Lecture 9: Multimedia Information Retrieval

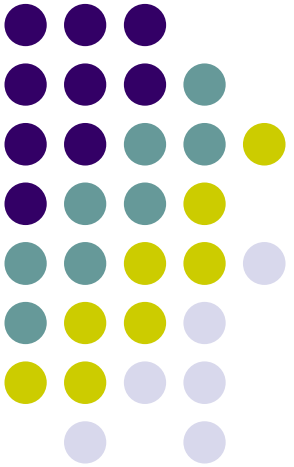
A/Prof. Jian Zhang

NICTA & CSE UNSW

COMP9519 Multimedia Systems

S2 2009

jzhang@cse.unsw.edu.au



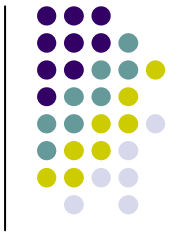
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY · AUSTRALIA

NICTA

Reference Papers and Resources

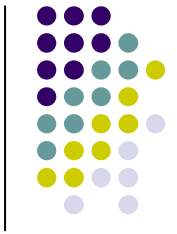
- Papers:
  - Colour spaces-perceptual, historical and applicational background: An overview of colour spaces used in image processing.
  - Colour indexing: using Histogram Intersection for object identification and Histogram Back-projection for object location.
  - Comparing Images Using Color Coherence 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
  - MPEG-7 homepage: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
  - IBM QBIC system homepage: http://wwwqbic.almaden.ibm.com/
  - UIUC CBIR system homepage: http://www.ifp.uiuc.edu/~qitian/MARS.html

9.1 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

9.1 Image Indexing and Retrieval based on Shape



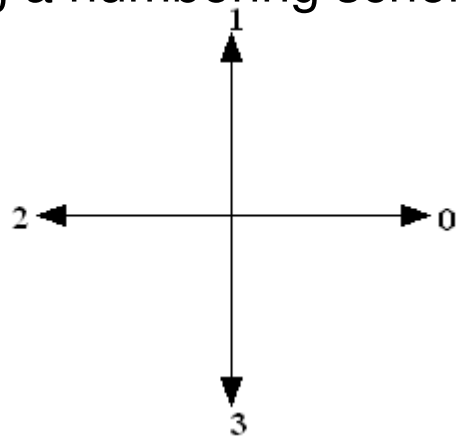
- Shape
  - Basic concept on shape
    - A good shape representation and similarity measurement for recognition and retrieval purposes should have the following two important properties:
      - Each shape should have a unique representation, invariant to translation, rotation and scale;
      - Similar shapes should have similar representations so that retrieval can be based on distance among shape representation

9.1 Image Indexing and Retrieval based on Shape

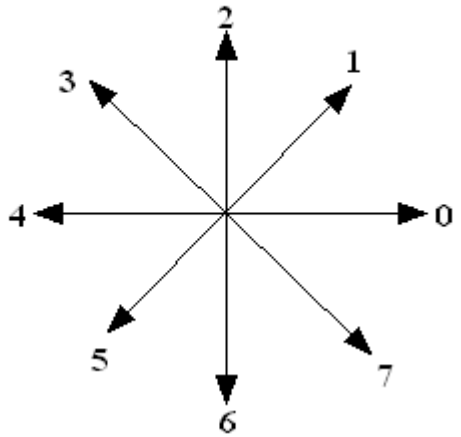
- Shape Representation
  - Boundary-based methods
    - Chain Codes, fitting line segmentation, Fourier description…
  - Region-based methods
    - Moments, orientation …
  - Geometry-based methods
    - Perimeter measurement, area attribute …
  - Structure-based methods
    - Medial axis transform (MAT)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

9.1 Image Indexing and Retrieval based on Shape

- Boundary-based methods -- Chain Code
  - Chain codes are used to represent a boundary by a connected sequence of straight-line segments of special length and direction
  - Typically, this representation is based on 4- or 8-connectivity of the segments. The direction of each segment is coded by using a numbering scheme



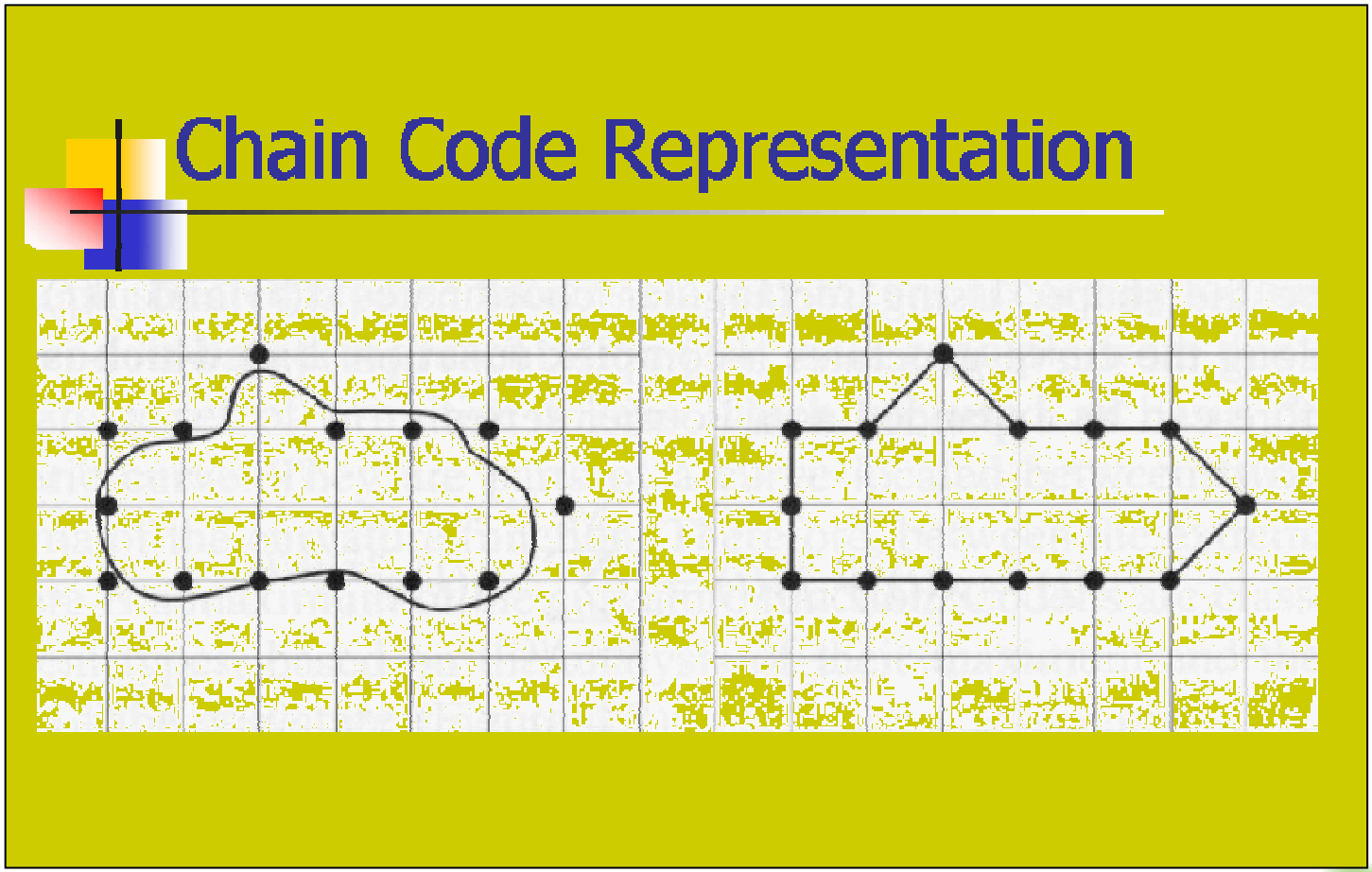
Direction numbers for 4-directional chain code



Direction numbers for 8-directional chain code

9.1 Image Indexing and Retrieval based on Shape

- Boundary-based methods -- Chain Code



9.1 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.

$$F_n = \frac{1}{N} \sum_{i=0}^{N-1} f(i) \cdot exp\left[\frac{-j2\pi ui}{N}\right]$$

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

9.1 Image Indexing and Retrieval based on Shape

- Boundary-based methods -- Fourier Descriptors (FDs)
  - The Radius-based signature – consists of a number of ordered distance from the shape centroid to boundary points (called radii). The radii are defined as

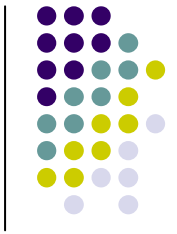
$$r_i = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2}$$

Where (x_c, y_c) *are the coordinates of the centroid and* (x_i, y_i) *for i=0 to 63 are the coordinates of the 64 sample points along the shape boundary and the number of pixels between each two neighboring points is the same*

  - A feature vector which is invariant to start point (p), rotation (r) and scale (s) should be calculated.

$$x = \left[\frac{|F_1|}{|F_0|}, \ldots \frac{|F_{63}|}{|F_0|} \right]$$

9.1 Image Indexing and Retrieval based on Shape



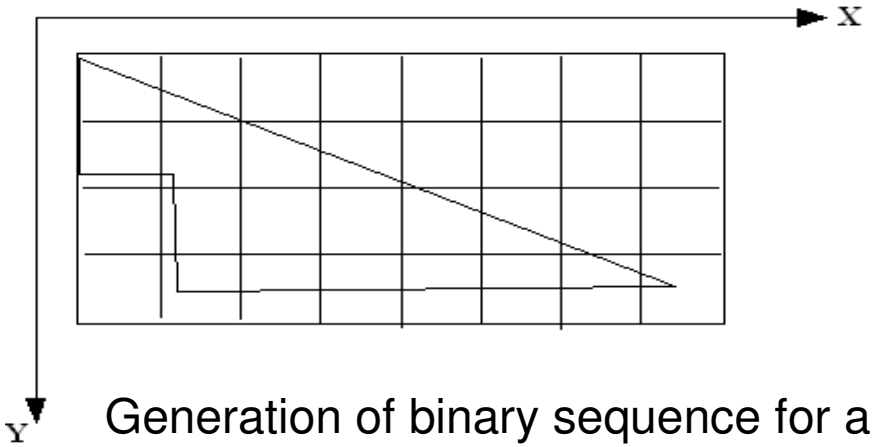
- Boundary-based methods -- Fourier Descriptors (FDs)
  - The distance between shapes is calculated as the Euclidean distance between their feature vectors.
  - Using FDs is to convert the sensitive radius lengths into the frequency domain where the data is more robust to small changes and noise.
  - The FDs capture the general features and form of the shape instead of each individual detail

9.1 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

9.1 Image Indexing and Retrieval based on Shape

- Region-based shape representation and similarity measure
  - 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.

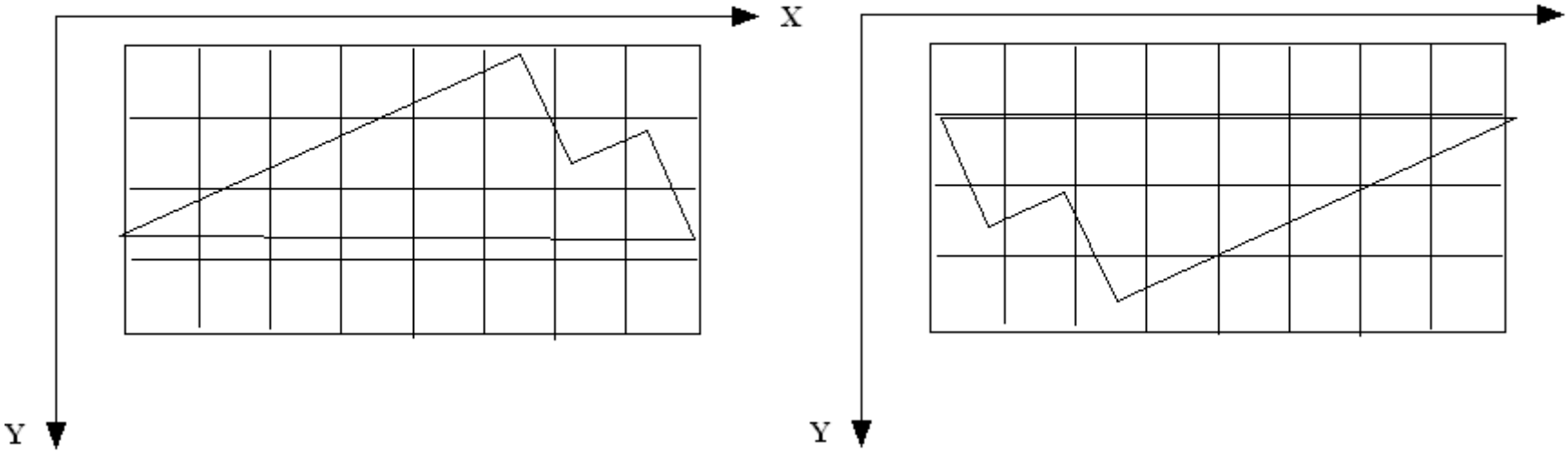


Generation of binary sequence for a shape

9.1 Image Indexing and Retrieval based on Shape

- Rotation normalization
  - Rotate the shape so that its major axis is parallel with the x-axis including two possibilities:

Two possible orientations with the major axis along the x direction



  - 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

9.1 Image Indexing and Retrieval based on Shape

- 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 a 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.

9.1 Image Indexing and Retrieval based on Shape

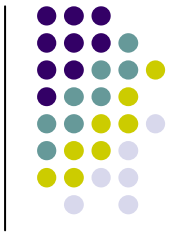
- Similarity measure between two shapes based on their indexes
  - Based on the shape eccentricities, there are three cases for similarity measurement
    - Same basic rectangle of two normalized shapes: bitwise compare and distance calculation between the shape point position values, For example:
      - 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)

9.1 Image Indexing and Retrieval based on Shape

- 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

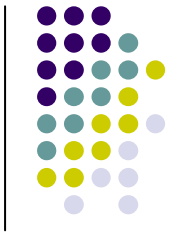
    B = (3, 11111111 11111000 11100000), then the shape A binary number is extended to the same length of B. Hence A = (3, 11111111 11110000 00000000). The distance of A and B is 4

9.2 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

9.2 Data Structure for Efficient Multimedia Similarity 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.

9.2 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 $\bar{x} = (R_{avg}, G_{avg}, B_{avg})^T$

$$A_{avg} = \frac{\sum_{p=1}^{p} A(p)}{p}$$ where A = {R,G, B}

  - Given the average color vectors $\bar{x}$ and $\bar{y}$ of two images. The Euclidean distance: $$d_{avg}(\bar{x}, \bar{y}) = \sqrt{\sum_{i=1}^{3} (\bar{x}_i - \bar{y}_i)^2}$$

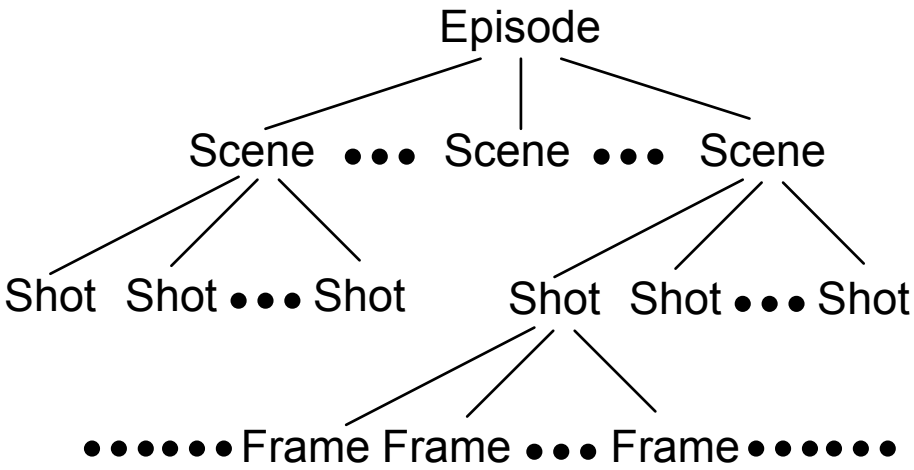
9.3 Video Indexing and Retrieval Introduction

- Video is information rich, a combination of metadata, text, audio, and images with a time dimension.
- Video indexing and retrieval methods:
  - Metadata-based method
    - video title, author/producer/director, date-of-production …
  - Text-based method
    - transcripts and subtitles
  - Audio-based method
    - audio is segmented into speech and non-speech groups

      speech signal → spoken words; non-speech signal → sound effect
  - Content-based method
    - individual frame or image indexing and retrieval
    - shot-based indexing and retrieval
  - Integrated approach
    - combine two or more of the above methods

9.3 Shot-based Video Indexing and Retrieval

- Video is normally made of a number of logical units or segments, which are called *video shots*:
  - The frames depict the same scene
  - The frames signify a single camera operation
  - The frames contain a distinct event or an action
  - The frames are chosen as a single indexable entity by user



A general video model

9.3 Shot-based Video Indexing and Retrieval

- Shot-based video indexing and retrieval consists of the following main steps:
  - Segment the video into shots
    - commonly called video temporal segmentation, partition, or shot detection
  - Index each shot
    - identify key frames or representative frames (*r frames*)
    - index r frames
  - Retrieve shots
    - based on indexes or feature vector similarity measurement

9.3 Video Shot Detection or Segmentation

- Consecutive frames on either side of a shot transition generally display a significant quantitative change in content
  - Automatic partitioning requires suitable difference metrics and the techniques for applying them
- Shot transition pattern
  - Simplest transition: a *camera break*
  - More sophisticated camera operations: *dissolve, wipe, fade in, fade out*, etc
    - much more gradual changes than in a camera break
    - requires more sophisticated approach other than a single threshold

9.3.1 Basic Video Segmentation Techniques

- Simple measure: pixel-to-pixel differences between neighboring frames (Similarity Distance –SD)
  - Problem: object movement causes many false detections

$$SD_i = \sum_{x,y} \left| I_i(x,y) - I_{i+1}(x,y) \right|$$

- Color histogram distance
  - To reduce computation, usually only two or three most significant bits of each color component are used to compose a color code.

$$SD_i = \sum_{j} \left| H_i(j) - H_{i+1}(j) \right|$$

- χ^2 test

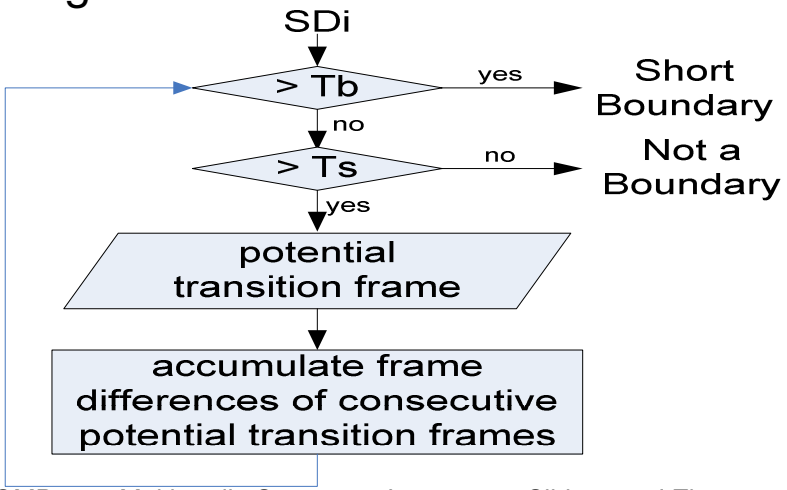
$$SD_i = \sum_{j} \frac{(H_i(j) - H_{i+1}(j))^2}{H_{i+1}(j)}$$

9.3 Similarity Comparison

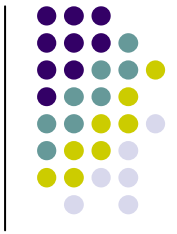
- Given two feature vectors, I, J, the distance is defined as D(I,J) = f(I,J)
- Typical similarity metrics
  - Lp (Minkowski distance)
  - X2 metric
  - KL (Kullback-Leibler Divergence)
  - JD (Jeffrey Divergence)
  - QF (Quadratic Form)
  - EMD (Earth Mover's Distance)

9.3.2 Detecting Shot Boundaries with Gradual Change

- The basic single threshold technique is not enough
  - Cannot detect shot boundaries with gradual change
  - Cannot recognize a boundary between frames of two different scenes but with similar color histograms
- *Twin-comparison technique* to rescue
  - Double thresholds
    - T_b for normal camera break detection
    - T_s for potential gradual transition detection



9.3.3 Preventing False Shot Detection



- Sources of false detection
  - Camera motion
    - camera panning and zooming → gradual changes
      → falsely interpreted as segment boundaries
    - solution: remove false alarm by *motion analysis*
      panning → a single strong modal motion vectors distribution
      zooming → motion vectors converge or diverge at focus center
  - Illumination change
    - a cloud moving across the sky, an actor walking into a spotlight, special lighting effects, etc
    - solution: *normalization*

9.3.3 Preventing False Shot Detection

1. Normalize each of the R, G and B channels of each frame separately

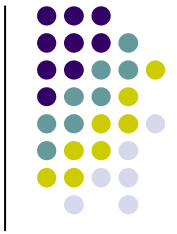
$$R_i^{'} = R_i \Big/ \sqrt{\sum_{i=1}^{N} R_i^2} \qquad G_i^{'} = G_i \Big/ \sqrt{\sum_{i=1}^{N} G_i^2} \qquad B_i^{'} = B_i \Big/ \sqrt{\sum_{i=1}^{N} B_i^2}$$

2. Convert pixel values into chromaticity

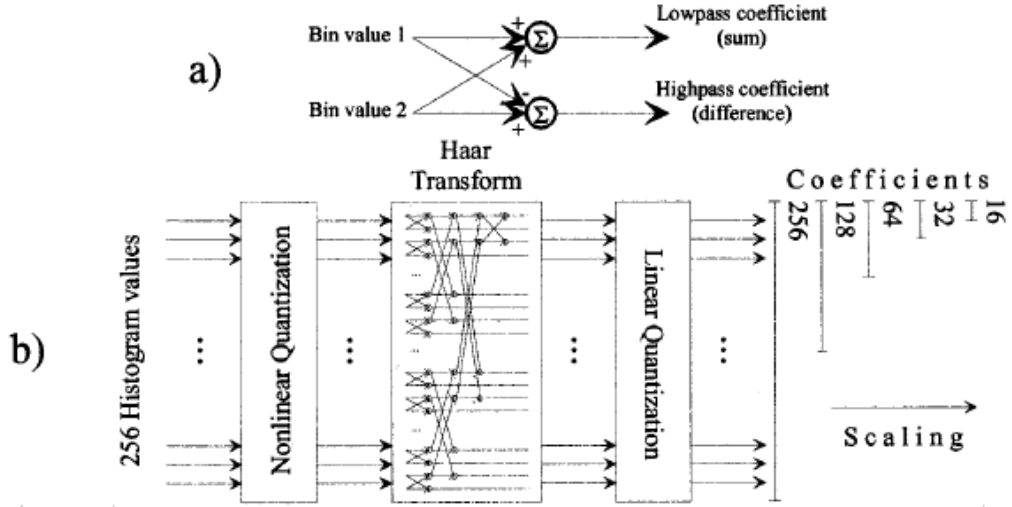
$$r_i = R_i^{'} \Big/ (R_i^{'} + G_i^{'} + B_i^{'}) \qquad g_i = G_i^{'} \Big/ (R_i^{'} + G_i^{'} + B_i^{'})$$

3. Build a combined histogram for *r* and *g*, called a *chromaticity histogram image (CHI)*
4. The resolution of each CHI is reduced to 16x16 using wavelet-based compression technique
5. A two-dimensional DCT is applied to the reduced CHI to obtain 256 DCT coefficients
6. Only 36 significant DCT coefficients are selected
7. Detect shot transition based on the distances calculated from the corresponding 36 coefficients between frames

9.3.3 Low level Feature Extraction -- Color Representation (Ref. to lecture note 8.3)



- Scalable color descriptor - wavelet-based compression technique



- 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

9.3.4 Other Shot Detection Techniques

- Ideal distribution for frame-to-frame distances
  - Close to zero with little variation within a shot
  - Significantly larger than zero between shots
- In common videos, object and camera motion and other changes introduce complexity
  - Pre-filtering to remove the effects of object and camera motion
  - Edge-based shot detection
    - compute the percentage of edges that enter and exit between the two frames, percentage > threshold →shot boundary
    - dissolves and fades are identified by the relative values of entering and exiting edge percentages
  - Advanced cameras provide extra information
    - position, time, orientation, etc

9.3.5 Segmentation of Compressed Video

- Most videos are in compressed form, indexing directly on compressed data would be advantageous
- Indexing on MPEG compressed video
  - DCT coefficients based
    - A DC image is formed by combining the DC coefficients of each image block
    - The DC image is 64 times smaller than the original image
    - Frame-to-frame distance on DC images
  - Motion information based
    - Determine camera operations based on directional information of motion vectors
    - Perform shot detection based on the number of bidirectionally coded macroblocks in B frames
    - Cannot determine the exact location of the shot boundary

9.3.5 Segmentation of Compressed Video

- Indexing on VQ compressed video
  - SD_i – frame-to-frame codevector histogram distance
    - $SD_i > T_1$ →declare a shot boundary
  - SD_j – distance between the current frame and the first frame of the current shot
    - $SD_j > T_2$ → declare a gradual transition
  - It is reported VQ-based segmentation method has higher shot detection performance than methods based on DCT coefficients

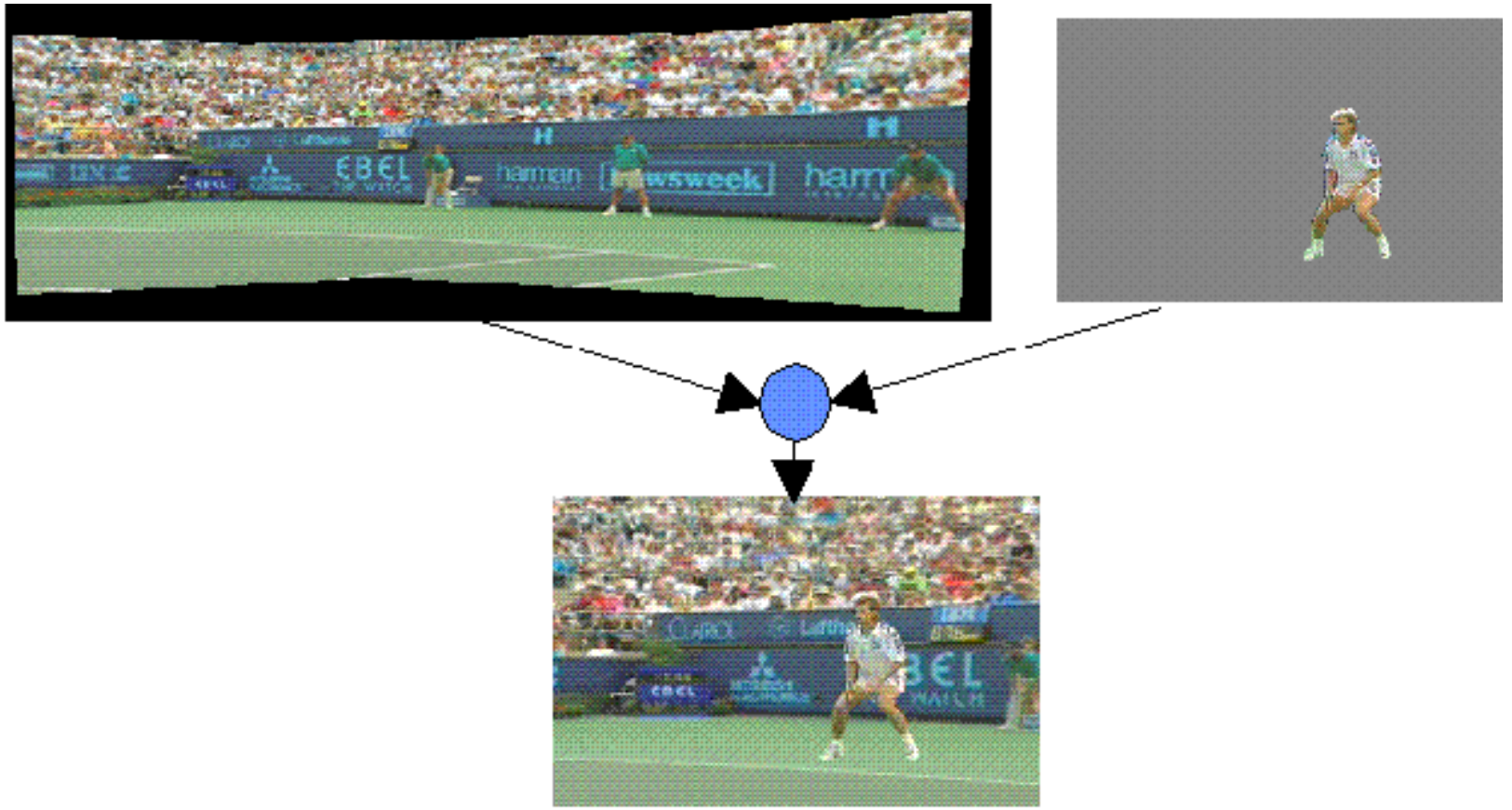
9.4.1 Indexing and Retrieval based on R Frames

- Indexing and retrieval based on Reference (r) frames of video shots
  - How many r frames should be used in a shot
  - How to select these r frames within a shot
- Determine the number of r frames
  - Use one r frame per shot
    - does not consider the length and content changes of shots
  - Assign the number of r frames according to shot length
    - shot length <= 1s → one r frame for the shot
    - shot length > 1s → one r frame for each second of video
  - Divide into subshots and assign one r frame to each subshot
    - subshots are detected based on changes in contents, such as motion vectors, optical flow, frame-to-frame distance, etc

9.4.1 Indexing and Retrieval based on R Frames

- Select r frames
  - One r frame per shot
    - the first frame is normally used as the r frame
  - One r frame per second/subshot
    - compute the average frame of all the frames in the segment, and the frame most similar to the average frame is selected as the r frame
    - or compute the average histogram of all the frames in the segment, and the frame whose histogram is closest to this average histogram is selected as the r frame
  - Image mosaic as r frame
    - each frame is divided into background and foreground objects
    - a large background is constructed from the background of all frames with main foreground objects superimposed

9.4.1 Indexing and Retrieval based on R Frames



example of image mosaic

9.4.1 Indexing and Retrieval based on r Frames

- Combined approach of determining the number of r frames and selecting them
  - The first frame of each shot is automatically an r frame
  - Each of the subsequent frames in the shot are compared with the previous r frame, if distance > threshold, the frame is marked as a new r frame
  - Problem: the final number of r frames is unpredictable
  - Solution: set an upper limit of r frames and the number of r frames are assigned to shots proportional to their amount of content
- In general, the choice of r frame selection method is application dependent

9.4.2 Indexing and Retrieval based on motion information

- R frame indexing ignores temporal or motion information
- Motion information is normally derived from optical flow or motion vectors
  - Motion content
    - a measure of total amount of motion within a video
    - a talking head video → small motion content

      a explosion or car crash → high motion content
  - Motion uniformity
    - a measure of the smoothness of the motion within a video
    - a smooth panning shot → high motion smoothness

      a staggered panning → low motion smoothness

9.4.2 Indexing and Retrieval based on motion information

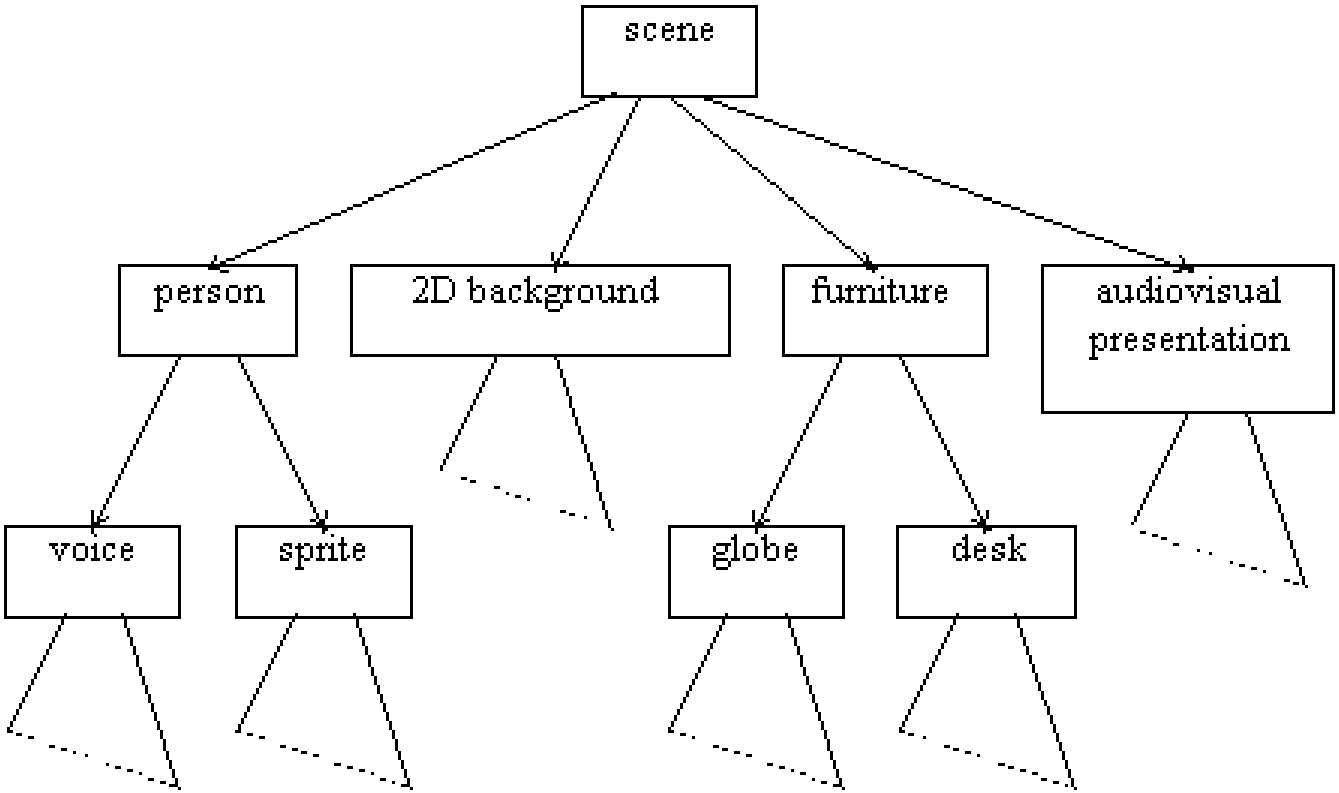
- Motion information is normally derived from optical flow or motion vectors
  - Motion panning
    - a measure of the panning motion
    - a pan shot → high motion panning

      a zoom shot → low motion panning
  - Motion tilting
    - a measure of vertical motion
- Associate motion information with shots
  - A fixed number of pairs of subimages or windows is decided for all r frames
  - Two bits are used to store motion for each window pair
    - e.g. 01 → motion in the second window and no motion in the first window

9.4.3 Indexing and Retrieval based on objects

- Drawback of shot based indexing and retrieval
  - Does not lend itself to content based representation
    - the content can drastically change within a single shot, or it might stay virtually constant over a series of successive shots
- Determine content change is the key question
  - A scene is a complex collection of parts or objects
    - the location and physical qualities of each object, and their interaction with others
  - If we could identify and track individual objects throughout the sequence, then we can index and retrieve based on information about each object

9.4.3 Indexing and Retrieval based on objects



Logical structure of a scene

9.4.3 Indexing and Retrieval based on objects

- In a still image, object segmentation and identification is normally difficult, but it is more tangible in a video sequence

(a) Image from an underground station

(b) Identified and tracked people

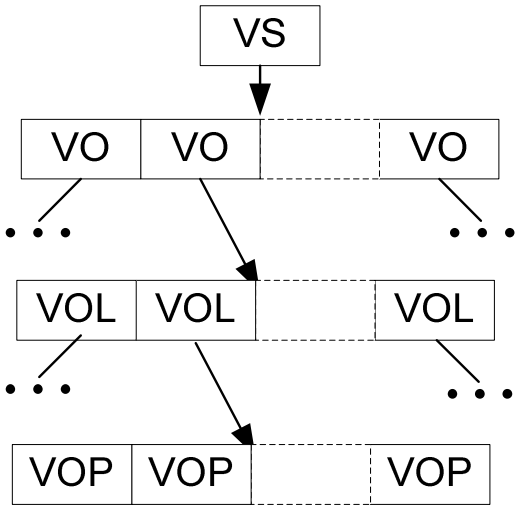
9.4.3 Indexing and Retrieval based on objects

- MPEG-4 object-based coding standard
  - Lend itself directly to object-based indexing and retrieval
  - VS – Video Session

    VO – Video Object

    VOL – Video Object Layer

    VOP – Video Object Plane
    - VOP is a semantic object containing shape and motion information



9.4.3 Indexing and Retrieval based on Metadata

- Metadata for video is available in some standard video formats. Indexing and retrieval can be based on Metadata using conventional DBMS.
  - *Program Specific Information (PSI)* of MPEG-2
    - four tables that contain information to demultiplex and present program via a transport steam
    - copyright and language can be obtained from *program map table*
  - *Service Information Table* of DVD *(DVD-SI)*
    - DVD-SI is organized into six tables
    - the most useful for video indexing are *service description table* and *event information table*, which contain items such as a title, video type, and directors.

9.4.3 Indexing and Retrieval based on Annotation

- Annotation is important as it can capture the high level contents of video
  - Interpret and annotate video manually
    - time-consuming but is still widely used because high-level video understanding is currently not possible for general video
    - to simplify manual annotation process
      - provide a well-defined framework for manual entry
      - make use of domain knowledge of specific types of video
  - Directly use associated transcripts and subtitles
  - Speech recognition to help
    - extract spoken words for indexing and retrieval
    - still very challenging because speech and nonspeech are normally mixed, and there is background music and noise

9.4.5 Effective Video Representation and Abstraction

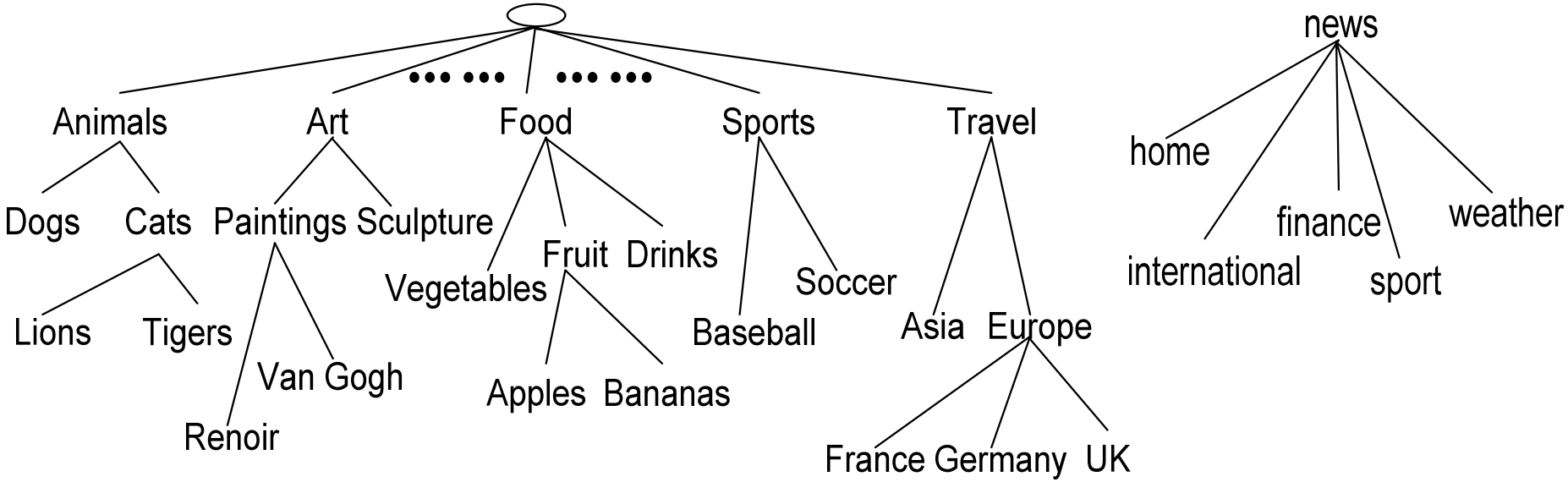
- Video sequences are rich in information, large in storage requirements, and have a time dimension
- To represent video content compactly, effective video representation and abstraction tools are needed
  - Video browsing
    - locate a relevant video segment
    - traditional video operations (e.g. fast forward) is sequential
  - Video retrieval results presentation
    - allows the results to be displayed in a limited display window
  - Reduce network bandwidth requirements and delay
    - compact representation is normally many times smaller than the video itself
    - makes quick browsing possible and reduce network bandwidth and delay

9.5.1 Topical or Subject Classification

- Subject classification is one of the most effective ways to organize large amount of information
  - Has been proven in traditional library systems and many on-line search engines, such as Yahoo and Infoseek
- Subject classification allows the combination of two powerful information retrieval approaches: browsing and search
- For video organization, two levels of topical or subject classification are used
  - First level classification divides different videos into different subject classes
  - The second level classification divides the shots of each individual video into different subclasses or topics

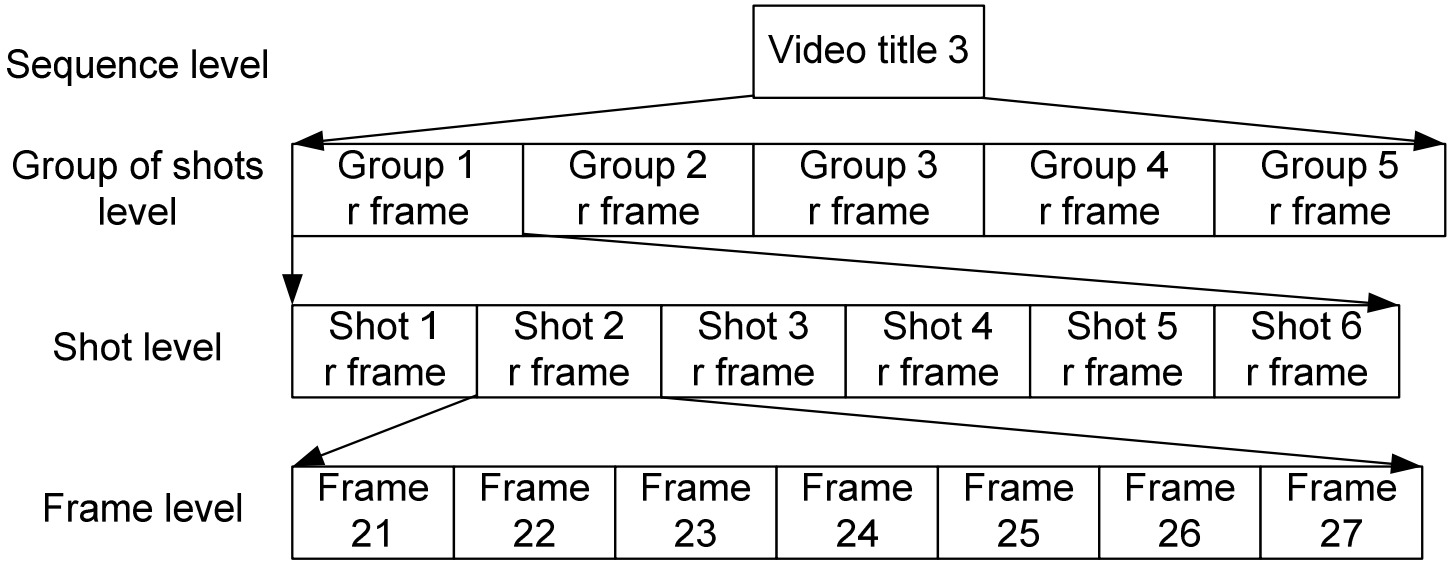
9.5.1 Topical or Subject Classification

- Most WWW search engines use the first level subject classification
- Many videos are very well structured based on topic
  - news programs, movies, etc



9.5.2 Hierarchical Video Browser

- To be able to browse a video sequence efficiently is important
- A hierarchical video browser consists of a number of levels; Representative frames are displayed at each level



9.6 Measurement of Retrieval Effectiveness

- Efficiency
  - Concerned with response time of the system
  - Determined by data structure used for organizing feature vectors
- Effectiveness
  - Concerned with presentation quality
  - Concerned with the system's ability to retrieve relevant items and discard irrelevant items
  - No need for traditional DBMSs, which are based on exact match
  - For multimedia database, it is difficult to achieve a good effectiveness measurement
    - It is difficult to design a similarity metric that exactly conforms to human perception
    - different levels of similarity or relevance between information items
    - Multimedia items are information rich → hard to judge relevance between them

9.6.1 Collection of Human Judgment Data

- As perceptual relevance varies from person to person, relevance data must be collected from a large subjective test group
- Three common methods
  - Method 1 (counting)
    - $C_{i,j} = 0, \quad \forall i, j$
    - For query i, if item j from database is selected as relevant, then $C_{i,j} = C_{i,j} + 1$
    - For query i, if $C_{i,j} > T$, then $R_{i,j} = 1$, else $R_{i,j} = 0$
  - Method 2 (weighting)
    - $W_{i,j} = 0, \quad \forall i, j$
    - For query i, if item j is selected as relevant, then $W_{i,j} = W_{i,j} + 1$
  - Method 3 (ranking)
    - $Q_j(i, k) = 0, \quad \forall i, j, k$
    - For query i, if item j is ranked as the k^{th} relevant, then $Q_j(i, k) = Q_j(i, k)+1$

9.6.2 Effectiveness Measurement Methods

- *Recall And Precision Pair (RPP)*
  - Use the results of subjective test Method 1
  - Recall measures the ability to retrieve relevant items;
    Precision measures the ability to reject irrelevant items
  - Recall and precision must be used together;
    A good system should have both high recall and precision
  - Problem
    - does not consider the different degrees of relevance and thus may not reflect human judgment accurately

$$recall = \frac{number\ \ of\ \ relevant\ \ items\ \ retrieved}{total\ \ number\ \ of\ \ relevant\ \ items\ \ in\ \ database}$$

$$precision = \frac{number\ \ of\ \ relevant\ \ items\ \ retrieved}{total\ \ number\ \ of\ \ retrieved\ \ items}$$

9.6.2 Effectiveness Measurement Methods

- *Percentage of Weighted Hits (PWH)*
  - Use the results of subjective test Method 2
  - Similar to recall
  - To indicate average performance, perform many queries (M2) and average PWH over these queries
  - Problem
    - assume a fixed number of return items, while different queries may have different numbers of relevant items
    - does not measure the ability to reject irrelevant items

$$P = \sum_{i=1}^{n} w_i \Big/ \sum_{j=1}^{N} w_j$$

where n is the number of items returned,
N is the total number of items in the database

9.6.2 Effectiveness Measurement Methods

- *Modified RPP*
  - Each item has a weight for each query, as in PWH
    - the weight is equal to the number of people selecting the item as relevant
  - Measures retrieval effectiveness more accurately because it takes into account the different degrees of relevance in calculating the ability to retrieve relevant items

$$recall = \frac{sum\ \ of\ \ the\ \ weights\ \ of\ \ retrieved\ \ items}{sum\ \ of\ \ the\ \ weights\ \ of\ \ all\ \ the\ \ items\ \ in\ \ database}$$

$$precision = \frac{number\ \ of\ \ relevant\ \ items\ \ retrieved}{total\ \ number\ \ of\ \ retrieved\ \ items}$$

9.6.2 Effectiveness Measurement Methods

- *Percentage of Similarity Rankings (PSR)*
  - Use the results of subjective test Method 3
  - From $Q_j(i, k)$, calculate
    - mean value $\overline{p}_j(i)$ - representing the average ranking of the i[th] image for query j
    - standard deviation $\delta_j(i)$ - degree of disagreement of the ranking among subjects
  - Percentage of similarity rankings $S_j(i)$

$$S_j(i) = \sum_{k=P_j(i)-\left\lceil \frac{\delta_j(i)}{2} \right\rceil}^{P_j(i)+\left\lceil \frac{\delta_j(i)}{2} \right\rceil} Q_j(i,k)$$

    $S_j(i)$ represents the degree of agreement between $P_j(i)$ and $\overline{p}_j(i)$
    - $P_j(i) = \overline{p}_j(i)$ → $S_j(i)$ is close to 1
    - $\left|P_j(i) - \overline{p}_j(i)\right| > \delta_j(i)$ → $S_j(i)$ is close to 0
  - In practice, we are interested in a number of highest ranked items
    - plot $S_j(i)$ as a function of rankings $P_j(i)$ (being 1, 2, 3 ……), which shows the degree of agreement between the subjects and the system in ranking the i[th] item in the $P_j(i)$ position

9.6.3 Factors Affecting Retrieval Effectiveness

- Three main factors that determine the retrieval effectiveness of a system
  - Features chosen to represent multimedia information items
    - e.g. color, shape, texture or a combination of these features
  - How to represent or describe the chosen features
    - e.g. for color there are many different color spaces and histogram representations
  - Distance or similarity metric
    - e.g. L1 and L2 norm