Multimedia Storage & Multimedia Information Retrieval (MIR)

Lectures 9-10

BIL464 Multimedia Systems

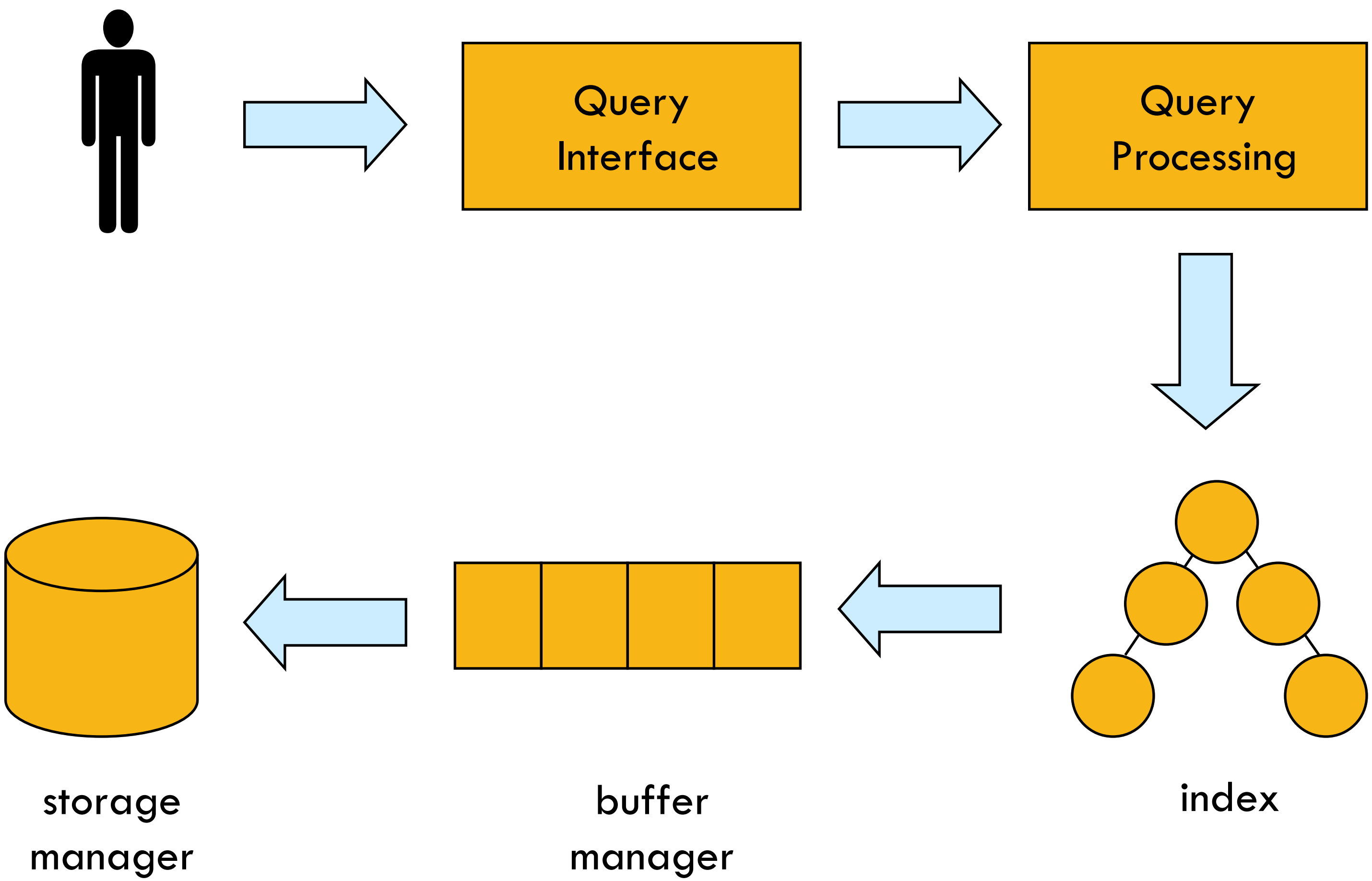
Mustafa Sert
Asst. Prof.
msert@baskent.edu.tr

Department of Computer Engineering, Başkent University
Ankara 06810 TURKEY

Outline

- Storage support of DBMSs for MM data
  - Main components of a typical DBMS
  - Requirements for MM data
  - Data types for MM data
- Multimedia Information Retrieval (MIR)
  - A typical architecture
  - Challenges in MM retrieval
  - Content-Based Retrieval (CBR) of Multimedia Data
  - Querying
  - Relevance Feedback
  - Evaluation of Results

Basic DBMS Components..



Requirements of MM-DBMS

- **Persistence**
  - Data objects can be saved and re-used by different transactions and program invocations
- **Privacy**
  - Access and authorization control
- **Integrity control**
  - Ensures database consistency between transactions
- **Recovery**
  - Failures of transactions should not affect the persistent data storage
- **Query support**
  - Allows easy querying of multimedia data

Requirements of MM-DBMS (*cont.*)

- In addition, an MM-DBMS should:
  - have the ability to uniformly **query** data (media data, textual data) represented in different formats.
  - have the ability to **simultaneously query** different media sources and conduct classical database operations across them.

    ⇒ *query support*
  - have the ability to **retrieve media objects** from a local storage device in a continuous manner.

    ⇒ *storage support*
  - have the ability to take the answer generated by a query and develop a **presentation** of that answer in terms of audio-visual media.
  - have the ability to **deliver** this presentation in a way that satisfies various Quality of Service requirements.

    ⇒ *presentation and delivery support*

MM Database Storage Issues

- DBMS typically provide three different kinds of domain for multimedia data:
  - **Large object** domains, sequences of data often of two kinds
    - Binary Large Objects – **BLOBs** – which are an unstructured sequence of bytes
    - Character Large Objects – **CLOBs** – which are an unstructured sequence of characters
  - **File references (BFILEs)** – instead of holding the data, a file reference contains a link to the data (e.g., OLE in Access).
  - Genuine **multimedia data types** – (Oracle and Jasmine)

Domain Types for MM Data

- The **BFILE** datatype provides access to BLOB files of up to 4 gigabytes that are stored in file systems outside an Oracle database.
  - The BFILE datatype allows read-only support of large binary files; you cannot modify a file through Oracle. Oracle provides APIs to access file data.

Large Object Types in Oracle and SQL3

- Oracle and SQL3 support three large object types:
  - **BLOB** - The BLOB domain type stores unstructured binary data in the database. BLOBs can store up to four gigabytes of binary data.
  - **CLOB** – The CLOB domain type stores up to four gigabytes of single-byte character set data
  - **NCLOB** - The NCLOB domain type stores up to four gigabytes of fixed-width and varying width multi-byte national character set data

* *SQL3 is a significant extension to standard SQL which turns into a full object-based language*

The Properties of LOBs

- **Concatenation** – making up one LOB by putting two of them together
- **Substring** – extract a section of a LOB
- **Overlay** – replace a substring of one LOB with another
- **Trim** – removing particular characters (e.g. whitespace) from the beginning or end
- **Length** – returns the length of the LOB
- **Position** – returns the position of a substring in a LOB
- **Upper and Lower** – turns a CLOB or NCLOB into upper or lower case
- **LOBs** can only appear in a **where** clause using "=", "<>" or "like" and not in **group** by or **order by** at all

Data types for MM Data – Example

- If we want to store students' pictures in a table, we can create a table as follows:

```
create table STUDENT (
          stu# integer,
          name char(20),
          address char(100)
          picture BLOB);
```

- BLOBs are normally just bit strings, so that
  - Operations such as comparison can not be carried out on them.
  - A DBMS does not know the contents or semantics of a BLOB. All it knows is a block of data

Data types for MM Data – Example

- Another type of DBMSs is object-oriented database management systems (OODBMSs)
- In such a system, objects are properly defined in the object-oriented sense
  - Each object contains attributes and methods
  - So that methods can operate on atributes
- **The main difference between the BLOB and the object is that**
  - The object is properly defined, including its attributes and allowed operations on the attributes, while the BLOB is not

Data types for MM Data – Example

- An example would be:

```
create type IMAGE (
private
        size integer,
        resolution integer,
        content float[],
public
     ...
);
```

```
create table STUDENT (
        stu# integer,
        name char(20),
        address char(100)
        picture IMAGE);
```

- While objects contain some simple attributes, mANY more capabilities should be developed to handle content-based multimedia retrieval, such as:
  - Feature extraction
  - Similarity metrics/calculation
  - Indexing structures,
  - Etc.

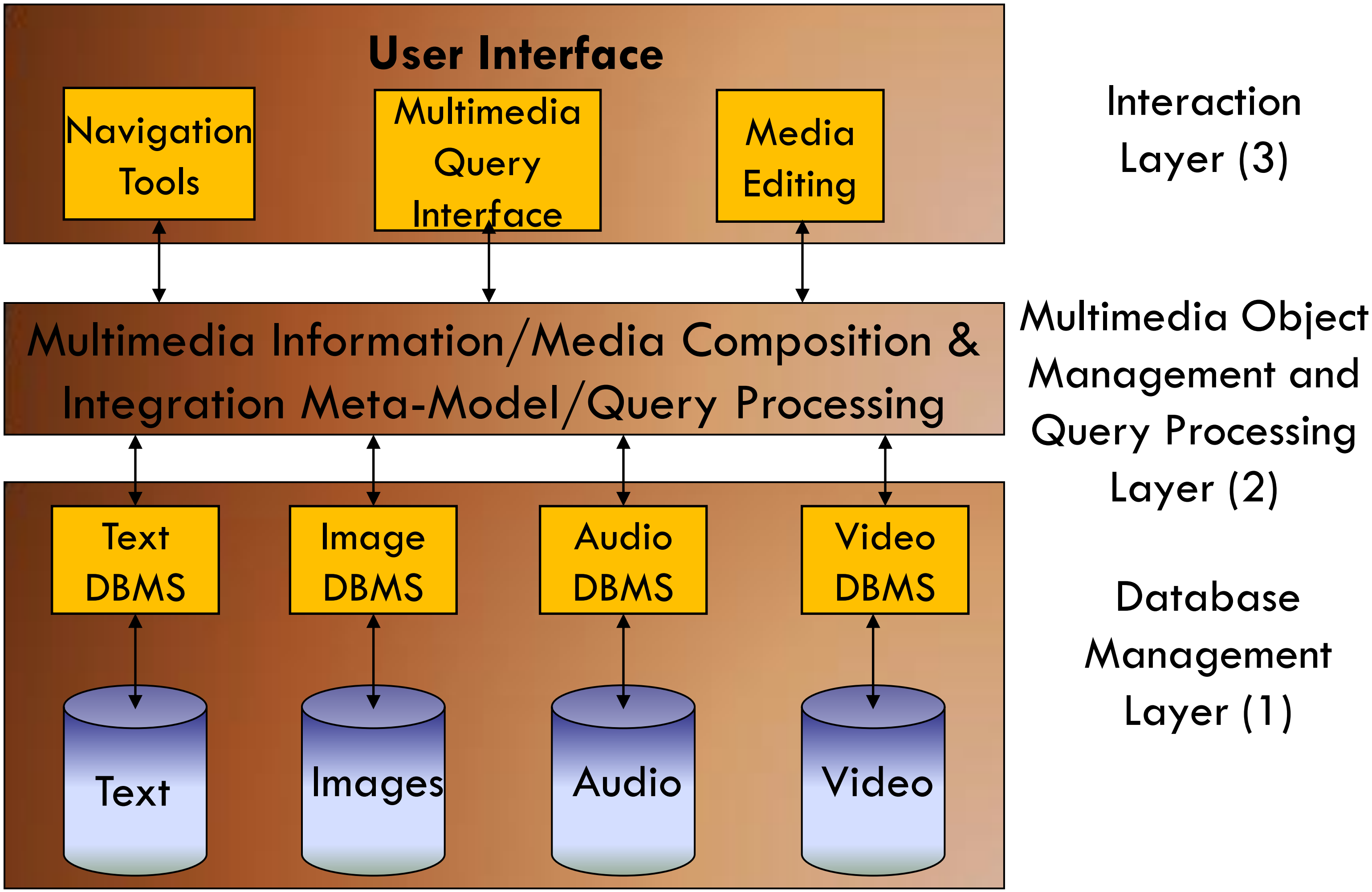
Outline

- Storage support of DBMSs for MM data
  - Main components of a typical DBMS
  - Requirements for MM data
  - Data types for MM data
- **Multimedia Information Retrieval (MIR)**
  - A typical architecture
  - Challenges in MM retrieval
  - Content-Based Retrieval (CBR) of Multimedia Data
  - Querying
  - Relevance Feedback
  - Evaluation of Results

Multimedia Information Retrieval (MIR)

- Motivation
  - With the explosive growth of digital media data, there is a huge demand for new tools and systems that enables average users to more efficiently and more effectively search, access, process, manage, author and share these digital media contents.

A Typical MIR Architecture



A Typical MIR Architecture – *Cont.*

- A typical multimedia information retrieval system has to address the followings:
  - **archive**
    - text, video, images, speech, music, combinations
  - **query**
    - text, stills, sketch, speech, humming, examples
    - **content-based**
  - **present results**
    - browsing, summaries, story boards
    - document clustering, cluster summaries
  - **utilise relevance feedback**

MIR Types

- Text-based Information Retrieval
  - Too many MM data to annotate
  - High cost of human interpretation
  - Subjectivity of AudioVisual content, e.g., “A picture is worth a thousand words”
- Content-based Retrieval (CBR)
  - automatically retrieves images, video, and audio based on the visual and audio content

MIR – *Challenges.*

- Data Types
  - Text, hypertext, image, audio, graphics, animation, paintings, video/movie, rich text, spread sheet, slides, combinations of these and user interaction
- Research Problems
  - Systems, content, services, user, evaluation, implementation, social/business, applications
- Methodologies
  - Database, information retrieval, signal and image processing, graphics, vision, human-computer interaction, machine learning, statistical modeling, data mining, pattern analysis, data fusion, social sciences, and domain knowledge for applications

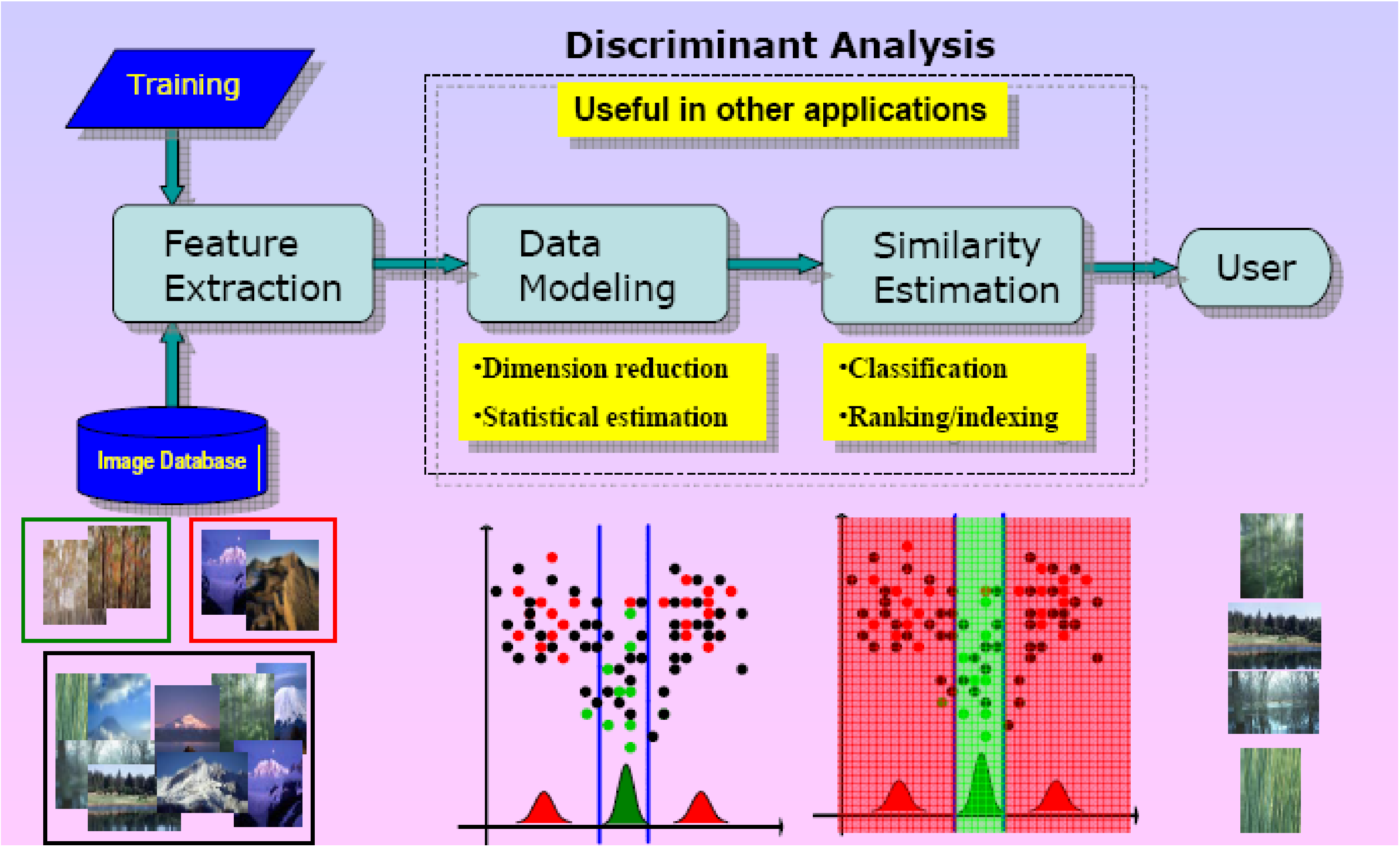
Content-based Retrieval (CBR) of Multimedia Data

- Content-based Image Retrieval (CBIR)
- Content-based Video Retrieval (CBVR)
- Content-based Audio Retrieval (CBAR)

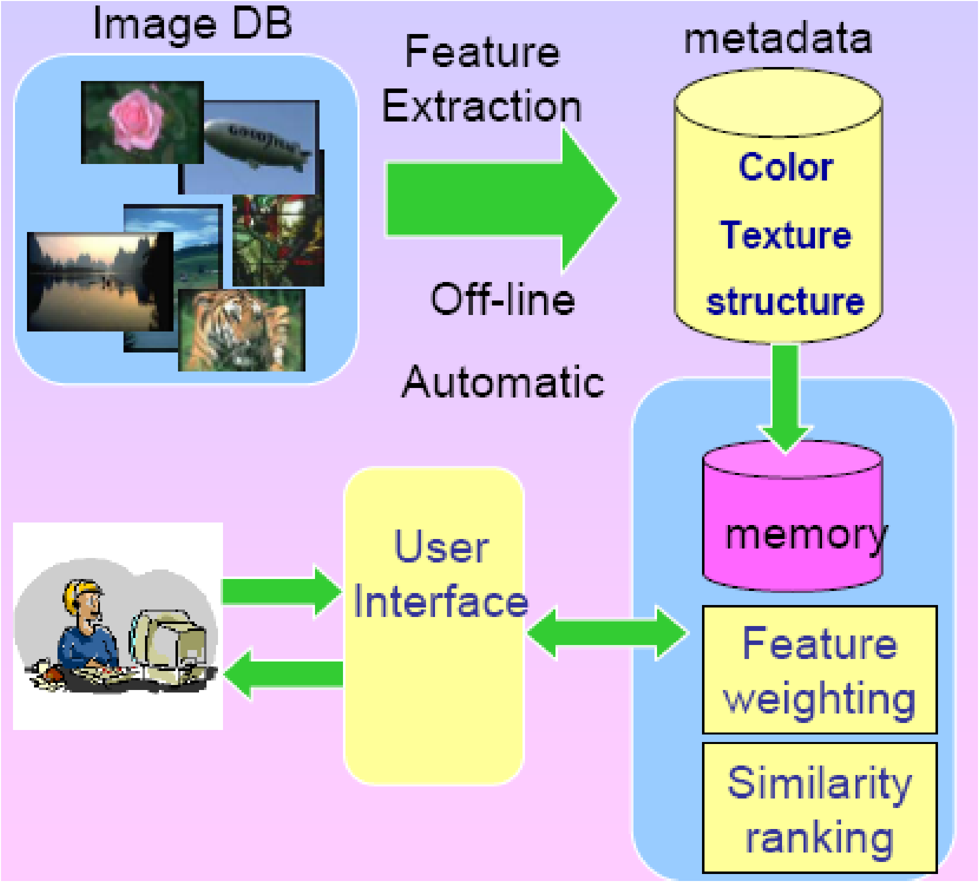
CBR – *Hierarchical Levels*

- High-level
  - Bridge the semantic gap, integration of context and content, hybrid (text and content) approaches
- Mid-level
  - Active Learning, Boosting, Incremental Learning
- Low-level
  - Feature Extraction and Representation, Dimension Reduction and Selection.

A Typical CBR System



Content-based image retrieval (CBIR)



- Color
  - Color histogram,
  - Color localization, …
- Texture
  - Co-occurence matrices
  - Gabor features, …
- Shape
  - 2-D shapes, …
- Structure
  - Edge-based features

Content-based video retrieval

- Traditional Video Retrieval
  - Query-by-textual keyword
- Automatic Visual Concept Detection
  - e.g., indoor/outdoor, Sky, Car, Building, country flags
  - Example concepts: Airplane, Building, Car, Crowd, Desert, Explosion, Outdoor, People, Vehicle, Violence
- Video Retrieval – Scene
  - How to recognize a scene? Context
    - Use Photo-Concepts to describe context
    - Machine learning to link context to concepts

Content-based audio retrieval

- To search sounds by their features in the waveform, statistics, or transform domains
  - Speech, Music, Environment Audio, Silence
- Applications
  - **Entertainment**
    - Film making - searching sound effects
    - TV/radio studio - editing programs
    - Karaoke, music stores, or online shopping
    - query by humming the melody
  - **Audio/video archive management**
    - Segmenting and indexing of raw recordings
    - Searching and browsing audio/video clips
  - **Surveillance**
    - Monitoring criminal or emergent events
    - Film rating

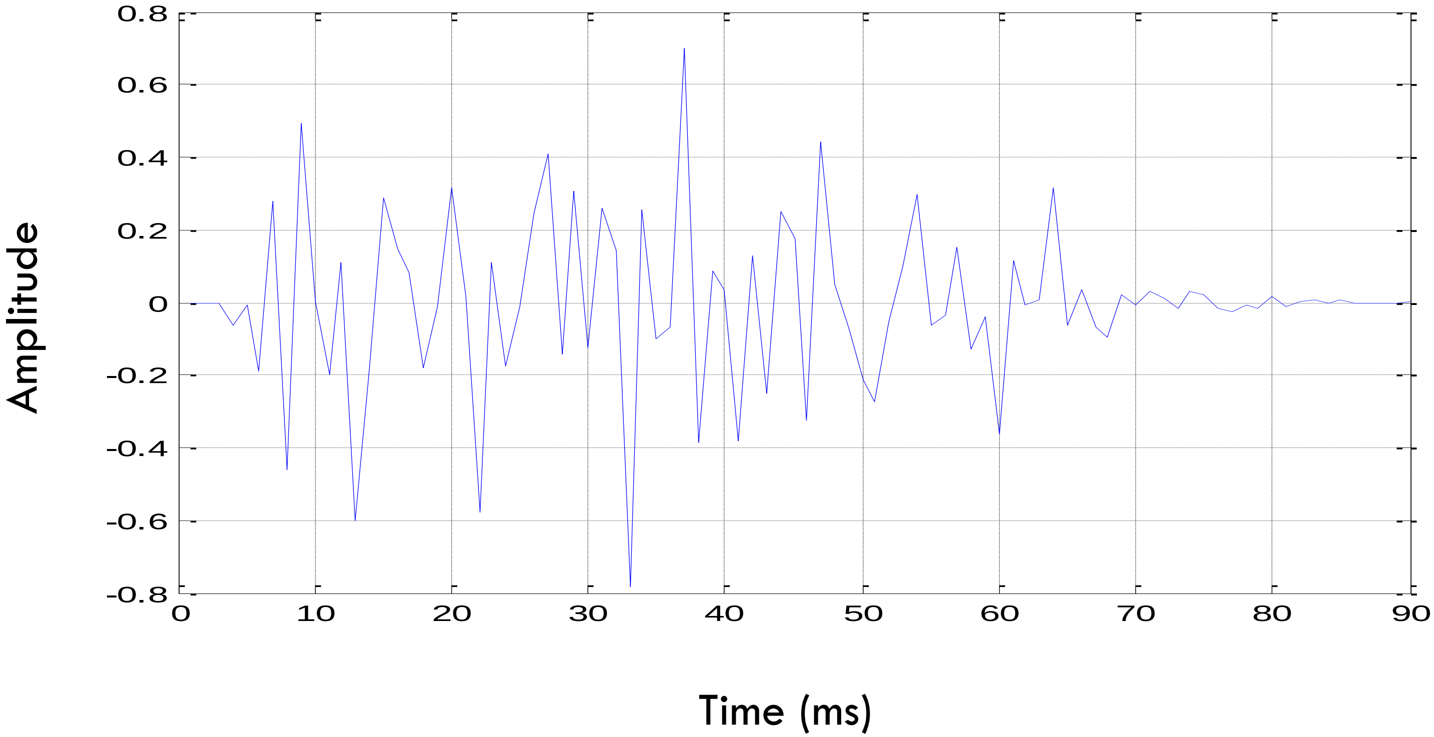
- Short-time Energy (STE)
- Zero-crossing rate (ZCR)
- Pitch Period
- MFCC
- Spectogram
- MPEG-7 features

Main Audio Features

- Time-Domain Features
  - Average Energy
  - Zero Crossing Rate
  - Silence Ratio
- Frequency-Domain Features
  - Sound Spectrum
  - Bandwidth
  - Harmonicity
  - Pitch
- Spectogram

Time-Domain Features

- Amplitude-time representation of an audio signal



Time-Domain Features – *Cont.*

- Average Energy
  - Indicates the loudness of the audio signal

$$E = \frac{\sum_{n=1}^{N-1} x(n)^2}{N}$$

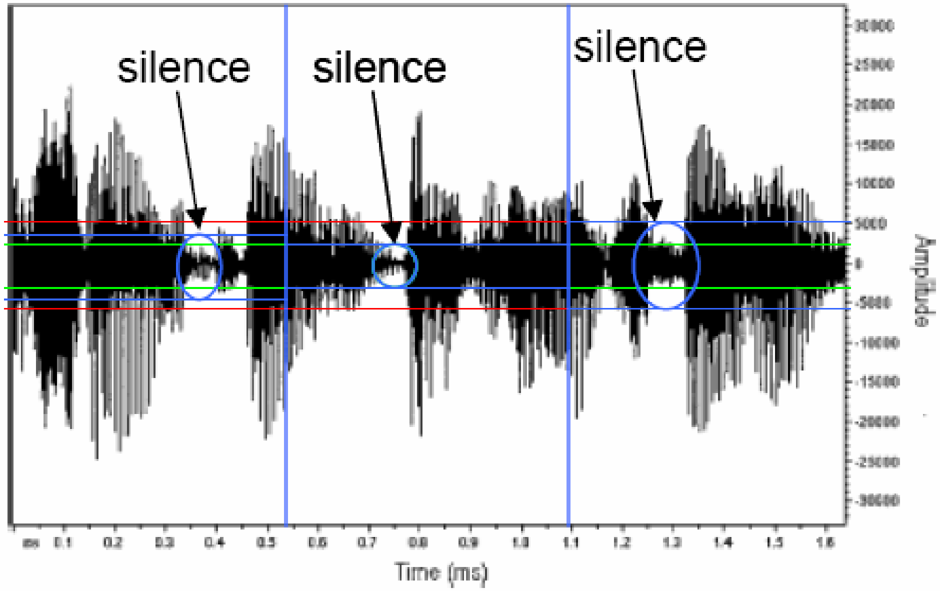
- Zero-crossing Rate (ZCR)
  - Indicates the frequency of signal-amplitude's sign change

$$ZC = \frac{\sum_{n=1}^{N-1} \operatorname{sgn}[x(n)] - \operatorname{sgn}[x(n-1)]}{2N}$$

$$\operatorname{sgn}(a) = \begin{cases} 1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases}$$

Time-Domain Features – *Cont.*

- Silence Ratio
  - Indicates the proportion of the sound piece that is silent
  - Silence is a period within which the absolute amplitude values of a certain number of samples are below a certain threshold

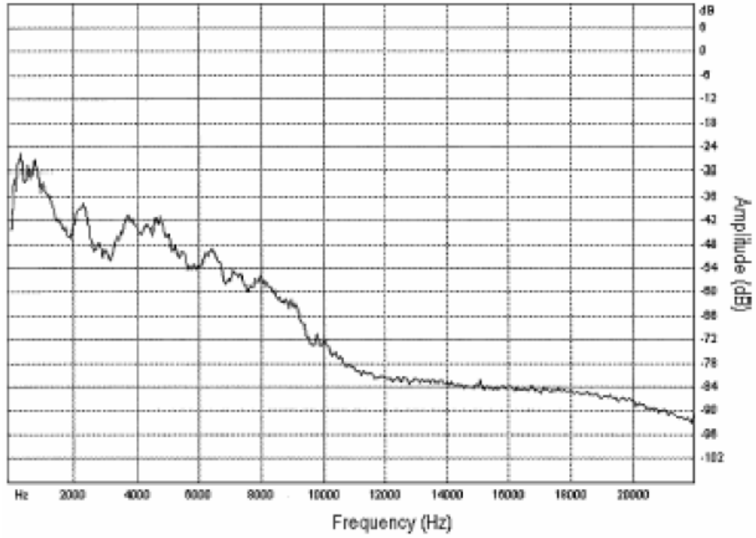


  - Silence ratio is calculated as the ratio between the sum of silence periods and the total length of signal

Frequency-Domain Features

- Sound Spectrum
  - Discrete Fourier Transform

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi nk}{N}}$$



  - For large values of N, the signal is often broken into blocks called frames and DFT is applied to each of the frames (STFT)

Frequency-Domain Features – *Cont.*

- Bandwidth
  - Indicated the frequency range of a sound
  - Can be defined as the difference btw the highest frequency and lowest frequency of non-zero spectrum components (at least 3dB above the silence level)
- Harmonicity
  - In harmonic sound, the spectral components are mostly whole number multiples of the lowest and most often loudest frequency (fundamental frequency)
  - Music is normally more harmonic than other sounds

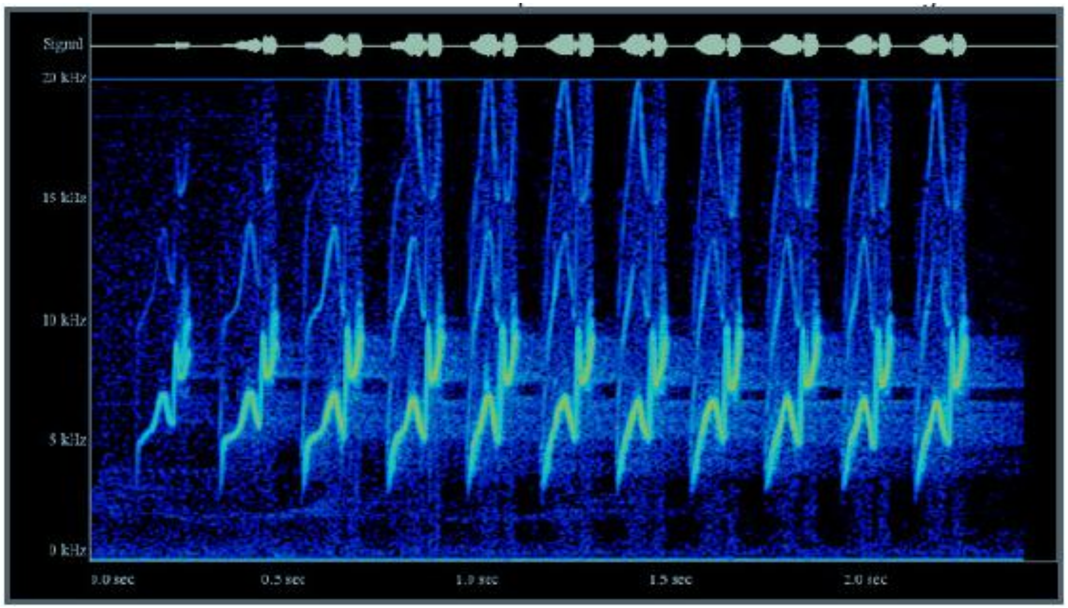
Frequency-Domain Features – *Cont.*

- Pitch
  - The distinctive quality of a sound, dependent primarily on the frequency of the sound waves produced by its source, such musical instruments and the voice
  - In practice, we use the fundamental frequency as the approximation of the pitch.

Spectrogram

- Time and frequency components are shown in the same representation (Intensity).
  - Intensity is defined as the power of a frequency component at a particular time interval

Frequency



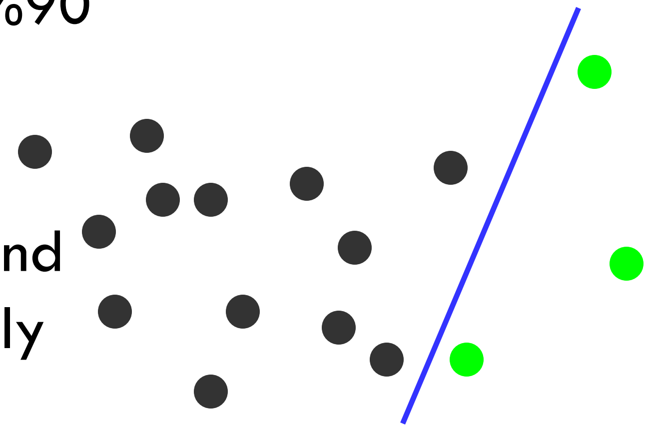
Time (ms)

Querying

- Three Types
  - Point
    - Exact matching
  - K-Nearest
    - Retrieve best K matching
  - Range
    - E.g., retrieve between %80 and %90
- Feature selection & learning

  ADA-Boost, K-NN, SVM, …

  eg, compute separating hyper-plane and rank all images in database accordingly



The Semantic Gap

- 96000 pixels with a particular spatial color distribution
- Human faces, white and red clothes
- Victory, UEFA cup

Bridging the Semantic Gap

- region segmentation + region classification (grass, water, ...)
- using simple models for complex concepts (grass+plates+people = barbeque)

Relevance Feedback

- System needs to learn from user
  - Change the parameters
  - System needs flexibility
  - Multimedia data can be quickly assessed and user can inform system either implicitly or explicitly.
- The basic advantage of putting the user into the loop by using relevance feedback is that the user need not provide a completely accurate initial query.
- For a specific example of relevance feedback with respect to the image search engine is called MARS.

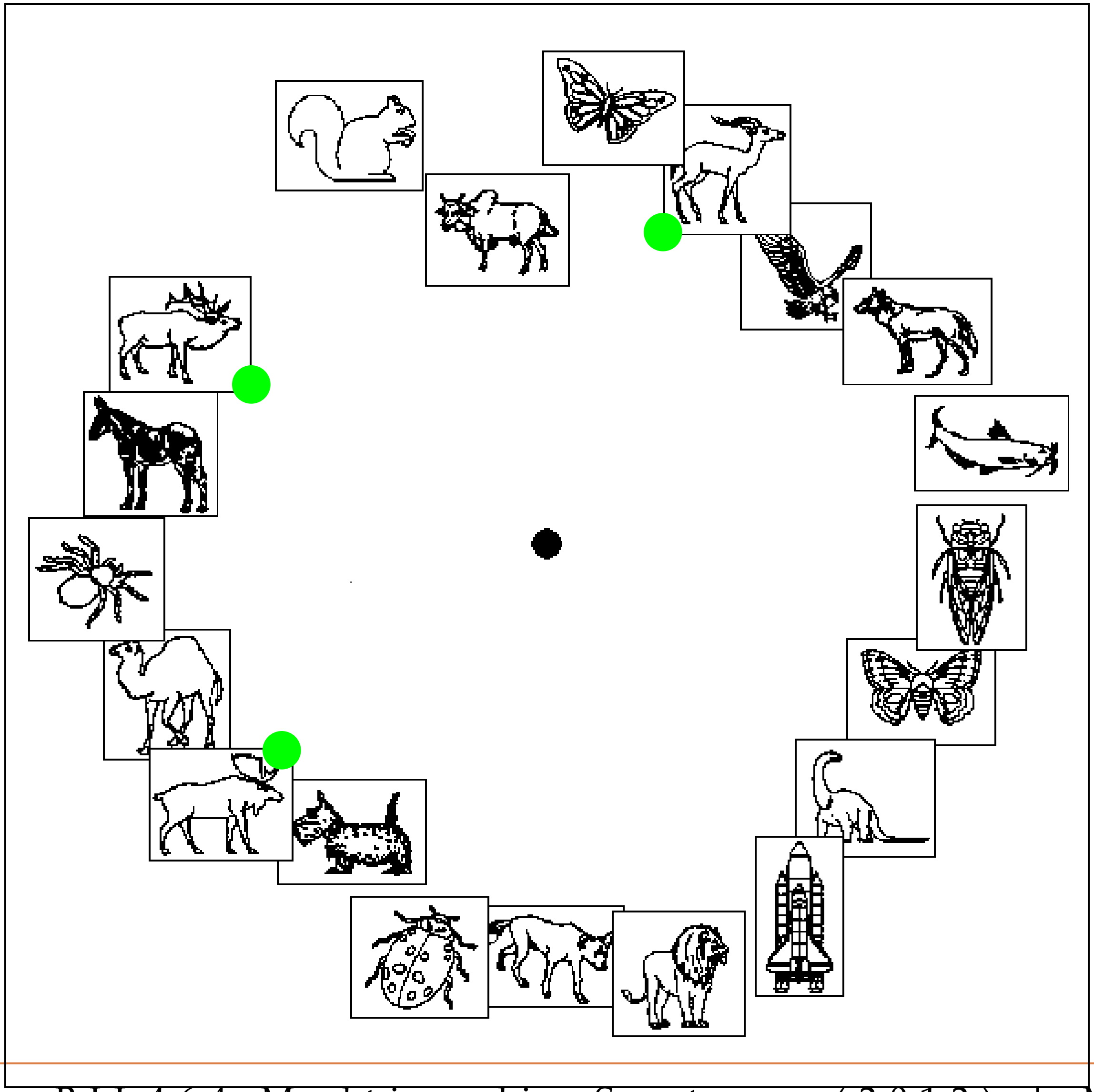
Relevance Feedback Mechanism

- Relevance Feedback Approaches
  - The usual situation: the user identifies images as good, bad, or don’t care, and weighting systems are updated according to this user guidance.
  - Another approach is to move the query towards positively marked content.
  - An even more interesting idea is to move every data point in a disciplined way, by warping the space of feature points.

Relevance Feedback Mechanism

- centre = query = ideal result
- Results are displayed such that distance to centre is the dissimilarity to the query
- User indicates her/his idea of similarity by rearranging the displayed results
- System recomputes optimal parameters for this specific query automatically

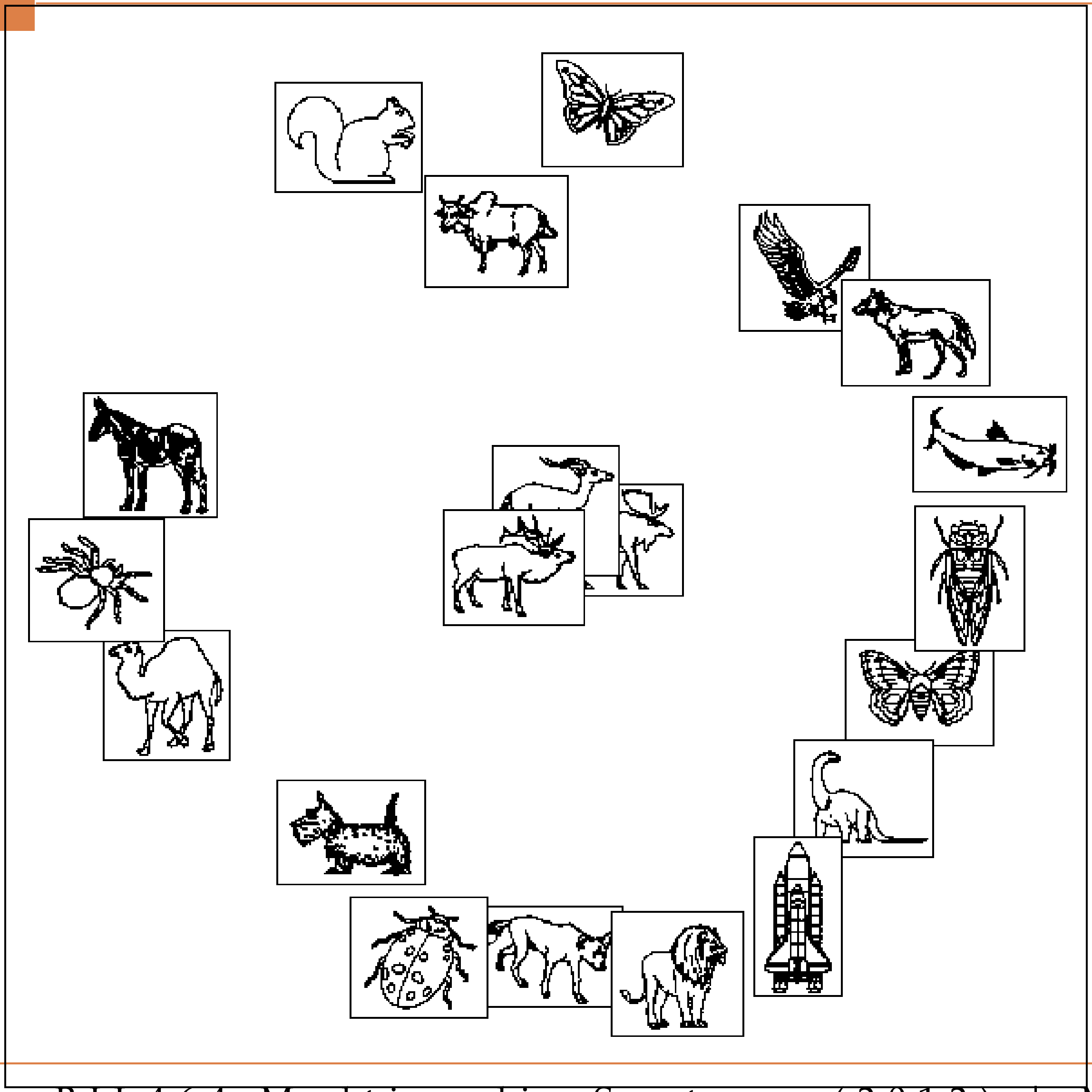
Relevance Feedback – *Example*



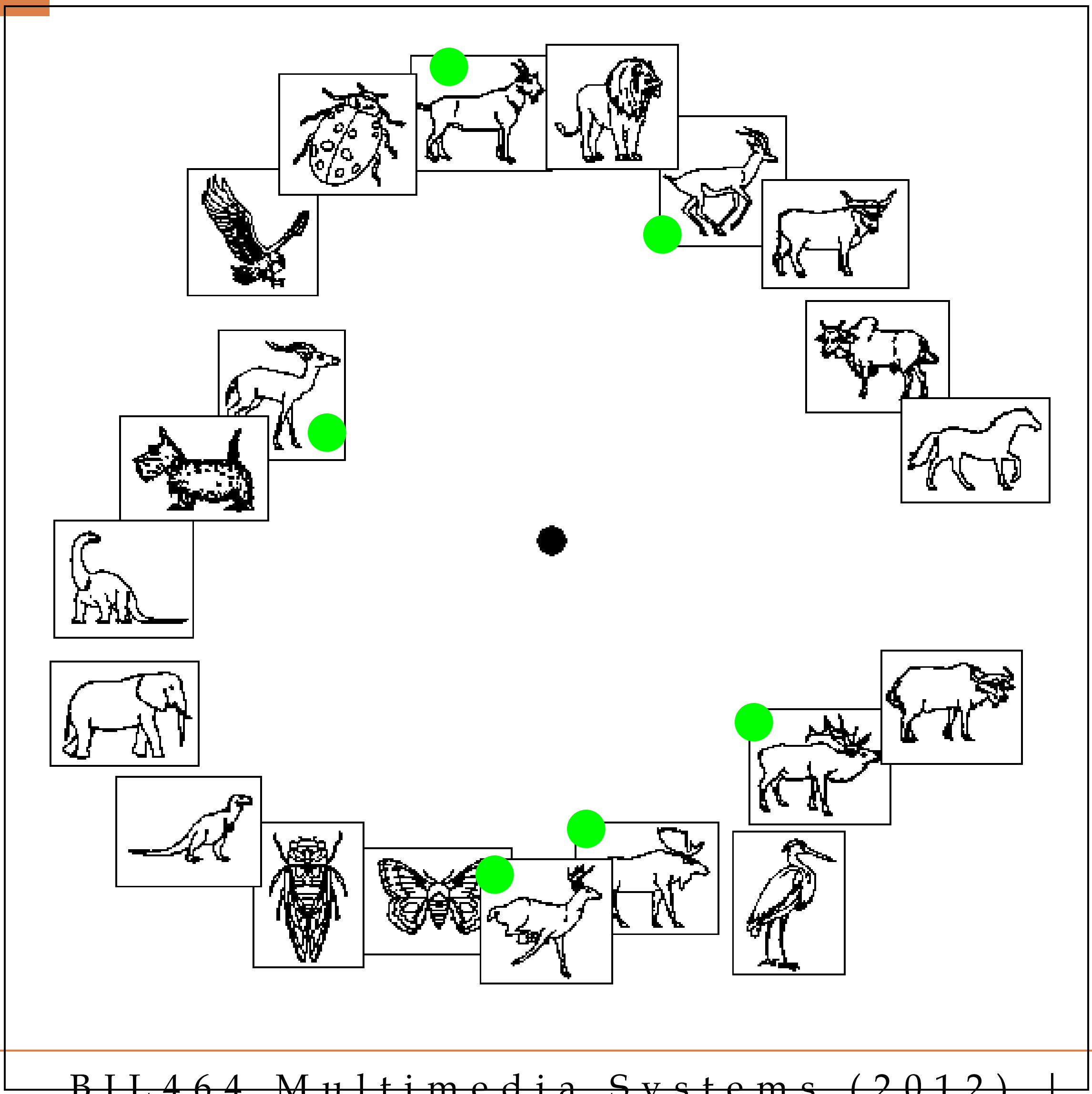
initial result

query

User action



After relevance feedback



number of relevant images has doubled

Evaluation of Results

- How do we measure performance of a retrieval system?
- Recall is the percentage of relevant documents retrieved out of all relevant documents.

$$Recall = \frac{Desired\ images\ returned}{All\ desired\ images}$$

- Precision is the percentage of relevant documents retrieved compared to the number of all the documents retrieved.

$$Precision = \frac{Desired\ images\ returned}{All\ retrieved\ images}$$

- These measures are affected by the database size and the amount of similar information in the database, and as well they do not consider fuzzy matching or search result ordering.

Open Research Directions in MIR

- Bridge the Semantic Gap
  - high level concept (sites, objects, events) and low-level visual/audio features (color, texture, shape and structure, layout; motion; audio – pitch, energy, etc.).
- How to Best Combine Human Intelligence and Machine Intelligence.
  - Keep human in the loop, e.g. Relevance Feedback
- New Query Paradigms
  - Query by keywords, similarity, sketching an object, sketching a trajectory, painting a rough image, etc. Can we think of useful new paradigms?

Open Research Directions in MIR – *Cont.*

- Multimedia Data Mining
  - Searching for interesting/unusual patterns and correlations in multimedia has many important applications, including Web Search Engines and dealing with intelligence data.
  - Work to date on Data Mining has been mainly in Text data.
- How to Use Unlabeled Data
  - Active learning, e.g., in Relevance Feedback
  - Label propagation, e.g., image/video annotation

Current Image Search Systems

- Some well-known current image search engines are listed as follows:
  - QBIC (Query By Image Content)
  - UC Santa Barbara Search Engines
  - Berkeley Digital Library Project
  - Chabot
  - Blobworld
  - Columbia University Image Seekers
  - Informedia
  - MetaSEEk
  - Photobook and FourEyes
  - MARS
  - Virage
  - Viper
  - Visual RetrievalWare