JD (Jeffrey Divergence)
  - QF (Quadratic Form)
  - EMD (Earth Mover’s Distance)