I-SEARCH
A Unified Framework for Multimodal Content Search

Dimitrios Tzovaras
CERTH-ITI
I-SEARCH Project Coordinator

Thomas Steiner
Google, Germany
The mission

• To provide a novel unified framework for multimodal content indexing, sharing, search and retrieval

• To handle specific types of multimedia and multimodal content (text, 2D image, video, 3D objects and audio) along with real world and emotional information

• To develop a user-centric search-engine (only the content of interest will be delivered to the end-users)

• To optimally present the retrieved results to the user via novel visualization schemes

• To dynamically adapt the search engine to end-user’s device (varying from a simple mobile phone to a high-performance PC)
What is STREP I-Search?

Objectives

- Research and development of an innovative Rich Unified Content Description (RUCoD)
- Development of intelligent content interaction mechanisms so that only the content of interest will be delivered to the users
- Provide a novel way for the presentation of the multimodal data retrieved by the search engine, by utilizing visual analytics technologies
The aim of the I-SEARCH project is the development of a search engine able to handle multimedia and multimodal content (text, 2D image, video, 3D objects, audio and combination of the above), which can be used as queries and retrieve any available relevant content of any of the aforementioned types.
I-Search Use Cases

UC1 - Music retrieval through expressive embodied queries
UC2 - Social music retrieval through expressive embodied queries
UC3 - Search and retrieval of a piece of furniture
UC4 - Search and retrieval of multimedia content using the end-user’s smartphone
UC5 - Personalized approach to customer needs
UC6 - 3D game component retrieval
UC7 - Game avatar retrieval
I-Search Architecture

Layer 1: Descriptor Extraction - RUCoD Formulation

- **Networked Media**
  - L(low-level): Content-related Descriptors
- **Real-world Information**
  - R(real World): RW-related Descriptors
- **Emotional, Expressive, Social Information**
  - U(user): User-related Descriptors

Multimodal Annotation Propagation

RUCoD: [L, R, U]

Layer 2: Interaction

- **Recommendations**
- **Individual/Social Relevance Feedback**

Layer 3: Visualisation

- **Visual Analytics Technologies**
  - Data Management
  - Data Analysis
  - Data Visualisation

User Interfaces
RuCoD Specifications

<table>
<thead>
<tr>
<th>Header</th>
<th>Low-level descriptors</th>
<th>Real World descriptors</th>
<th>User-related descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification</td>
<td>Version</td>
<td>Length</td>
<td>L-part</td>
</tr>
<tr>
<td>R-Header</td>
<td>Time</td>
<td>Position (GPS/Galileo, etc.)</td>
<td>Weather (temperature, sensors, etc.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low-level descriptor</th>
<th>User-related descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Header</td>
<td>Expressive / Emotional</td>
</tr>
<tr>
<td>Text-based descriptors</td>
<td>Social / Collaborative</td>
</tr>
<tr>
<td>Image Descriptors</td>
<td>...</td>
</tr>
</tbody>
</table>
The Concept of Content Object (CO)

Content Object: a container of multimedia objects (Images, 3D objects, audio) that carry the same semantics information.

My_Dog

3D Descriptors:
- 0.9439
- 0.0498
- 0.6849
- 0.1346

2D Descriptors:
- 1.1366
- 0.0224
- 0.8209
- 0.1195

Audio Descriptors:
- 1.0552
- 0.0524
- 0.6994
- 0.2062
Multimodal Retrieval Example

Image Similarity

Multimodal Similarity

CO Dataset

Query CO

CO A1

2D

CO A2

3D

CO A3

3D

CO B1

2D

CO B2

3D

2D
Construction of Multimodal Feature Space

Multimedia DB

- CO₁
- CO₂
- ... COₙ

Content Objects

- 3D Object
- 2D Image
- Sound

Content Object - 1

3D Object

3D Object Descriptor Extractor

3D Object Descriptor set

2D Image

2D Image Descriptor Extractor

2D Image Descriptor set

Sound

Sound Descriptor Extractor

Sound Descriptor set

Multimedia Indexing

Construction of n x n Adjacency Matrix

Laplacian Eigenmaps

Multimodal Descriptor-1
Multimodal Descriptor-2
...
Multimodal Descriptor-n
Multimodal Feature Space

**CATEGORY A**
- \( CO_{A1} \)
- \( CO_{A2} \)
- \( CO_{A3} \)
- \( CO_{A4} \)

**CATEGORY B**
- \( CO_{B1} \)
- \( CO_{B2} \)
- \( CO_{B3} \)

**Overall Dissimilarity between 2 COs**

**LE: Laplacian Eigenmaps**

Search Computing Event, Brussels, 25-26 September 2012
Results Presentation

Menu to switch modalities

Stack effect denotes multi-modality

Tooltip popup on hover (or tap)

Summary text

Search Computing Event, Brussels, 25-26 September 2012
Results Presentation

- Images: thumbnails
- 3D: rendered image thumbnails (look-around effect)
- Video: slideshow of keyframes
- Audio: generation of synthetic visual previews from aural features.
Audio Visualisation

- MFCC spectral features
  - Dimensionality Reduction
  - Parametric shape generation

Parameters:
- shape and number of internal/external petals +
- internal/external colors
Cluster Visualisation

Classic

Hyperbolic Tree

Panel

Treemap

Search Computing Event, Brussels, 25-26 September 2012
Thumbnail arrangement

Grid

Smart

Smart Grid

Multiple pages
Real-world info

Location

Time
Timeline of documents

The Fender Precision Bass (often shortened to "P Bass") is an electric bass designed by Leo Fender and brought to market in 1951. It is one of Fender's most popular bass models and has been used by many famous musicians.

Location: 12 Flatholm Street, Portsmouth, PO2, UK.
Adaptation

Filtering (modality, tags, date, time, relevance)

Query settings

Clustering

Thumb size

Thumb layout
Advanced interactive feedback

Exploit user feedback to provide more informative visualisations
MuSeBag - Multimodal Search Interface

- To support: Audio, Video, 3D, Sketch, Image, Real world data, texts
- Tokens for different modalities representation
- HTML5 based
- Adaptiveness and Device independence
I-Search Demonstrators

Demonstrators

- I-SEARCH User Interface
  [http://isearch.ai.fh-erfurt.de](http://isearch.ai.fh-erfurt.de)

- CoFetch tool for semi-automatic creation of Content Objects (HSF)

- Multimodal Search (CERTH)
  [http://vcl.iti.gr/is/isearch/client/musebag/](http://vcl.iti.gr/is/isearch/client/musebag/)

- Demo presentation of UC3 - Furniture retrieval (CERTH)
Thank you for your attention!