Πολυμεσικές ΒΔ

Multimedia Databases: Where are we?

Βασική πηγή διαφανειών: Vincent Oria, Department of Computer Science,
New Jersey Institute of Technology

Εργαστήριο Πληροφοριακών Συστημάτων, Παν/μιο Πειραιώς (http://infolab.cs.unipi.gr/)

έκδοση: Ιανουάριος 2010

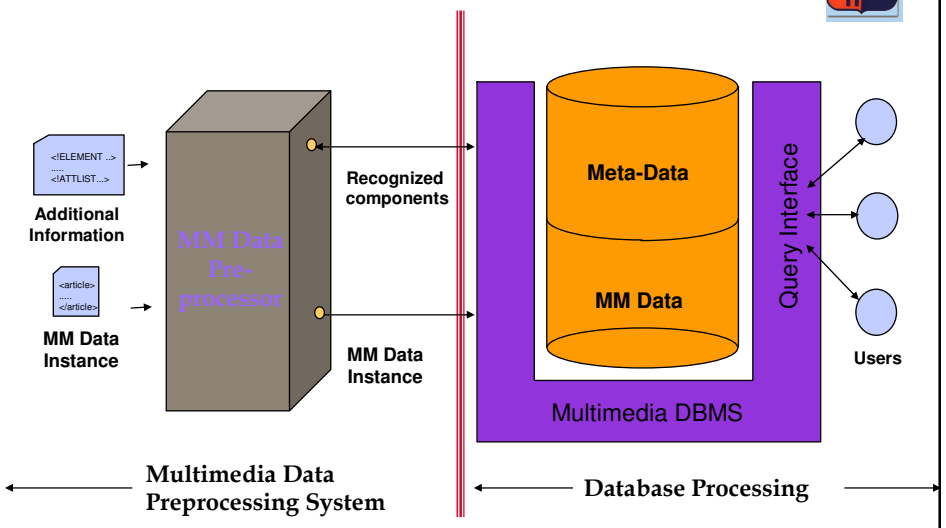
Multimedia databases: Challenges

- Retrieval and Indexing
  - "Give me all video sequences showing Steve laughing"
- Storage, Communication and performances
  - Standards (MPEG)
  - Streaming and Resource Scheduling
    - Continious data such as audio and video need continuous transport (QoS)
- Authoring
- Presentation

Outline

- Mono or Multi-media databases (or data repositories)
- Image Content Analysis and Description
- Image Querying
- Multimedia Metadata Standards
- Conclusion

Multimedia Database Architecture



Document Database Architecture

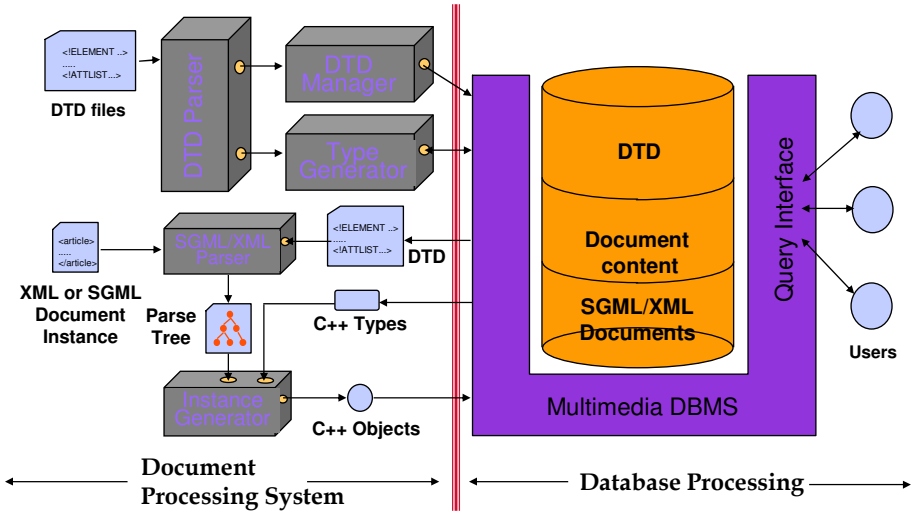
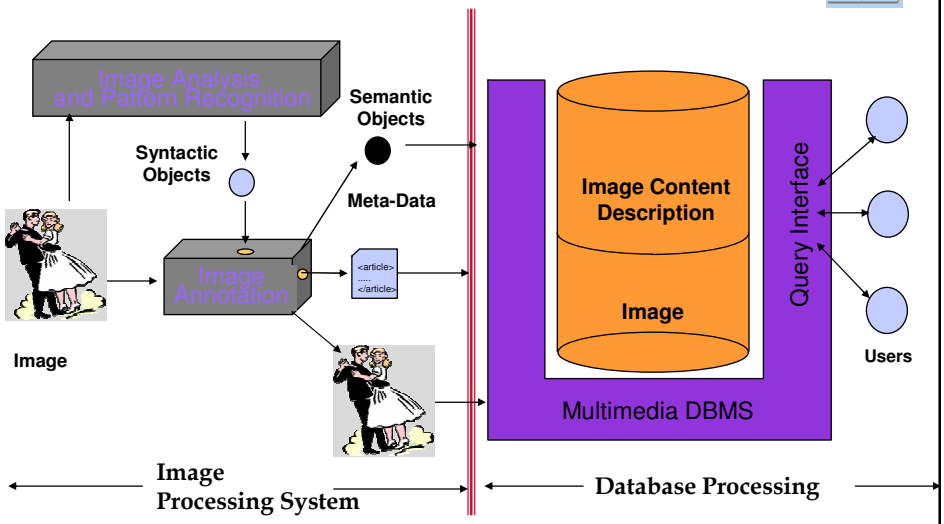
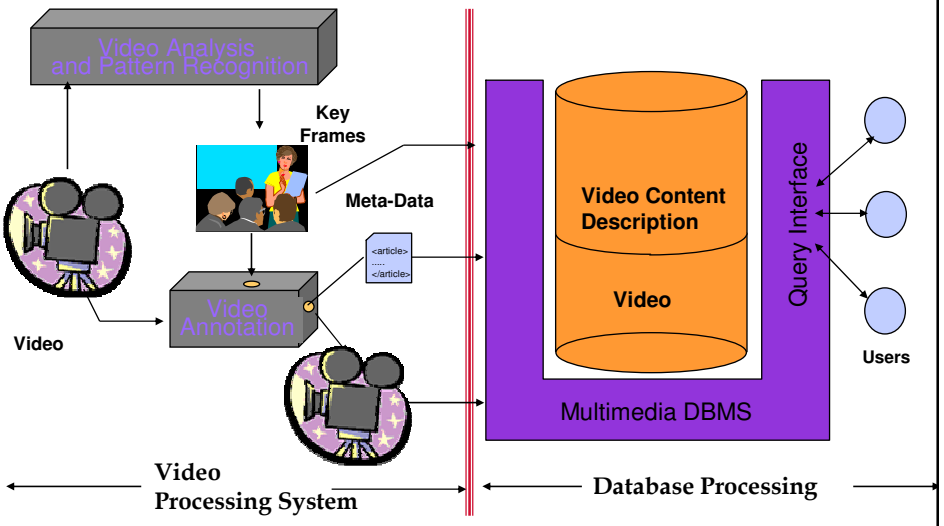


Image Database Architecture



Video Database Architecture



Multimedia at the Confluence of Many Disciplines

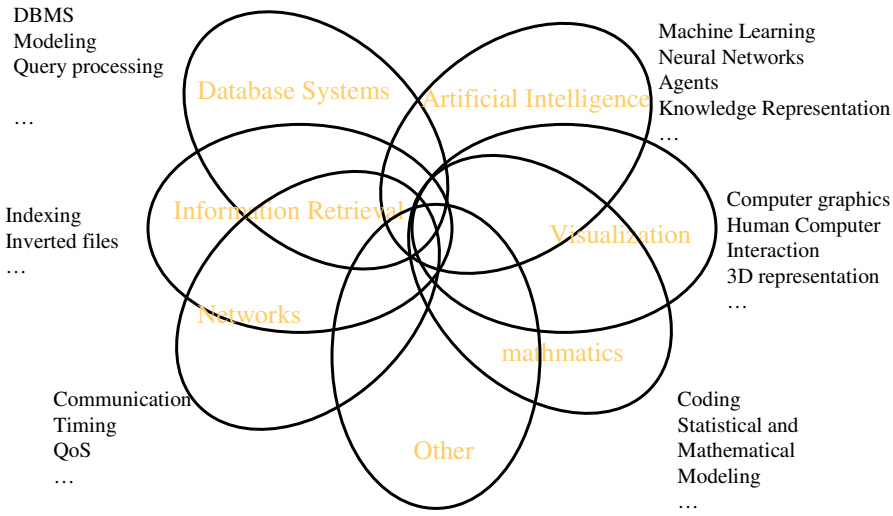


Image Content Analysis

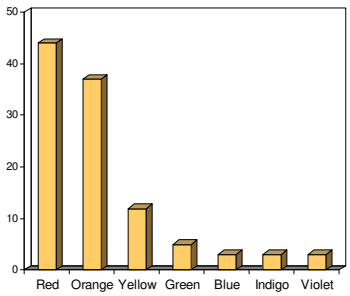
- Image content analysis can be categorized in 2 groups:
  - Low-level features: vectors in a multi-dimensional space
    - Color
    - Texture
    - Shape
  - Mid- to high-level features: Try to infer semantics
  - Semantic Gap

Image Content Analysis: Color

- Color space:
  - Multidimensional space
  - A dimension is a color component
  - Examples of color space: RGB, CIELAB, CIEL*u*v*, YCbCr, YIQ, YUV, HSV
  - RGB space: A color is a linear combination of 3 primary colors (Red, Green and Blue)
- Color Quantization
  - Used to reduce the color resolution of an image
- Three widely used color features
  - Global color histogram
  - Local color histogram
  - Dominant color

Color Histograms

- Color histograms indicate color distribution without spatial information
  - Color histogram distance metrics



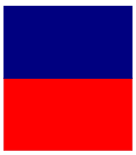


Image Content Analysis: Texture

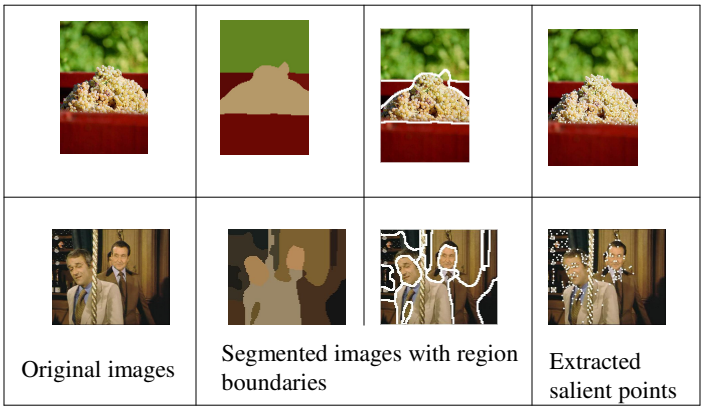
- Refers to visual patterns with properties of homogeneity that do not result from the presence of only a single color
- Examples of texture: Tree barks, clouds, water, bricks and fabrics
- Texture features: Contrast, uniformity, coarseness, roughness, frequency, density and directionality
- Two types of texture descriptors
  - Statistical model-based
    - Explores the gray level spatial dependence of texture and extracts meaningful statistics as texture representation
  - Transform-based
    - DCT transform, Fourier-Mellin transform, Polar Fourier transform, Gabor and wavelet transform

Image Content Analysis: Shape

- Object segmentation
  - Approaches:
    - Global threshold-based approach
    - Region growing,
    - Split and merge approach,
    - Edge detection app
  - Still a difficult problem in computer vision. Generally speaking it is difficult to achieve perfect segmentation

Salient Objects vs. Salient Points

Generic low-level description of images into salient objects and salient points



Original images | Segmented images with region boundaries | Extracted salient points

Modeling Images – Principles

- Support for multiple representations of an image
- Support for user-defined categorization of images
- Well-defined set of operations on images
- An image can have (semantic, functional, spatial) relationships with other images (or documents) which should be represented in the DBMS
- An image is composed of salient objects (meaningful image components)

Salient Object Modeling

- Multiple representations of a salient object (grid, vector) are allowed
- A salient object *O* is of a particular type which belongs to a user defined salient object types hierarchy
- An image component may have some (semantic, functional, spatial) relationships with other salient objects

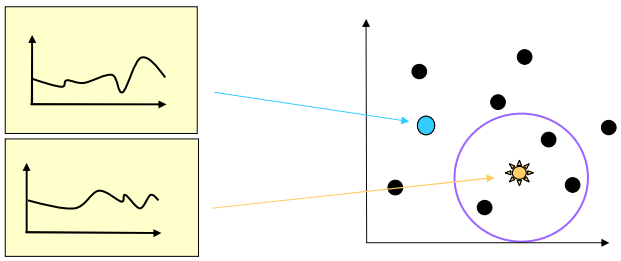
General Approach to Similarity Queries

- A descriptor is a point in a multidimensional space
- Querying consists in defining a metric in the space and computing distances between a query point and the points in the space.

The vector-based Similarity Searches

1) extract from each object N *numerical features* and map objects into points of a N-dimensional space

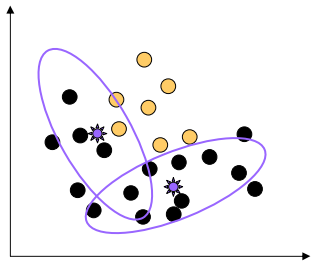
2) use a suitable *distance* (e.g., Euclidean) over such a space, and search for "close" objects using a multi-dimensional ("spatial") index
(low distance = high similarity)



User-defined similarity

- Using the same distance function is not always appropriate

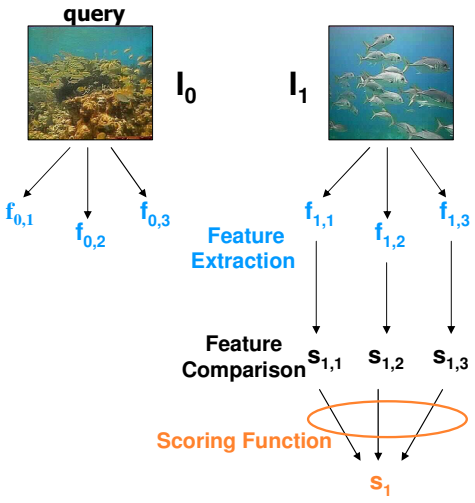
Example: retrieve (only) black points



Integration of Sub-Query Results

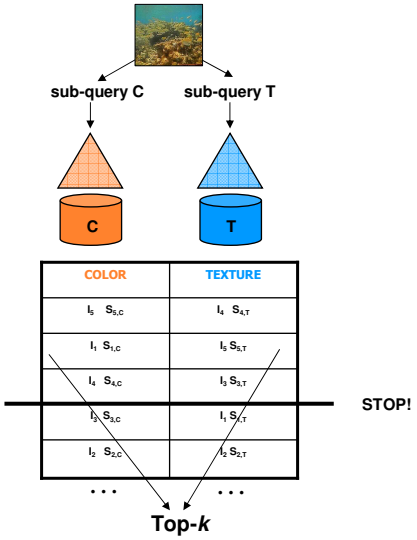
- Image similarity queries are processed by splitting the overall query into sub-queries. *How to obtain an effective and efficient integrating result*?

- The common approach is to define a monotonic scoring function (e.g. min and avg) that associates to each object I_i a numeric value (overall score) s_i
- An object is better (preferred to) than another iff its s is higher



Common Approach: Efficiency

- In order to minimize the number of DB accesses, "middleware" algorithms are applied to return only the Top-*k* highest scored objects
  - TA [FLN01]
  - MEDRANK [FKS03]
  - Upper [BGM02]



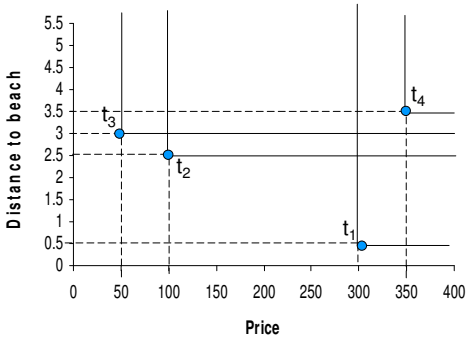
Skyline Operator

- The Skyline filters out the set of objects that are not dominated by any other object

  "An object dominates another object if it is equal or better in all its dimensions and better in at least one dimension"

	Hotel name	Distance to beach (km)	Price ($)
t_1	Opera	0.5	300
t_2	Chariot	2.5	100
t_3	Star	3	50
t_4	Hilton	3.5	350

$t_3 \rightarrow t_4$, $t_2 \rightarrow t_4$, $t_1 \rightarrow t_4$



Best Operator

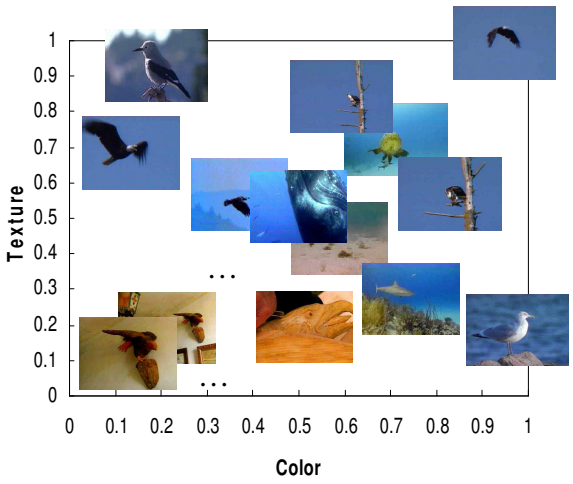
- Skyline filters out only the "first " better results

  …but what happens if the cardinality of Skyline is low and the user want more results?

- The Best operator has been recently proposed
  - Independent from the selected partial order
  - Given a PO and the level n (n>0), Best computes
    - the best results (level 1)
    - the "second choices" (level 2)
    - …
    - the "n-th choices" (level n)

t_3 t_2 t_1 t_4

MM Example: Best



MPEG-7 Objectives

- MPEG-7, formally called "Multimedia Content Description Interface", standardised:
  - A language to specify description schemes, i.e. a Description Definition Language (DDL).
  - A set of Description Schemes and Descriptors
  - A scheme for coding the description
- Developed by the International Standard Organization and the International Electrotechnical Commission (IEC)
- Standardized representation of multimedia metadata in XML (XML Schema Language)
- Describes audio-visual content as a number of levels (features, structures, semantics, models,…)

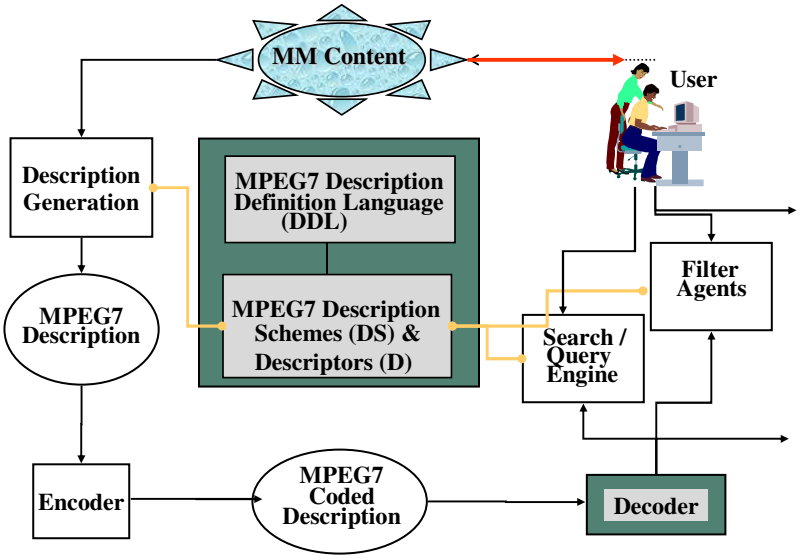
MPEG 7 Context and Objectives

- Content Description
  - format independent
  - may be applied to analogue media
  - different description granularities
- Supplementary Data
- Application Types

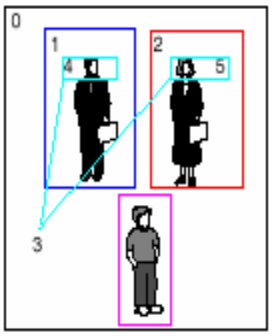
Content Description in MPEG-7

- Metadata
  - Author, producer, copyright
- Semantics
  - Objects, events, people
- Structure
  - Region, segments
- Features
  - Colors, textures, shapes

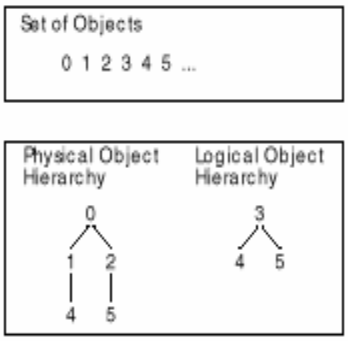
MPEG-7 Framework



MPEG-7 Example



a)



b)

MPEG-7 Example (cont)

- <object_hierarchy type="PHYSICAL"> <!-- Physical hierarchy -->
- <object_node id="10" object_ref="0"> <!-- Portrait -->
- <object_node id="11" object_ref="1"> <!-- Father -->
- <object_node id="12" object_ref="4"/> <!-- Father's face -->
- </object_node>
- <object_node id="13" object_ref="2"> <!-- Mother -->
- <object_node id="14" object_ref="5"/> <!-- Mother's face -->
- </object_node>
- </object_node>
- </object_hierarchy>
- <object_hierarchy type="LOGICAL"> <!-- Logical hierarchy: faces in the image -->
- <object_node id="15" object_ref="3"> <!-- Faces -->
- <object_node id="16" object_ref="4"/> <!-- Father's face -->
- <object_node id="17" object_ref="5"/> <!-- Mother's face -->
- </object_node>
- </object_hierarchy>

MPEG-21

- MPEG-7 does not take into account the aspect of the organization of the infrastructure of distributed multimedia systems
- Initiated in 2000 to provide mechanisms for distributed multimedia systems design and associated services
- A new distribution entity is proposed and validated: the Digital item

```
<ITEM>
   <DESCRIPTOR>
    <STATEMENT TYPE="text/plain">Album #1: Summer 2005</STATEMENT>
   </DESCRIPTOR>
   <ITEM>
    <DESCRIPTOR>
     <STATEMENT TYPE="text/plain">
        Visiting "Sagrada Familia" with my friends.
     </STATEMENT>
    </DESCRIPTOR>
     <COMPONENT>
      <RESOURCE REF="photo_1.jpg" TYPE="image/jpg" />
     </COMPONENT>
   </ITEM>
   <ITEM>
    <DESCRIPTOR>
     <STATEMENT TYPE="text/plain">
        We are waiting for the bus.
        My brother is trying to find our hotel in the map.
      </STATEMENT>
    </DESCRIPTOR>
     <COMPONENT>
      <RESOURCE REF="photo_2.jpg" TYPE="image/jpg" />
     </COMPONENT>
  </ITEM>
</ITEM>
```


Conclusion: Towards more Semantics

- Semantic Representation
  - MPEG-7
- Semantic Acquisition
  - Automated approach for all purpose images and videos have failed
  - Manual interpretation is tedious
- How to minimize human interventions?

Towards more Semantics: Challenges

- "A picture is worth a thousand words"
- "It is not so much that a picture is worth a thousand words, for many fewer words can describe a still picture for most retrieval purpose, the issue has more to do with the fact that those words vary from one person to another"
- ``In spite of almost fifty years of research, design of a general-purpose machine pattern recognizer remains an elusive goal....The best pattern recognizers in most instances are humans, yet we do not understand how humans recognize patterns".

Conclusions

- Gap between research and applications in general
  - Semantic Gap
- What is Next?
  - Capture more semantics
  - Real multimedia databases and application
- Indexing and dimensional reduction
- Integration of Multimedia and Classical Data