



Multimedia semantic analysis technologies and their potential uses

Industry Day SAMT 2006

Yiannis Kompatsiaris
Multimedia Knowledge Group
CERTH - Informatics and Telematics Institute



samt
2006

Evolution of Content

- 1-2 exabytes (millions of terabytes) of new information produced world-wide annually
- 80 billion of digital images are captured each year
- Over 1 billion images related to commercial transactions are available through the Internet
- This number is estimated to increase by ten times in the next two years.
- 4 000 new films are produced each year
- 300 000 world-wide available films
- 33 000 television stations and 43 000 radio stations
- 100 billions of hours of audiovisual content



Personal Content



Sport - News

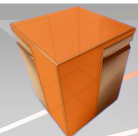


AllPosters.com
OVER
100,000
Posters & Prints
Click Here
To Buy Now!
The World's Largest
Poster and Print Store!

Web
Mobile

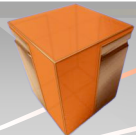


Movies



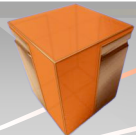
Need for annotation + medatata

“The value of information depends on how easily it can be found, retrieved, accessed, filtered or managed in an active, personalized way”



Text-based indexing

- **Manual annotation**
 - + Straightforward
 - + High/Semantic level
 - + Efficient during content creation
 - Most commonly used
 - Necessary in a number of applications
 - - Time consuming
 - - Operator-application dependent
 - - Text related problems (synonyms etc)
- **Annotation using captions and related text**
 - **Web, Video, Documents etc**
 - + Straightforward
 - + High/Semantic level
 - + Multimodal approach
 - - Text processing restrictions and limitations
 - - Captions must exist



Text-based indexing

Google Αναζήτηση: hockey - Microsoft Internet Explorer

Αρχείο Επεξεργασία Προβολή Αγαπημένα Εργαλεία Βοήθεια

Πίσω - Αναζήτηση - Αγαπημένα - Μέσα









http://images.google.com.gr/images?hl=el&lr=&ie=UTF-8&q=hockey

Google hockey Search Web 390 blocked AutoFill Options hockey

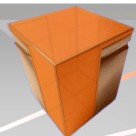
Google Εικόνες

hockey Αναζήτηση Σύνθετη Αναζήτηση Ρυθμίσεις

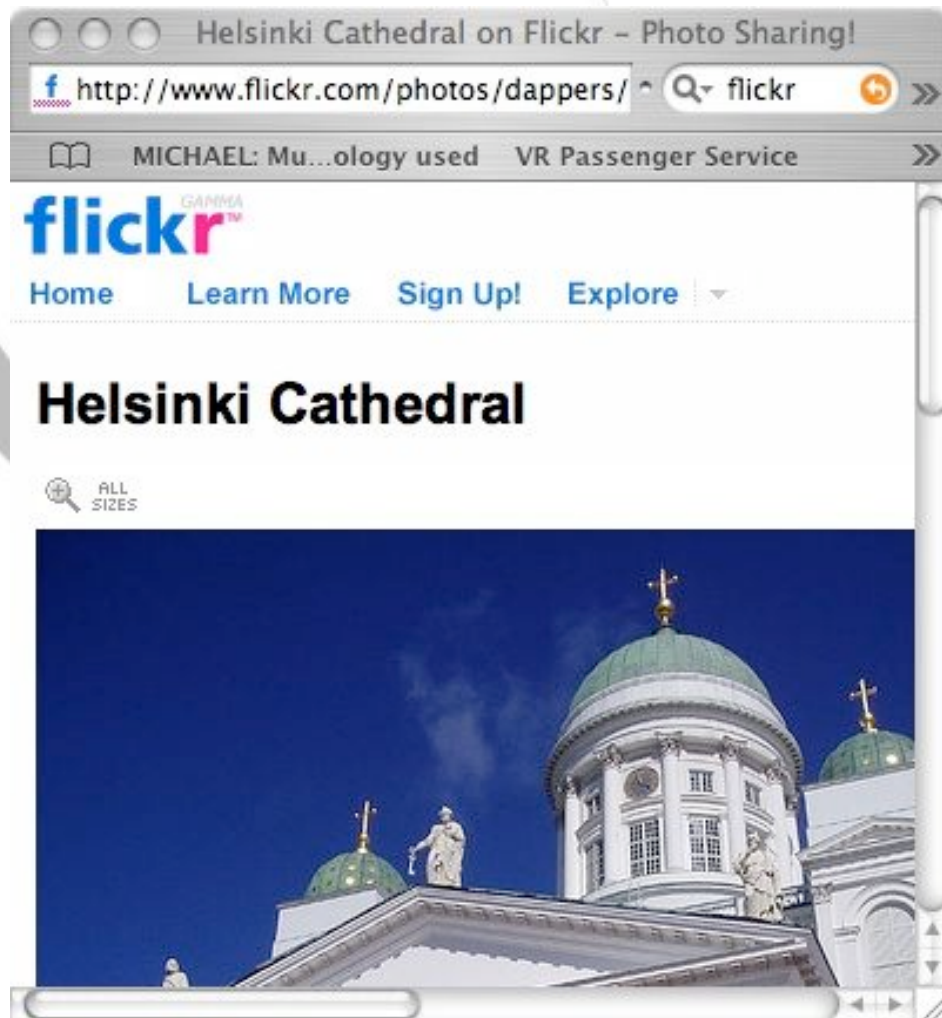
Εικόνες Αποτελέσματα 1 - 20 από περίπου 577.000 για **hockey**. (0,10 δευτερόλεπτα)
Εμφάνισε: Όλες οι διαστάσεις - Μεγάλο - Μεσαίο - Μικρό

 <p>hockey goalie wallp 1024.jpg 1024 x 768 pixels - 220k www.paralympic.ca/english/products/wallpapers...</p>	 <p>hockey wallp 1024.jpg 1027 x 768 pixels - 126k www.paralympic.ca/.../hockey%20wallp%201024.jpg</p>	 <p>a-harris.jpg 150 x 250 pixels - 12k umterps.ocsn.com/.../md-w-fieldh-body.html</p>	 <p>a-Falgowski1107.jpg 150 x 250 pixels - 7k tarheelblue.ocsn.com/.../unc-w-fieldh-body.html [Περισσότερα αποτελέσματα από το graphics.fansonly.com]</p>
 <p>Roller Hockey Ball 006.jpg 800 x 604 pixels - 55k www.soton.ac.uk/~rollhock/autoGallery/Ball200...</p>	 <p>hockey player.jpg 300 x 300 pixels - 48k www.merrittsbakery.com/.../hockey%20player.jpg</p>	 <p>bertuzziapunch.jpg 410 x 328 pixels - 27k www.barse.org/blog/jack/bertuzziapunch.jpg</p>	 <p>HOCKEY.jpg 1024 x 768 pixels - 439k www.torino2006.org/.../WALLPAPER_1024/HOCKEY.jpg</p>

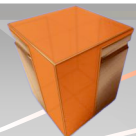
Internet



Collaborative tagging



e.g. Flickr
Web 2.0 applications



Addressing the Semantic Gap

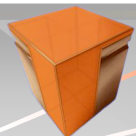
- **Semantic Gap** for multimedia: To map automatically generated numerical low level-features to higher level human-understandable semantic concepts

```
<?xml version='1.0' encoding='ISO-8859-1' ?>
<Mpeg7 xmlns...>
  <DescriptionUnit xsi:type = "DescriptorCollectionType">
    <Descriptor xsi:type = "DominantColorType">
      <SpatialCoherency>31</SpatialCoherency>
      <Value>
        <Percentage>31</Percentage>
        <Index>19 23 29 </Index>
        <ColorVariance>0 0 0 </ColorVariance>
      </Value>
    </Descriptor>
  </DescriptionUnit>
</Mpeg7>
```

Dominant Color Descriptor of a
sky region

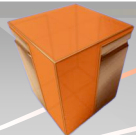


This image
contains a sky
region and is a
holiday image



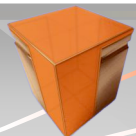
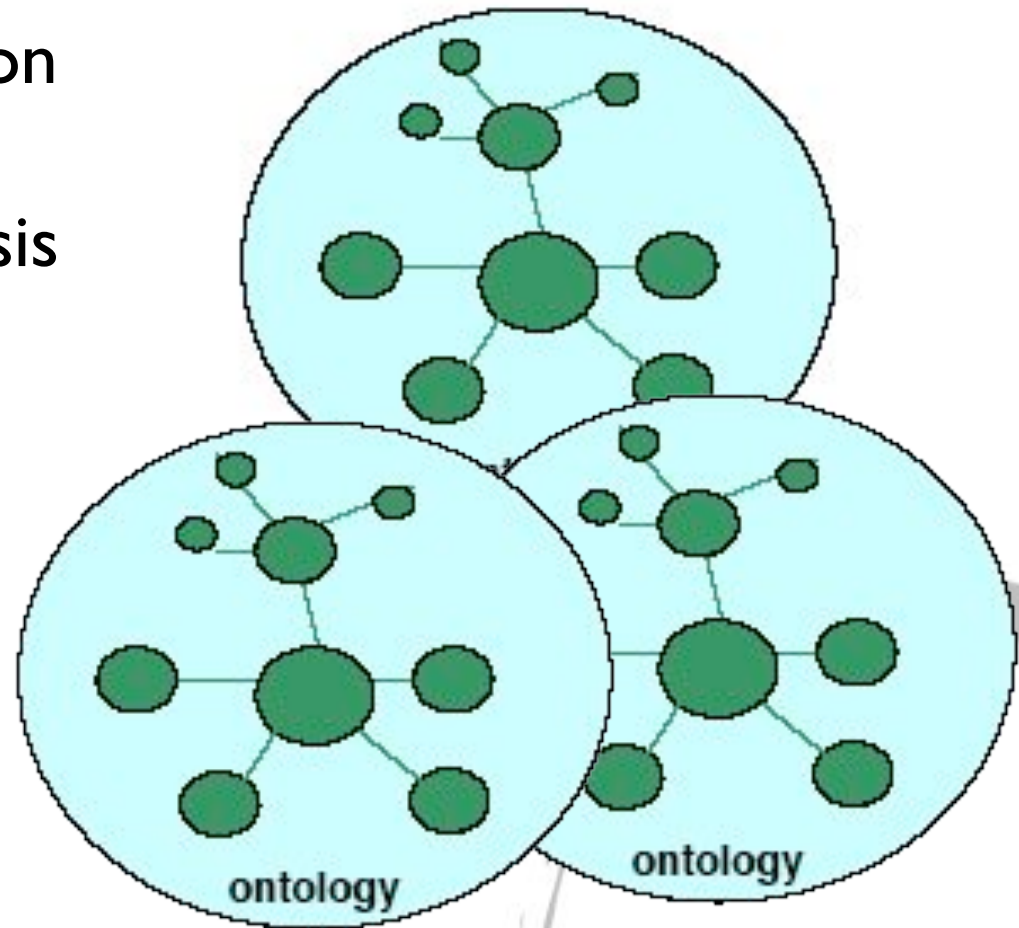
Problem definition

- **Semantic image analysis:** *how to translate the automatically extracted visual descriptions into human like conceptual ones*
- **Low-level features** provide **cues** for *strengthen/weaken evidence based on visual similarity*
- **Prior knowledge** is needed to support *semantics disambiguation*



Use of ontologies

- Metadata representation
 - interoperability
- Ontology-driven analysis
- Reasoning
 - Extracting higher-level annotations
- Retrieval
- Personalization
- Semantic Web



Indexing using Low-Level Visual Features

- **Low-level features (color, texture, shape, edges, motion, etc)**
 - + automatic extraction
 - + computation efficiency
 - Suitable for many applications
 - - not semantic
 - - algorithm complexity
- representation
 - features
 - color, texture space
 - invariance
 - compactness
- indexing (MPEG-7)
- database
- matching – distance
- global – local features (segmentation)



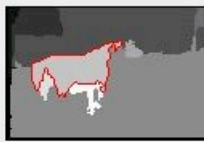
Region-Based Query-by-Example

Results:

25 items total [1 - 9]. Search completed in 4.660 secs...



Original Image



PFZL (TUM)



113025.jpg 99.89%
192x128px, 8KB
[Find Similar](#)



113029.jpg 82.27%
192x128px, 8KB
[Find Similar](#)



113040.jpg 81.86%
192x128px, 6KB
[Find Similar](#)



113001.jpg 81.37%
192x128px, 7KB
[Find Similar](#)



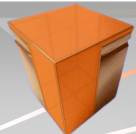
113036.jpg 80.89%
192x128px, 8KB
[Find Similar](#)



113030.jpg 80.25%
192x128px, 7KB
[Find Similar](#)

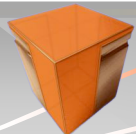
SCHEMA MPEG-7
XM based
Reference System

<http://media.itl.gr/site/SchemaXM>



Query-by-Example Application (Like.com)

The screenshot displays the Like.com search interface. At the top, there are three filter sections: "Which aspects are most important?" with sliders for Color, Shape, and Pattern; "Any Specific Color?" with a color palette; and "What is your price?" with a price range set to \$50 and an "Apply" button. Below these is a "Which part of this image do you like?" section with a drawing tool and a reference image of a brown leather high-heeled boot. The main area shows a grid of search results for similar boots, each with a "Likeness Search" button, a product name, price, and a "Buy at" button. A large image on the right shows a woman in a dress and boots on a red carpet, with a white box highlighting her boots and a blue "Likeness Search" button overlaid on it.



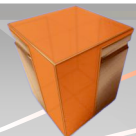
Scene Classification

- **General concept-based classification (people-no people / indoors – outdoors)**
- + automatic
- + computational efficient
- + semantic classes
- Appropriate for a number of applications
- - training and classification limitations
- - predefined restricted classes

People – No People

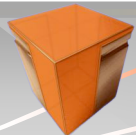


Outdoor - Indoor



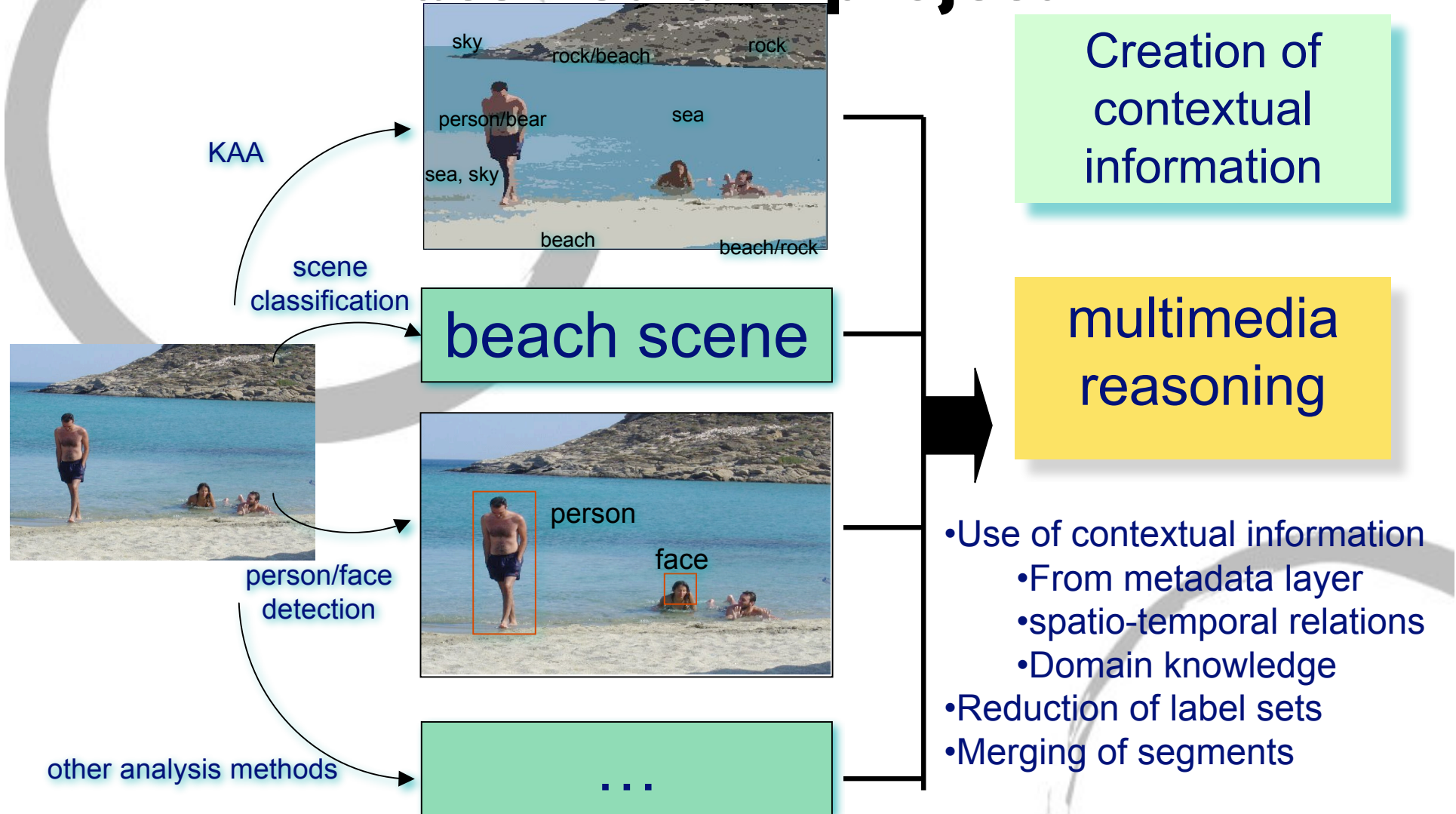
Semantic Analysis

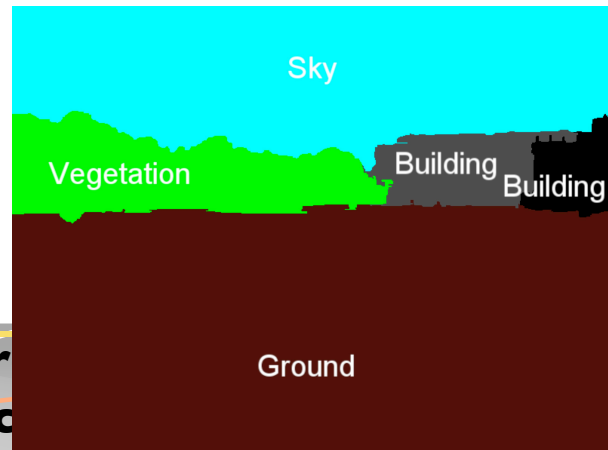
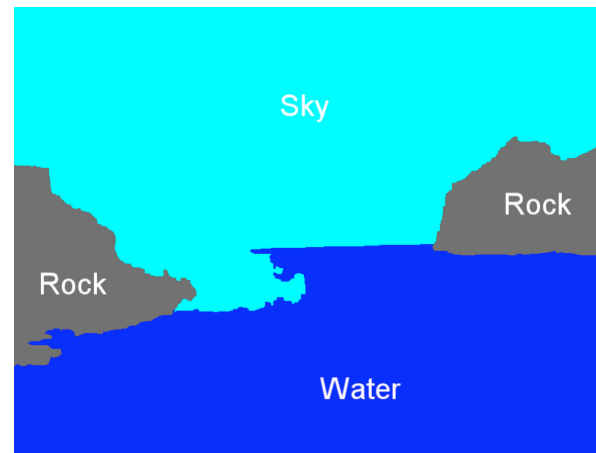
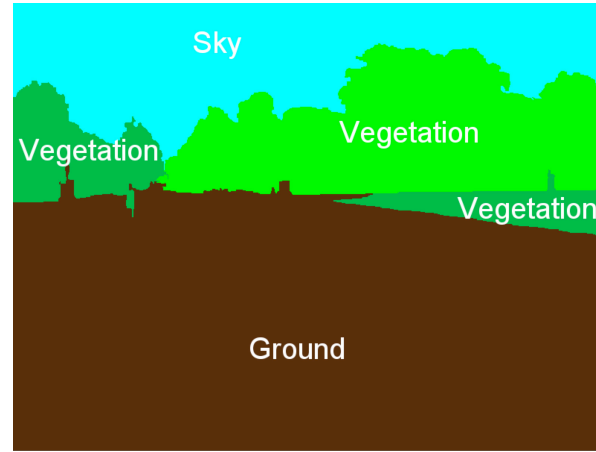
- Aims to develop automated methods for semantic annotation of multimedia content
 - LL features are analyzed to recognize **objects** and **events**
- Object/Events/Relationships **knowledge** is needed
- Techniques for knowledge extraction and representation
 - knowledge base
 - Learning techniques, classification, pattern recognition (implicit knowledge)
 - Model-based techniques (explicit knowledge)
- Specific domains (e.g. sports, news)
- **Multimodal, context-assisted** approaches are usually followed (e.g. audio-assisted video analysis: goal detection)



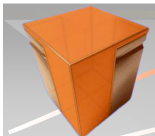
Context and Reasoning for Analysis

aceMedia IP project



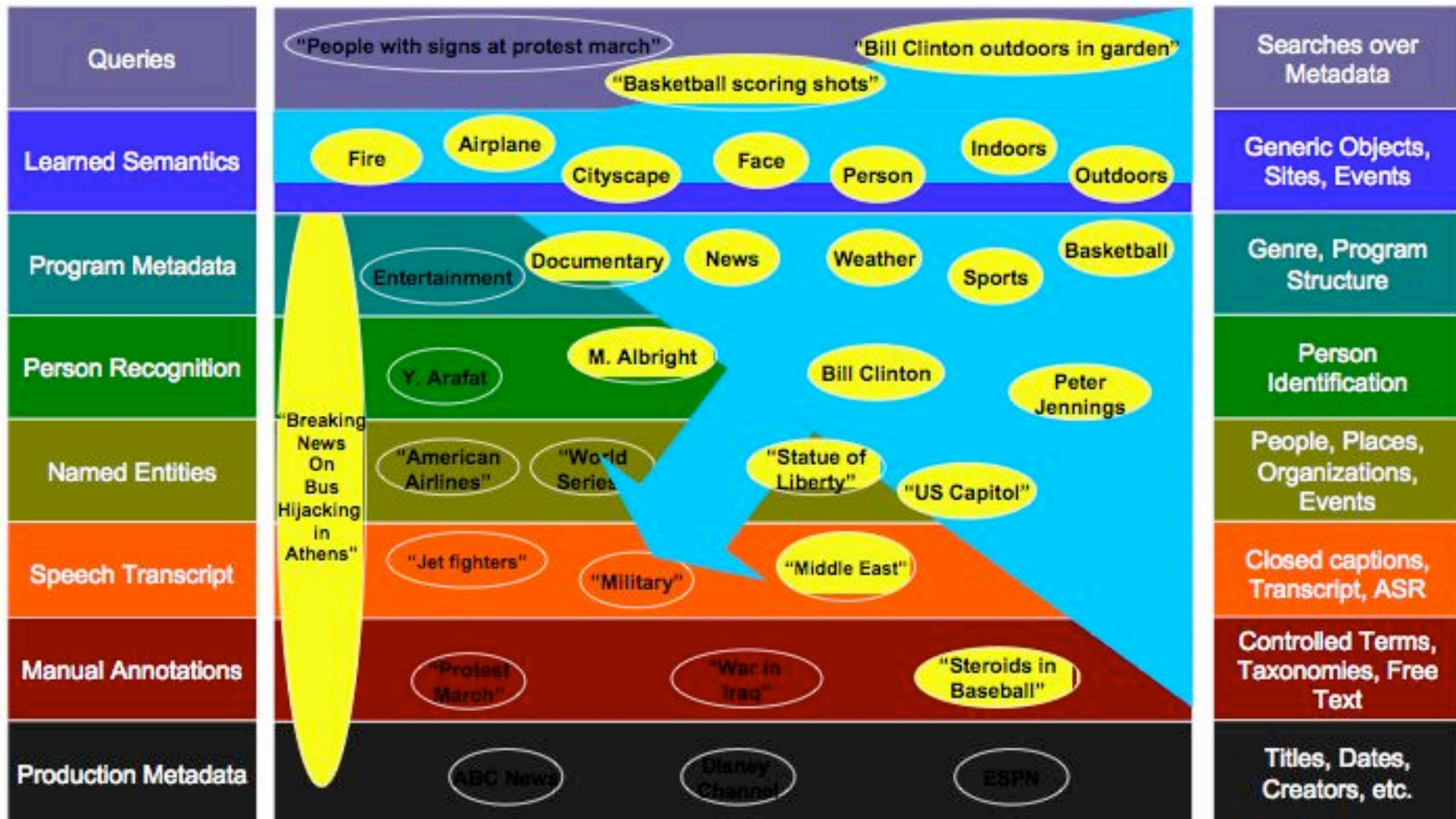


Automatic annotation of holiday images

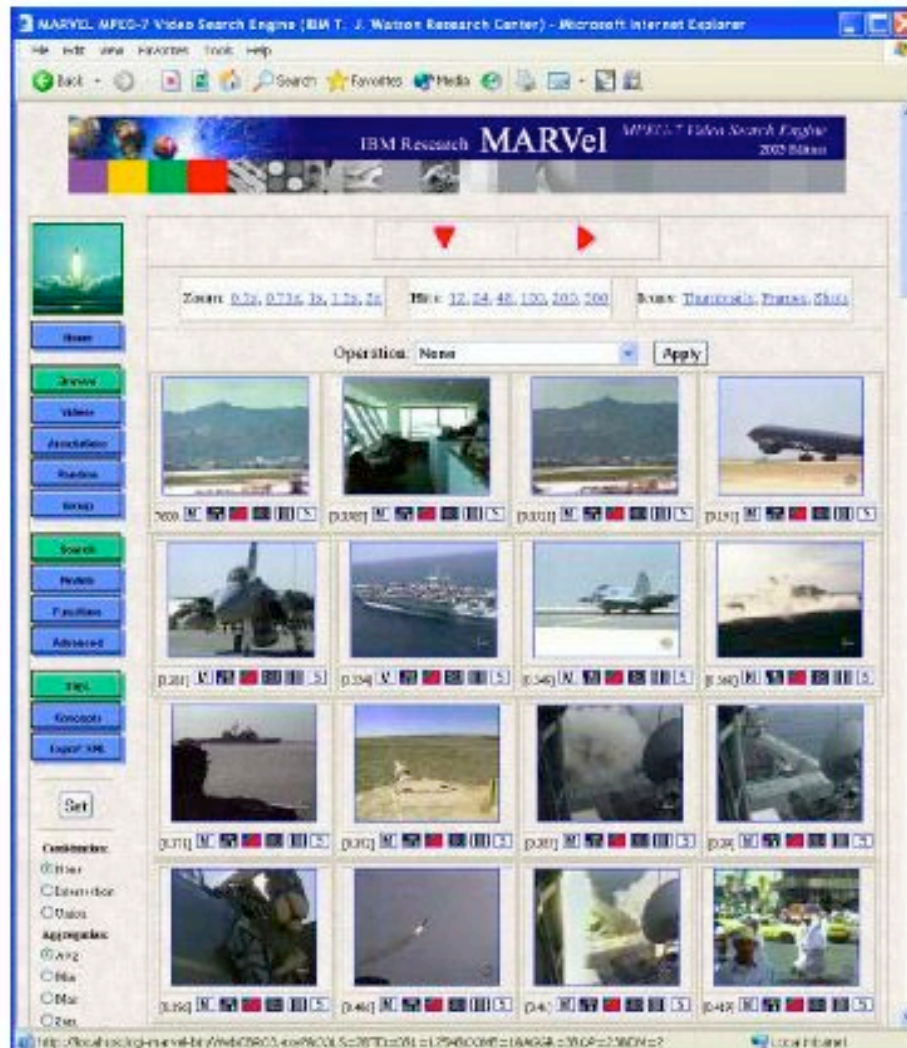


Gr
atic

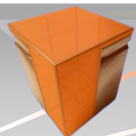
Level of automation keeps increasing (IBM-TRECVID)



IBM Marvel



- MPEG-7 Video Search Engine
- Automatic indexing:
 - Shot detection/key-frame extraction
 - Feature Extraction
 - Semantic Concept Detection
- Search methods:
 - **Model-based retrieval (MBR)** – statistical modeling and detection of semantic concepts - faces, people, outdoors, etc.
 - **Content-based retrieval (CBR)** - color, texture, edges, etc.
 - **Text-based retrieval (TBR)** – textual metadata, annotations, speech transcript
 - **Model-vector based retrieval (MVBR) = MBR + CBR**
- Interaction:
 - Multi-example relevance feedback searching
 - * Iterative searching (combination methods and aggregation functions)
- On-line demo:
 - <http://mp7.watson.ibm.com>



ALIPR™

Automatic Linguistic Indexing of Pictures - real time

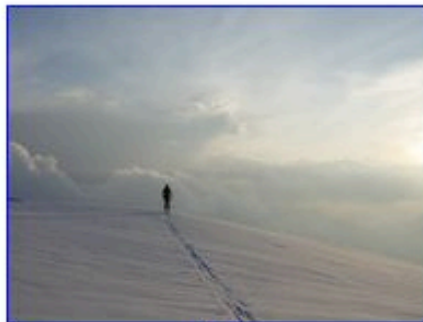
Image Upload no file selected

Or Image URL

http://..... - Try drag and drop from another website.

Keywords - [Vote](#) - [Most Voted](#) - [Random](#) - [My Pictures](#) **NEW**

or begin search with: [rock](#) [purple](#) [decoration](#) [bridge](#) [rose](#) [desert](#) [cloth](#) [dance](#) [old](#) [mouse](#) [leaf](#) [child](#) [fish](#) [vegetable](#) [poppies](#)



snow



flower plant rose

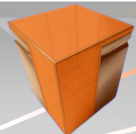


bird chicken toy



man-made face art composite
fluffy white dog unnatural
animal

©2006 alipr.com Patent Pending. [About Us](#) / [Why Alipr](#) / In the News: [Discovery](#), [MIT Tech Review](#), [Scientific American](#) / [Press Relations](#)
Do NOT upload objectionable images. Your images can be viewed by others.



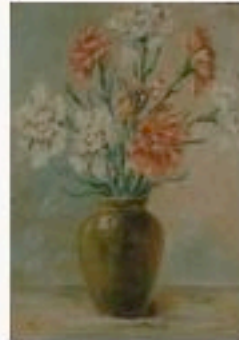
ITI REACH

<http://reach.iti.gr>



ΚΑΚΤΟΣ

[Εύρεση Ομοίων](#) [Πληροφορίες](#)
[Αγορά Συναφούς](#) [Προσθήκη σε](#)
[Υλικού](#) [προσωπικό](#)
[κατάλογο](#)



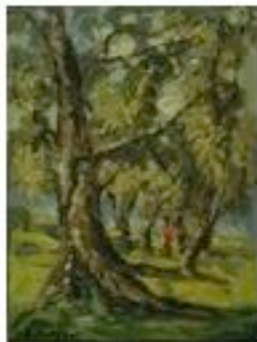
ΒΑΖΟ ΜΕ ΓΑΡΥΦΑΛΛΑ

[Εύρεση Ομοίων](#) [Πληροφορίες](#)
[Αγορά Συναφούς](#) [Προσθήκη σε](#)
[Υλικού](#) [προσωπικό](#)
[κατάλογο](#)



ΑΝΘΗ

[Εύρεση Ομοίων](#) [Πληροφορίες](#)
[Αγορά Συναφούς](#) [Προσθήκη σε](#)
[Υλικού](#) [προσωπικό](#)
[κατάλογο](#)



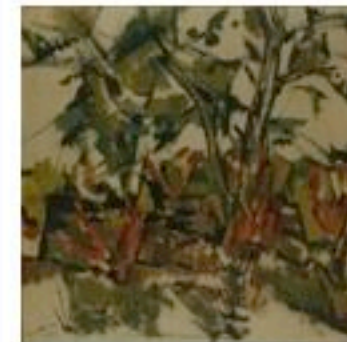
ΤΟΠΙΟΓΡΑΦΙΑ

[Εύρεση Ομοίων](#) [Πληροφορίες](#)
[Αγορά Συναφούς](#) [Προσθήκη σε](#)
[Υλικού](#) [προσωπικό](#)
[κατάλογο](#)



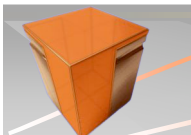
ΑΓΡΟΤΟΣΠΙΤΟ

[Εύρεση Ομοίων](#) [Πληροφορίες](#)
[Αγορά Συναφούς](#) [Προσθήκη σε](#)
[Υλικού](#) [προσωπικό](#)
[κατάλογο](#)



ΤΟΠΙΟ ΜΕ ΔΕΝΔΡΑ

[Εύρεση Ομοίων](#) [Πληροφορίες](#)
[Αγορά Συναφούς](#) [Προσθήκη σε](#)
[Υλικού](#) [προσωπικό](#)
[κατάλογο](#)



ITI REACH

Automatic query generation and recommendation



ΠΑΡΑΓΚΕΣ ΣΤΗΝ ΚΑΤΕΡΙΝΗ
[Εύρεση Ομοίων](#) [Πληροφορίες](#)



ΚΑΚΤΟΣ
[Εύρεση Ομοίων](#) [Πληροφορίες](#)



ΛΕΥΚΟΣ ΠΥΡΓΟΣ
[Εύρεση Ομοίων](#) [Πληροφορίες](#)



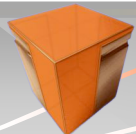
ΑΡΚΟΥΔΕΣ
[Εύρεση Ομοίων](#) [Πληροφορίες](#)



ΣΤΟΝ ΤΑΦΟ ΤΗΣ ΜΑΝΑΣ
[Εύρεση Ομοίων](#) [Πληροφορίες](#)



**ΟΔΙΠΟΡΟΣ ΠΟΥ
ΞΕΚΟΥΡΑΖΕΤΑΙ**
[Εύρεση Ομοίων](#) [Πληροφορίες](#)



use case / domain



news video



security

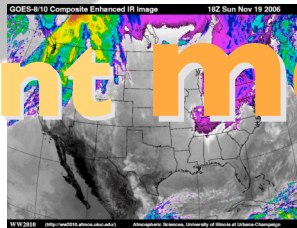


surveillance

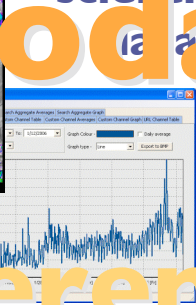
different requirements



personal photos



scientific data



medical images

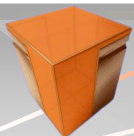
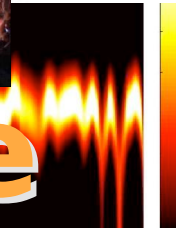
different modalities



sports



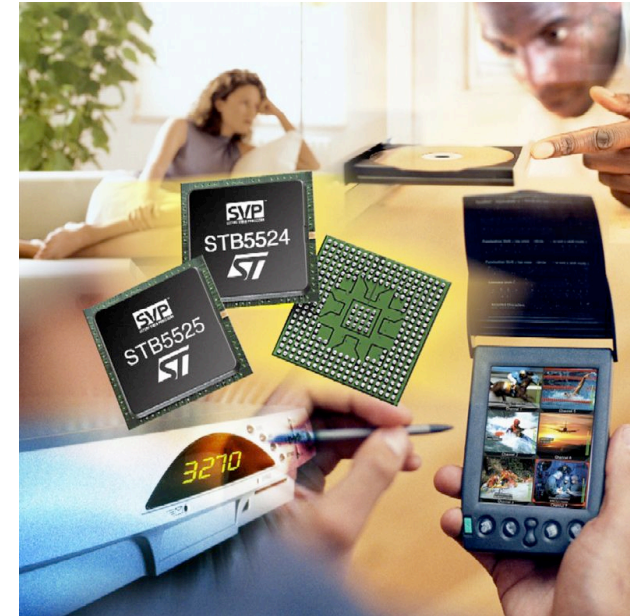
different knowledge



Users - Applications

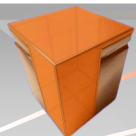
Audiovisual Sector (TV stations, digital - subscriber TV, advertisement, production, designers, photographers, etc)

- *Adaptation, summarization*
- *«I want images of the Prime Minister»*
- *«I want all goals of the national team in this game»*



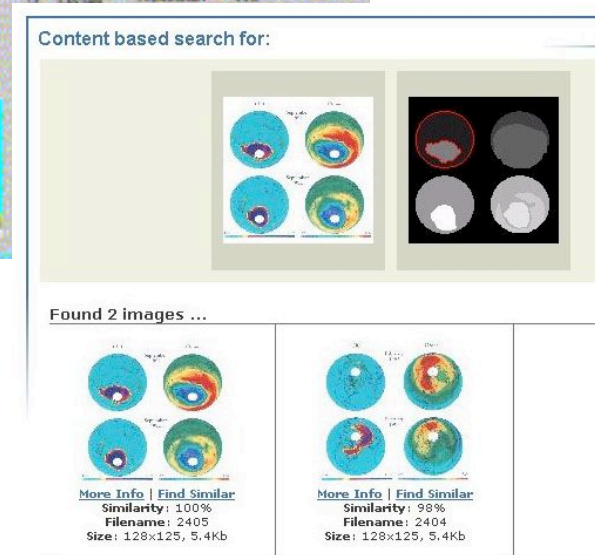
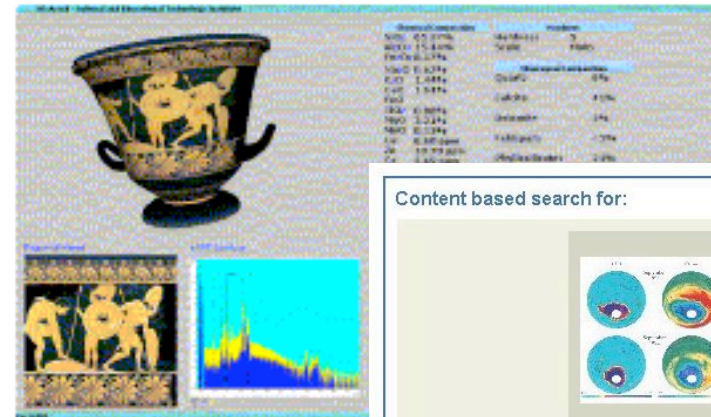
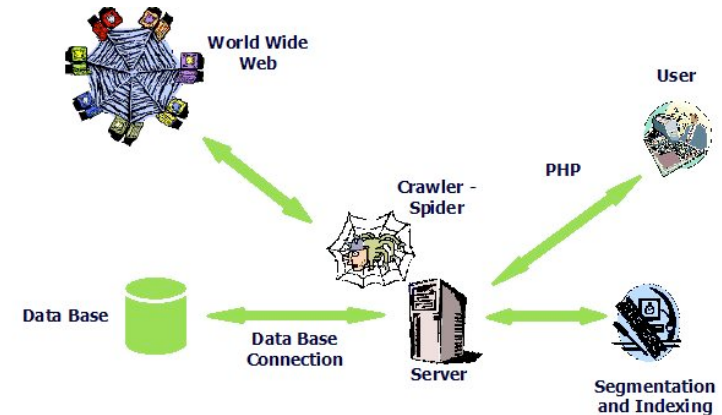
Mobile Sector

- *Personalized content transmission*
- *Recommendations*



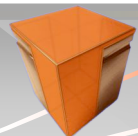
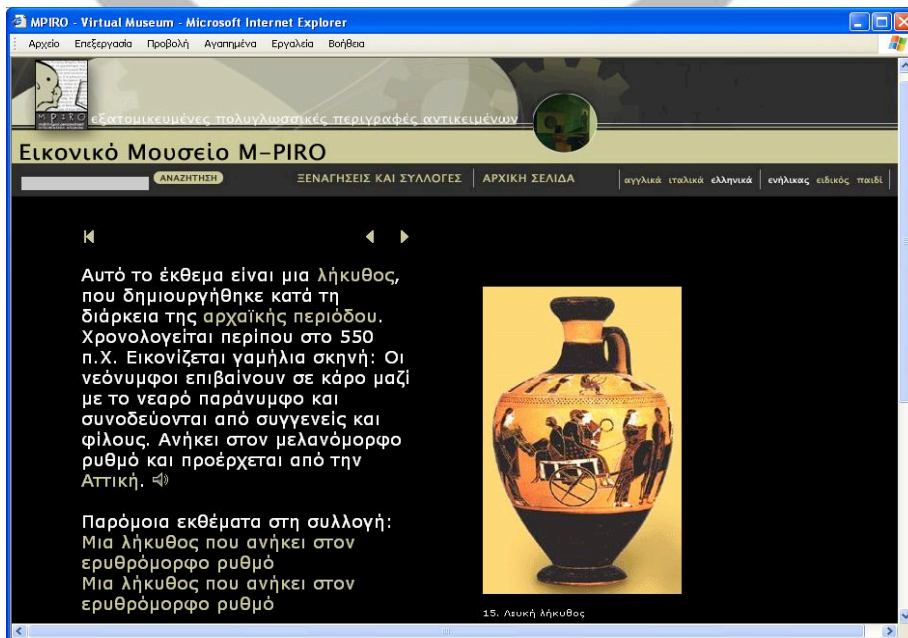
Users - Applications

- Search Services (Portals, news sites, libraries, museums, companies, etc)



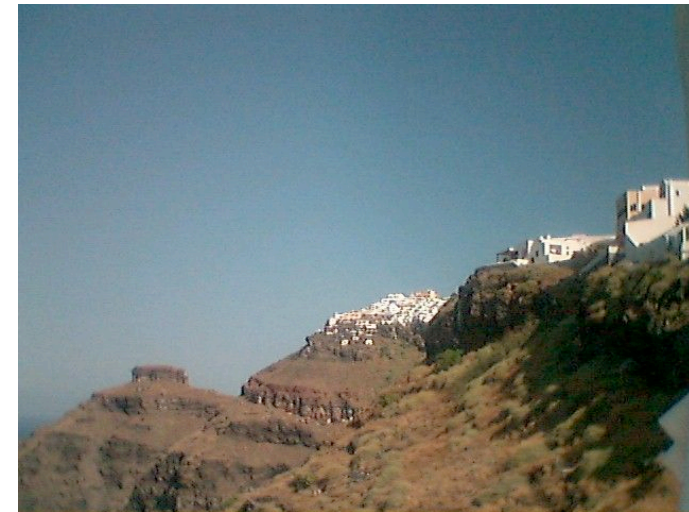
Medical Applications

Cultural Applications

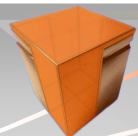


Users - Services

- Multimedia and Semantic Web
- Personal users and content:
 - «I want photos from my vacations in the Greek islands»
 - «I want video-clips from the Web with sports cars»



A screenshot of a Microsoft Internet Explorer browser window. The address bar shows 'http://kom/video1/start'. The page title is 'Object-based Video Retrieval using Ontologies' and the subtitle is 'Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki & Informatics and Telematics Institute, Centre for Research and Technology Hellas'. The main content area displays 'Query results for "BlueCar": objects 1 to 20 of 51'. Below this, there is a grid of 12 video thumbnails, each with a label like '1. VID-1334 SID:3' and a 'relevant' checkbox. On the left side of the page, there is a 'SCHEMA' section with text about high-level concepts and an ontology. The browser's taskbar at the bottom shows several open applications and the system tray.



Conclusions

- Semantic analysis of multimedia is already providing results
- There is a **gap** between generic technologies and specific applications
- In many cases automatic analysis can enhance existing applications and not generate new ones
- Have to be integrated as part of a complete system or application
- A lot of factors have to be considered: **users**, interfaces, infrastructure, scenario, business model



Thank you!

CERTH-ITI / Multimedia Knowledge Group

<http://mkg.itigr>

