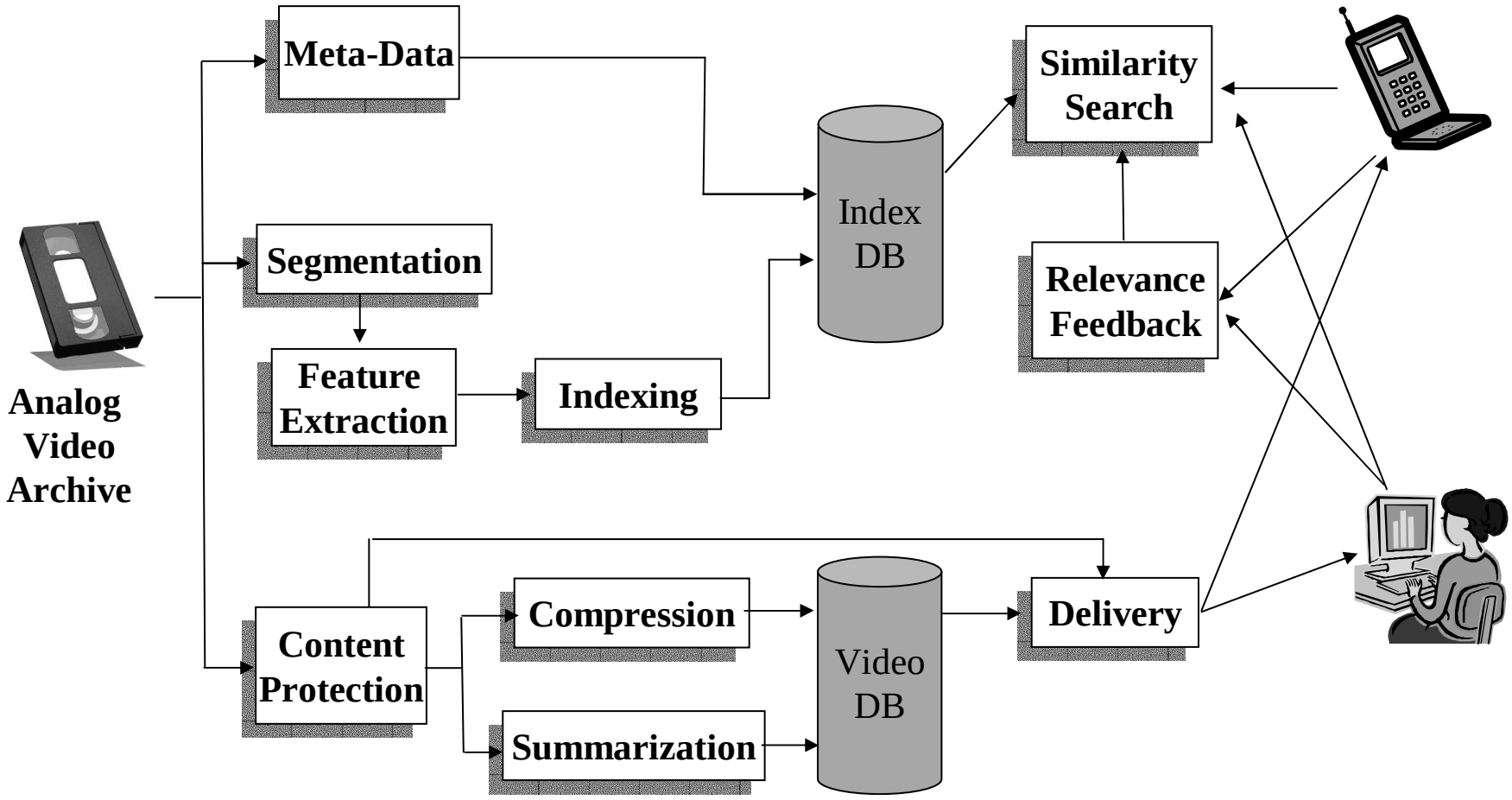
Multimedia Information Systems

Samson Cheung

EE 639, Fall 2004

Lecture 2: Course Overview

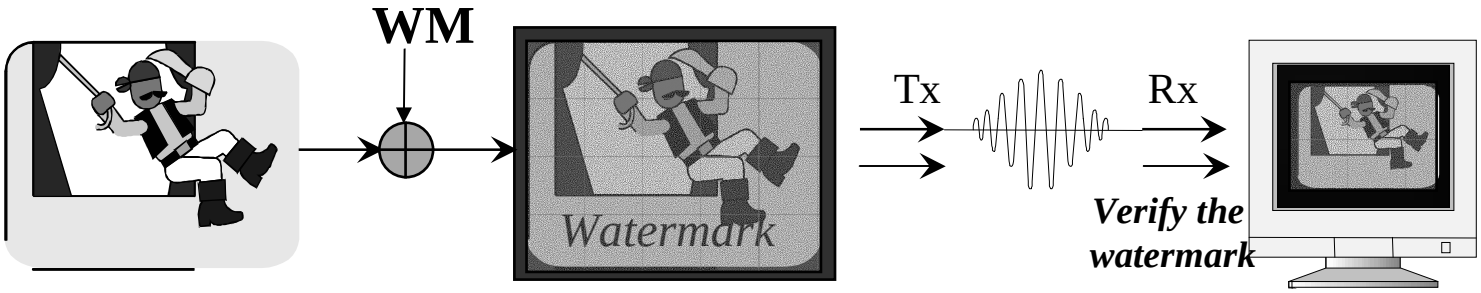
Digital Video Library



Content Protection

- **Problems:**
  - **Forgery detection**
  - **Authentication**
  - **Tracking**
  - **Secure distribution (scrambling)**
- **Cover:**
  - **Basic DRM concept and infrastructure**
  - **Digital Watermarking**

Watermarking and Content Protection



- Stenography vs. watermarking
- Watermark
  - Visible vs. invisible
  - Spatial vs. frequency domain
  - Fragile vs. robust vs. semi-fragile
- Applications: copyright, user ID, content ID, authentication, data hiding

Compression & Delivery

- **Problems:**
  - **Large Data Size: Digital video 720x480x24x30 → 249 Mbps, 112 GB per hour**
  - **Variation in capabilities in receivers: cell phone, PDA, PC, TV, HDTV**
  - **Variation in network: mobile, Wi-Fi, Internet, Satellite**
- **Cover:**
  - **Review image, video, audio, and graphics compression**
  - **Survey scalable compression, joint source-channel coding, multiple description, distributed compression**

Segmentation

- **Problem:**
  - **Break down complex audio-visual objects into "atomic" units for retrieval**
- **Cover:**
  - **Basic image segmentation**
  - **Video segmentation – shot and story level**
  - **Special "events" detection**
    - Face
    - Object Tracking
    - Video Structure Modeling

Feature Extraction

- **Problems:**
  - **Represent an audio-visual object as a (multiple) vector with a (multiple) distance function d(.,.) for measuring similarity**
- **Example:**
  - **Vector : Color histogram**
  - **Distance : L-1, KL Divergence, Earth-mover**
- **Cover:**
  - **Simple features for image, video, audio, 3-D objects**
  - **Various similarity functions**

Similarity Search & Indexing

- **Similarity Search : Given a query vector q, find v in a database D of feature vectors such that d(q,v) is the smallest (most similar).**
- **Problem:**
  - **D could be very large**
    - Sequential is slow. Logarithmic search is very desirable!
  - **The dimension of v is very high (or even uncountable, e.g. graph of spatial layout)**
  - **“Curse of dimensionality” – in high-dimensional space, most fast search methods become sequential**
- **Cover: feature selection, dimension reduction, clustering-based data structure (trees), and randomized search**

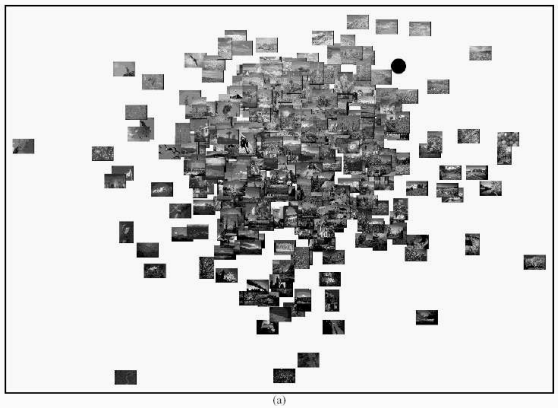
Relevance Feedback

- **Problems:**
  - **Simple distance function does not reflect human's notion of similarity → Learn from human**
- **Cover:**
  - **Classification: Define three different classes P=positive, N=negative, D=don't care with respect to a query q. Given a small human-labeled set for each class, identify all v in a database D that belongs to class P.**
  - **Learning: Minimize the size of labeled sets**

Summarization

- **Problem:**
  - **How to present 100 clips of video, each 3 hours long, to a user to select?**
  - **Limited capability of receivers**
- **Cover:**
  - **Keyframes, Mosaicing**
  - **Hierarchical Clustering**
  - **Spatial Summary and visualization**
  - **Skim**

Examples of Summarization



EMD Search

a. Level 1: Clustering of all the video shots by temporal variances

b. Level 2: Clustering of shots under the second class above by color histograms (Lu*v* space)

c. Anchorperson shots under the third class at level 2

Structural Parsing

Skim: Drastically condensed audio-video clips

Meta-Data

- **Problem:**
  - **Auxiliary information**
  - **Unified textual description of features and semantic concepts**
  - **Highly searchable**
- **Cover:**
  - **XML, MPEG-7, MPEG-21**
  - **Traditional IR techniques on texts and hypertext**

Auxiliary Information

```
<Event id="Goal">
    <Label> <Term CSLocation="http://www.CSs.com/Sports" CSTermId="43">
        <Label xml:lang="en-us"> Goal </Label>
    </Term> </Label>
    <SemanticLocation>
        <Place>    <PlaceName xml:lang='en'> Santiago Bernabeu </PlaceName>
                   <Country> Spain </Country>
                   <PostalAddress> C/Fabregas No. 26, Barcelona </PostalAddress>
    </Place>
    </SemanticLocation>
    <SemanticTime>
        <Definition>  <FreeText> 8:33pm, Saturday, March 20, 2000 </FreeText> </Definition>
    </SemanticTime>
    <MediaOccurence>
        <MediaLocator><MediaTime>      <MediaTimePoint> T0:0:0 </MediaTimePoint>
                                       <MediaDuration> PT5S </MediaDuration>
        </MediaTime></MediaLocator>
    </MediaOccurence>
</Event>
```

Content Semantic Description

```
<Semantic id="S1">
    <Object id="Forward-ob"> … </Object>
    <Object id="Ball-ob"> … </Object>
    <Object id="GoalKeeper-ob"> … </Object>
    <Object id="Goal-ob"> … </Object>
    <Event id="Goal-ev">
        <Event id="Kick-ev"> … </Event>
        <Event id="Not-Catch-ev"> … </Event>
        <Event id="Enter-ev"> … </Event>
    </Event>
    <SemanticGraph>
        <Edge name="agent" source="Kick-ev" target="Forward-ob"/>
        <Edge name="patient" source="Kick-ev" target="Ball-ob"/>
        <Edge name="destination" source="Kick-ev" target="Goal-ob"/>
        <Edge name="agent" source="Enter-ev" target="Ball-ob"/>
        <Edge name="patient" source="Enter-ev" target="Goal-ob"/>
    <SemanticGraph>
</Semantic>
```

Course format

- **3 lectures per week**
- **6 homeworks**
- **1 final project**
  - **Two types of projects**
    - Experimental or Implementation (up to three persons)
    - Survey (single person)
  - **mid-term proposal**
  - **final report and presentation**
  - **may be extended to “Advanced Projects”**