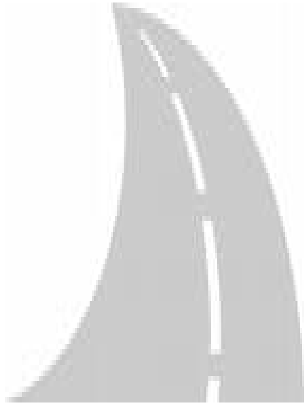
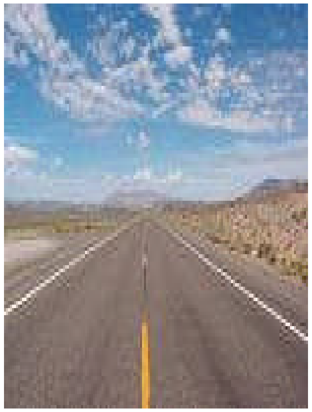
MMEDIA 2009
PANEL

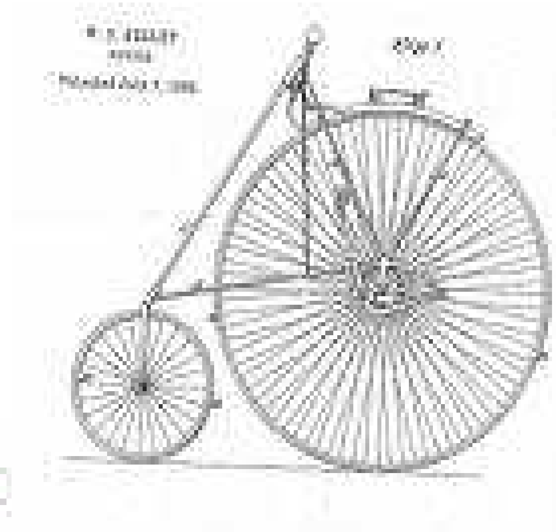
Thinking of a Picture and Finding it:
Evolving MMedia Computer
Interfaces

Road

-  

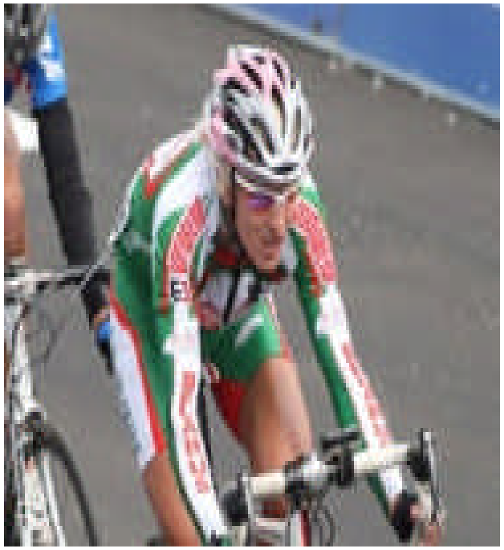
 

Bicycle



Tour de France

- http://www.youtube.com/watch?v=lmIJjR1kdYw
- Big crash 2007



Mental Query

- Thinking of all pictures with cycling accidents

  &
- Finding them

or

- Thinking of a Picture with a cycling accident

  &
- Finding it

Thinking + Finding

- Thinking [model, context, interface, etc.]
- Finding [picture/text, similarities, etc.]

Expert guests

- **Moderator:**
  Petre Dini, IARIA, USA / Concordia University, Canada

- **Expert panelists:**

- *Erwan Baynaud, Alcatel-Lucent France, Bell Labs, France*
- Laszlo Böszörmenyi, Klagenfurt University, Austria
- Dumitru Dan Burdescu, University of Craiova, Romania
- Philip Davies, BPC & Bournemouth University, UK

Evolving MMedia Computer Interfaces

Thinking of a Picture and Finding it

Erwan Baynaud

MMEDIA 2009 – July 20-25, 2009 – Colmar, France

Introduction

- Erwan Baynaud / Bell Labs Applications Domain / Alcatel-Lucent
- Thanks to Petre Dini for having invited me to participate.
- Question
  - ➔ Thinking of a Picture and Finding it!
- Methodology
  - ➔ Decomposition of the subject
  - ➔ Brainstorming around topics
  - ➔ Issues suggested

Mini-brainstorming

- Thinking
  - Thinking is something cerebral, isn't it?
  - Thinking mean to me a personal experience…
  - Oh… In fact we are trying to think together today. So let's say that it can also be a community experience.
- Picture
  - This is a kind of limitation when speaking of multimedia to only consider pictures.
  - Mainly because I spend my time to search around video applications!
  - Nevertheless, to use these media contents in a system, we have a need for indexation.
- Finding
  - To find anything on internet, I use Google.
  - It would be so easy if my stuff would be stored, sorted and labelled.
  - You can search for a long time before to find something into my girlfriend's handbag.

Considerations

- Thinking
  - Are we considering Brain Human interfaces only?

    (➔ I think I would start by simpler interfaces like voice or web-based for a demonstrator. :-P)
  - If not, questions are:
    - How to minimize the effort of the user to transcript his thoughts ? Which inputs can we use?
- Picture
  - Should we consider only pictures?

    ➔ We could consider other media contents and especially videos!

    ➔ Maybe also enlarging to things, thinks, dreams, contexts…
- Finding
  - MPEG-7 is probably the best standard to support picture's indexation.

    ➔ But who is using it? Is one standard the real solution? Or rather the manipulation of each system and descriptions through interfaces and dedicated engines.

Thinking of a Picture and Finding it!

cer nal

community

interface (Brain Human only?)

Adaptation system based on multimedia indexation

limitation

indexation

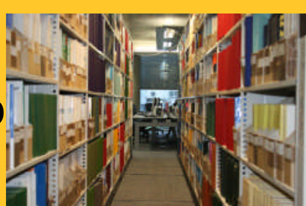
picture description (MPEG-7?)

video

my gi

Description matching engine

pictures and descriptions



Go

sorted and labelled

Adaptation system / Matching engine

- Thinking create the input
- Finding create the output

➔ What between?

- Adaptation: How to transcript the user request through the chosen interface to something being understandable by the system?
- Matching: System to correlate the description created with descriptions of the pictures in database.

➔ Keywords: user interfaces, Semantic, Indexation, Ontology…

To go further

- What are the works in progress about Brain-Human interfaces?
- How to describe a media content, an object, a context in a same way? Does anybody think to *an universal identification system* and later to an *universal format of description*? I mean not only tools to describe media contents but also to describe real things, thougths, situations.
- ➔ Maybe just create interfaces between all the existing descriptions?
- Which links can we envisage between real and virtual worlds?

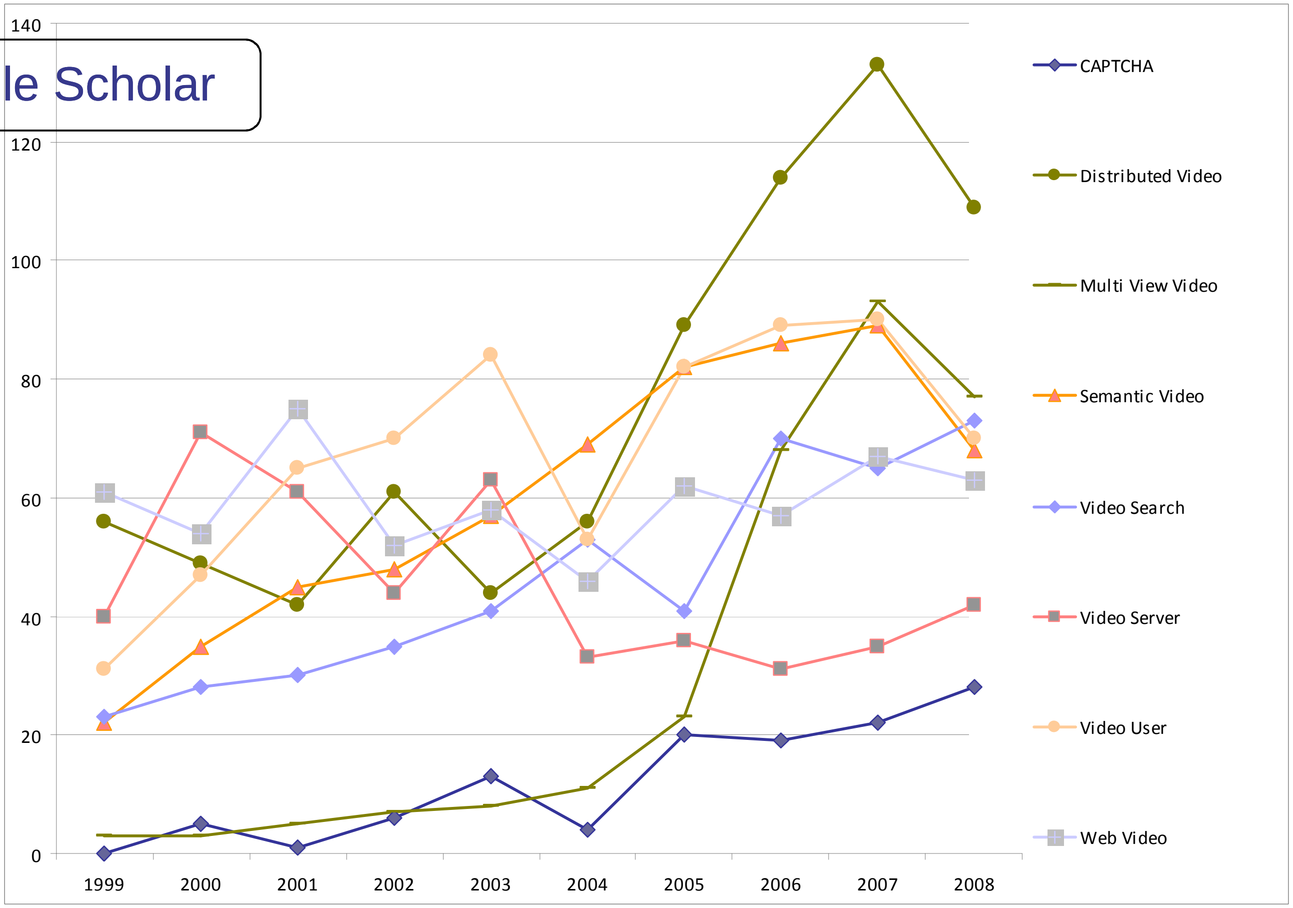
Thank you!

Is Multimedia as Pervasive as Graphics? Should it be?

Laszlo Böszörmenyi

Number of research papers in MM

from Google Scholar



A short history of computer graphics

- Up to the middle of the 70ties
  - ➢ Expensive, special hardware + software for special applications (e.g. CAD for architecture)
- End of 70ties the Alto computer (Xerox PARC)
  - ➢ Microcode supported pixel graphics
  - ➢ Cheap, software based solution (e.g. soft fonts)
- In the early 80ties
  - ➢ Graphics becomes pervasive
  - ➢ It was a process of years with many opponents
- A similar success story for Multimedia in sight?

Key Issues in the Success of Graphics

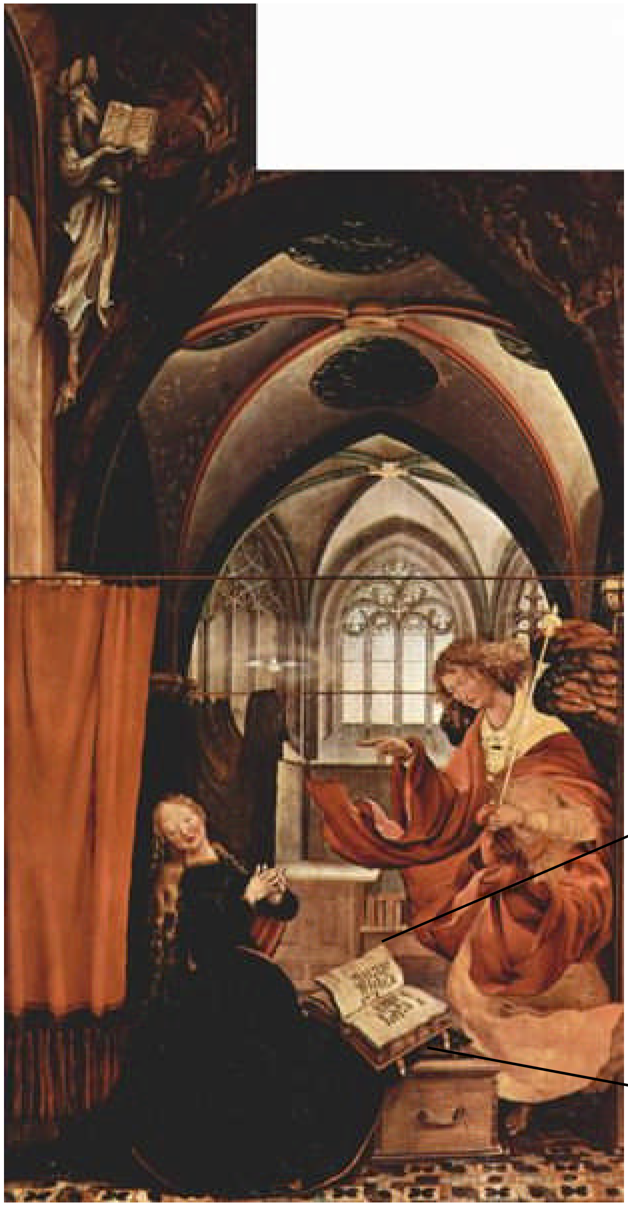
1. It was an integration of several technologies
2. Most of these were very innovative
   - Windows, mouse, object-orientation (reinvented) etc.
   - Successful business models
   - “Killer applications”, e.g. in text processing, slides
3. Took several years until it became pervasive
4. Strong enough to enforce changes in hardware, op. systems, middle-ware (X.11), applications
   - Applications had to be re-implemented, using an own *“window.redraw (width, height, parameters …)”*

Same Issues in Multimedia (incl. continuous)

1. Integration of much *more* technologies needed
   - Op. sys, networks, film, TV …
2. Innovation is disappointing
   - Hardly any successful business model
   - “Killer applications”?
3. Is still *not* pervasive
4. Not strong enough to enforce changes in hardware, op. sys, middle-ware, applications
   - Best effort instead of Quality of Service awareness

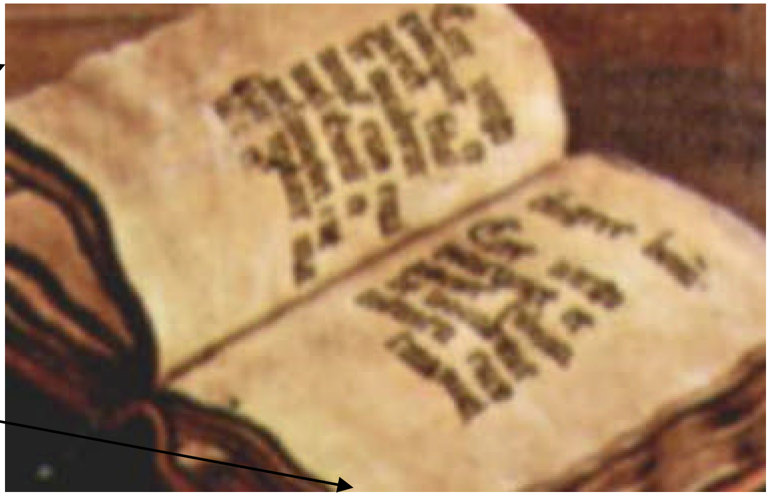
Further Issues in Multimedia

5. Semantic gap (pictures vs. arrays of pixels)
   - Gap is larger than in graphics (a spline is more “intelligent” than an array of pixels)
6. Delivery
   - From any source to any target – unsolved
7. Presentation
   - Sequential tape “paradigm” still leading
8. Software engineering
   - QoS-aware design methods and software tools rare
9. Digital right management



Behold, a virgin shall conceive (Isiah, 7, 14)

- Isenheim Altarpiece
- Grünewald (1512 – 1516)
- Unterlinden Mus., Colmar
- Ecce virgo concipit et pariet



University of Craiova

Romania

Software Engineering Department

Visual Information Retrieval

Dumitru Dan Burdescu

burdescu@software.ucv.ro

Selected concepts and principles from Information Retrieval

Information Retrieval (IR) deals with the representation, storage, organization of, and access to information items.

The key goal of an IR system is to retrieve information which might be **useful or relevant** to the user. The emphasis is on retrieval of *information* as opposed to the retrieval of *data*.

The notion of **relevance** is at the center of IR. The primary goal of an IR system is to retrieve as many the documents which are *relevant* to a user query (high recall) while retrieving as few *non-relevant* documents as possible (high precision).

The effective retrieval of relevant information is directly affected both by the *user task* and by the *logical view of the documents* adopted by the IR system.

[Christopher D. Manning 2008]

Selected concepts and principles from Information Retrieval

There are two different views of IR:

– Computer-centered

- Build efficient indexes
- Process user queries with high performance
- Develop ranking algorithms which improve the quality of the answer set

– Human-centered

- Study user behavior
- Understand user's needs
- Determine how such understanding impacts the design and operation of IR systems

Visual Information retrieval

More ambiguities arise when interpreting images than words, which makes user interaction more of a necessity;

Judging a document takes time, while an image reveals its content almost instantly to a human observer, which makes the feedback process faster and more sensible for the end user.

Motivation

- Taking pictures and storing, sharing, and publishing them has never been so easy and inexpensive like nowadays. If only we could say the same about *finding* the images we want and *retrieving* them.

- We are trying to create automated solutions to the problem of finding and retrieving visual information (images, videos) from (large, distributed, unstructured) repositories in a way that satisfies the search criteria specified by their users, relying (primarily) on the visual contents of the media.

- The fundamental difficulty in doing what we want to do is related to the need to encode, perceive, convey, and measure ***similarity*** (e.g., between two images).

[Oge Marques 2008]

Motivation

- **Challenge**: the elusive notion of similarity!

Let's take a look at a few examples:

- Are these two images similar?

Motivation

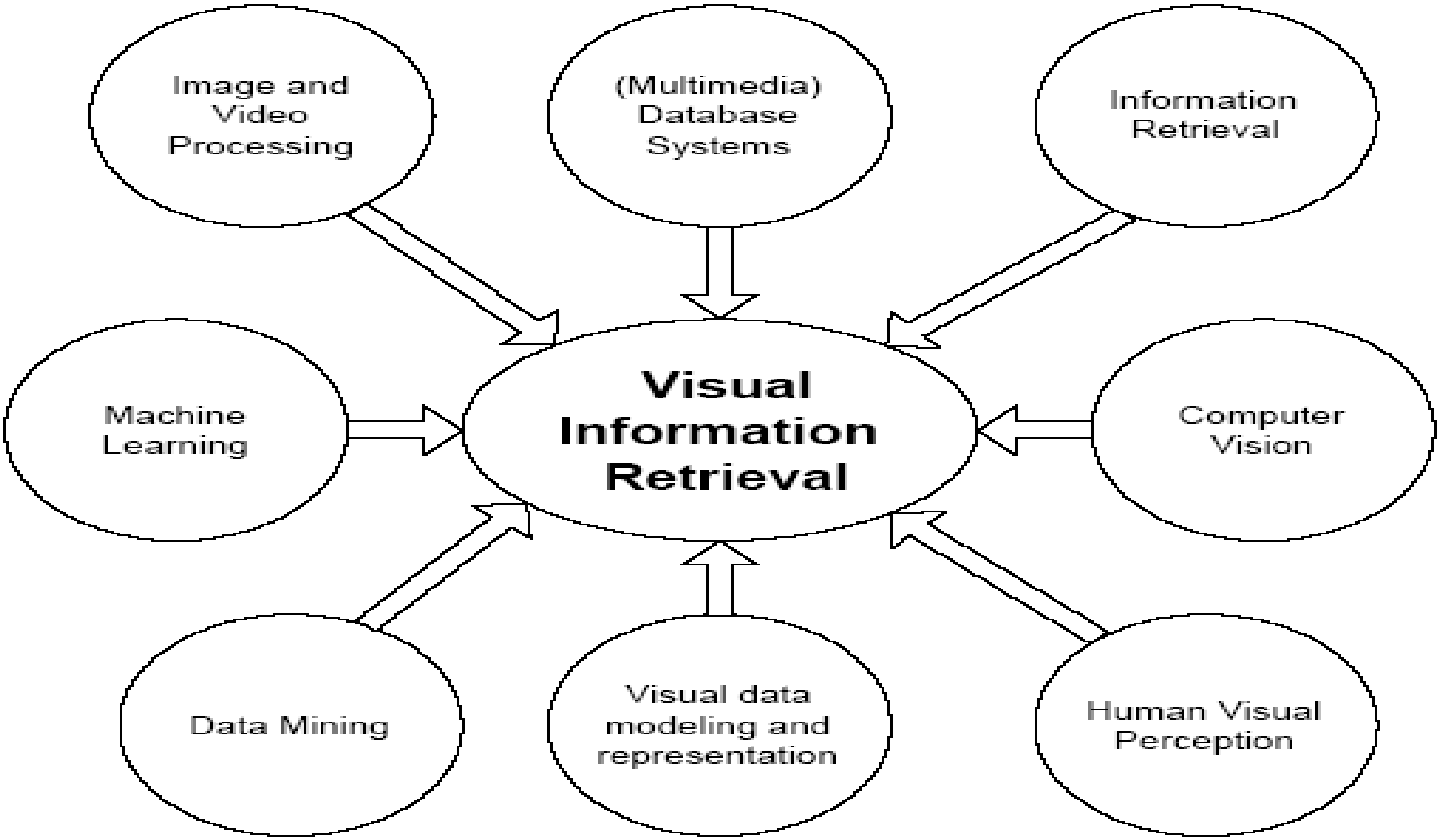
- **Challenge:** each application has different users (with different intent, needs, background, cultural bias, etc.) and different visual assets.
- **Challenge:** the *semantic gap*.
- **Challenge:** the same image evokes more than one (high-level) meaning (*polysemy*)

– According to WordNet, polysemy is “the ambiguity of an individual word or phrase that can be used (in different contexts) to express two or more different meanings”.

Motivation

- Semantic gap: Two images that are usually rated as <u>not</u> similar by Visual Image Retrieval systems. Are they similar?

Domains Related to VIR



Principles of Visual Information Retrieval

The early years (1995-2000):

– Query-by-example
– Emphasis on low-level features (mostly color)
– Sensory and semantic gaps
– Weak segmentation

Principles of Visual Information Retrieval

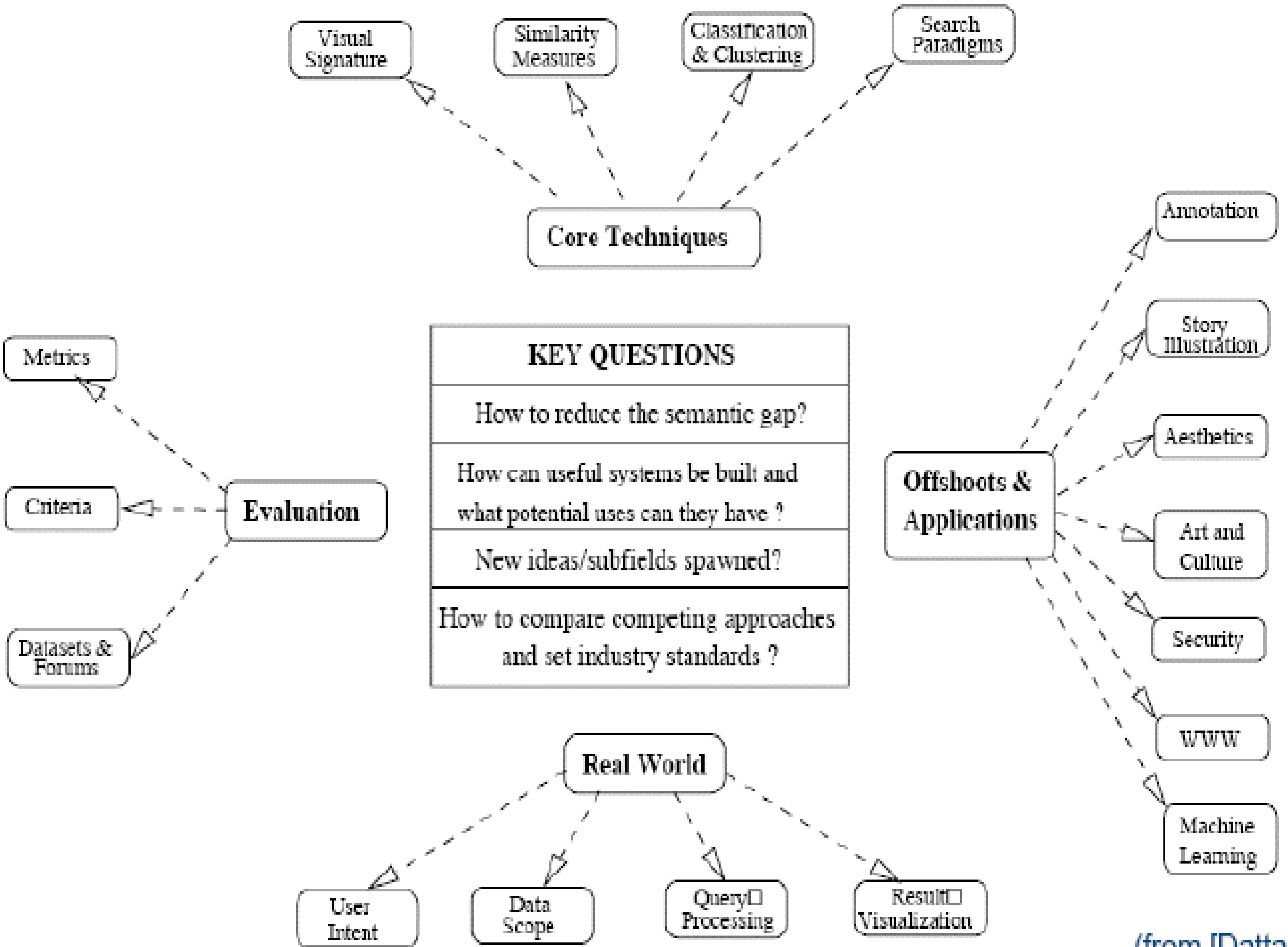
- Sensory gap and semantic gap

  – The *sensory gap* is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.

  – The *semantic gap* is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.

  “The pivotal point in content-based retrieval is that the user seeks semantic similarity, but the database can only provide similarity by data processing. This is what we called the semantic gap.” [Smeulders et al., 2000]

Principles of Visual Information Retrieval



(from [Datta et al, 2008])

Principles of Visual Information Retrieval

I think that successful VIR solutions will:

- combine content-based image retrieval (CBIR) with metadata (high-level semantic based image retrieval)
- only be truly successful in narrow domains
- include the user in the loop
    - Relevance Feedback (RF)
    - Collaborative efforts (tagging, rating, annotating)
- provide friendly, intuitive interfaces

Visual Features

Visual features may be classified as:

– **global** features (entire image) and

– **local** features (specific objects or regions within the image)

The problem of segmentation:

– **Strong segmentation**: a division of the image data into regions in such a way that *region T contains the pixels of the silhouette of object O in the real world and nothing else*.

□ Object segmentation for broad domains of general images is not likely to succeed, with a possible exception for sophisticated techniques in very narrow domains.

– **Weak segmentation**: a grouping of the image data in conspicuous regions T internally homogenous according to some criterion.

□ The criterion is satisfied if region T is within the bounds of object O, but there is no guarantee that the region covers all of the object's area.

□ **When occlusion is present in the image, weak segmentation is the best one can hope for.**

Visual features - Color

Color is one of the most widely used features.

- Advantages:
  – robust to background complications
  – independent of size and rotation
  – meaningful to human beings ("Color is one of the most obvious and pervasive qualities in our environment" - E. Bruce Goldstein in *Sensation & Perception*).
- Even simple color descriptors provide good initial results, but be careful with what you mean by "simple".

Visual features - Color

Color-based features usually rely on a properly chosen *color space*.

– A color model (also called color space or color system) is a specification of a coordinate system and a subspace within that system where each color is represented by a single point (color normalization and representation: RGB; color standardization: (CIE)XYZ; perceptual uniformity: L*u*v* (CIELUV); intuitive description: HSV).

Possible ways of representing / encoding the color contents of an image:

– Color histogram
– Color moments
– Color sets
– Color coherency vectors (CCVs)
– MPEG-7 color descriptors
– Many more…

Visual Features - Color

Global and Accumulating Features

– A simple but very effective approach to accumulating features is to use the **histogram**, that is, the set of features F(*m*) ordered by histogram index *m*. The original idea to use histograms for retrieval comes from

[Swain and Ballard, 1991].

– A histogram may be effective for retrieval as long as there is a uniqueness in the color pattern held against the pattern in the rest of the entire data set.

– Moreover, the histogram shows an obvious robustness to translation of the object and rotation about the viewing axis.

– Histogram's extensions and alternatives include:

- Joint histograms
- Geometric histogram
- Color moments

Visual features - Texture

Fundamentals:

– Even though the concept of texture is intuitive we recognize texture when we see it, but a precise definition of texture has proven difficult to formulate.

– Many different texture definitions are in the literature.

– Despite the lack of a universally accepted definition of texture all researchers agree on two points:

- within a texture there is significant variation in intensity levels between nearby pixels that is at the limit of resolution there is non-homogeneity
- texture is a homogeneous property at some spatial scale larger than the resolution of the image.

It is implicit:

– A single physical scene may contain different textures at varying scales.

Visual features - Texture

- Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of a single color or intensity.

- It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment.

- Research in texture analysis started in the early 70’s. The best-known approaches to describe / analyze textures can be divided into:
    - statistical,
    - structural, and
    - spectral

Visual features - Texture

Some of the most popular texture extraction approaches are:

- Haralick’s gray-level co-occurrence matrices (GLCM)
- Gray Level Difference (GLD) methods
- Fourier power spectrum and autocorrelation methods
- Edge density and edge histograms (MPEG-7)
- Global texture descriptors
- Gabor features
- Wavelet-based features
- Morphological features
- Fractal-based features
- Tamura features (uniformity, density, coarseness, roughness, regularity, linearity, directionality, frequency, and phase)
- Many more…

Visual features - Shape

Retrieving images based on the shape of its most important components seems to improve the capability of a VIR system.

– The problem: shape detection requires segmentation.

– Segmentation is an indispensable stage in most Computer Vision systems.

– Segmentation is not a completely solved problem. Good algorithms use a lot of *a priori* knowledge about the context. Extending this knowledge to a generic image repository is virtually impossible.

– VIR systems that attempt to use some type of segmentation usually rely on human assistance.

 Shape representations may be invariant to translation, rotation, and scaling, or not, depending on the application.

 Shape representations may be divided in two main categories: boundary-based and region-based.

 Restricted to narrow-domain applications and/or applications where segmenting the object of interest is relatively easy.

Visual features - Shape

Popular shape descriptors

– Global image transforms (e.g., wavelets)
– Moments and moment invariants
– Global object features (e.g., area, circularity, eccentricity, compactness, major axis orientation, Euler number, concavity tree, shape numbers, and algebraic moments)
– Fourier descriptors
– Decomposition into eigenvectors (PCA)
– Others

Visual features – Measuring Similarity

The issue of determining and measuring similarity is still an open (and very active) research topic.

Many **challenges**:

– Similarity is an elusive concept

– Mathematical convenience vs. true measure of similarity (as judged by humans)

– How far can we go based on pixel values alone?

– Can vision (and cognitive) science help?

Similarity between Features:

– While searching for a query image among the elements of the data set of images, knowledge of the domain will be expressed by formulating a similarity measure between the images (query, data set) on the basis of some feature set.

– The similarity measure depends on the type of features (Histogram distances, Similarity of Object Silhouettes, Similarity of two feature vectors or as a probabilistic concept, Similarity of Structural Features, Similarity at the Semantic Level)

Visual features – Measuring Similarity

Among the many unexplained aspects of similarity judgment, there is a claim that humans perceive visual similarity in two ways:

- pre-attentive: based on image features only
- attentive: uses features + interpretation based on previous knowledge and a form of reasoning

- Whereas computers (so far) can only afford the pre-attentive way.

Visual Information Retrieval

Thank you for your time!