# Limitations of Content-based Image Retrieval

## Or why it is hard to Google pictures



## Limitations of CBIR - Outline

- What is the true current state of the art?
- Methodological Problems of General CBIR
- Why is CBIR so Hard?
- What can we learn from the past?
- What can be done?

## The True Current State of the Art

- Title of Editorial in Special Issue of IEEE Proceedings (April 2008): ***"The Holy Grail of Multimedia Information Retrieval: So Close or Yet So Far Away?"***
- **Close** if we take published results at face value.
- **Far Away** if we evaluate results from online test sites or look closely at the published results.
- The answer also depends on what do we mean by CBIR.



## Two Kinds of CBIR

- **General:** We try to match a query image to an arbitrary collection of images, such as those found on the web. The goal of the query is to obtain images with the same object as the query. Such CBIR imitates web search engines for images rather than for text.
  - Given an image with a horse, find all images showing a horse (at least as their main subject).
- **Application Specific:** We try to match a query image to a collection of images of a specific type. For example, fingerprints, X-ray images of a specific organ, images of skin lesions, etc.

# Results from Online Tests of General CBIR

- Two types of tests are publicly available: image retrieval and auto-tagging.
- Fewer sites are actually available than advertised.
  - For example, *Cortina* (cited in a Nov. 2008 PAMI paper) is not operational except for already tagged images.
- Results are generally poor. Only one site (GazoPa) produced a good match and that was only once. See [Appendix A] and [Recent Tests]. In some sites the system failed to produce any results.

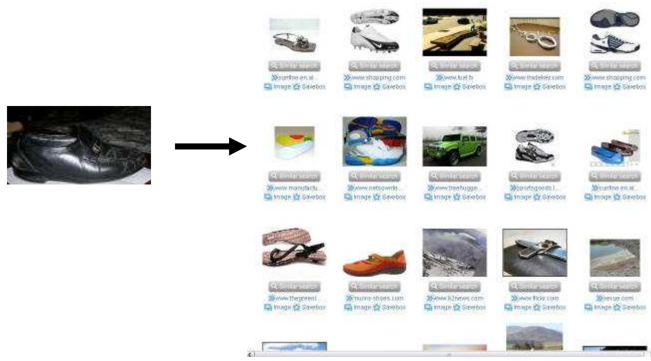


# Shopping for Shoes (or Cars?)

(GazoPa – November 25, 2008)

Capturing the overall shape is not enough!
There is a semantic abyss rather than just semantic gap.

## Shopping for Shoes (cont)
(GazoPa – November 25, 2008)



Adding a tag (SHOE) has not helped.



## An Auto-tagging Result
(From a new site, October 2008)

Rest, chairs, **architecture**, animals, Europe, church, boats, livestock, ports, **city**, Italy, the sea, **building**, boat, beach, **housing**, harbor, holiday

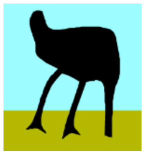
## Another Auto-tagging Result
(From the same site, October 2008)

Mammals, show, Business Woman, animals, black, business, attitude, full, office workers, business, computers, office, smiles, close-up, businessman, adults, parents



## Response of Test Site Owners*

1. "The Problem turned out to be much harder than we thought – we have given up on general CBIR …"
2. "We are still trying to improve our site …"
3. "Student who maintained the site has left", etc.
4. Silence!



---

* Also of Authors

## Limitations of CBIR - Outline

- What is the true current state of the art?
- Methodological Problems of General CBIR
- Why is CBIR so Hard?
- What can we learn from the past?
- What can be done?



Methodological Issues:

## Solutions in Search of a Problem - 1

Many papers describing CBIR methods use trivial queries, for example:

- "Show all pictures with buildings" (rather than a particular building)
- "Show all pictures with people" (rather than a particular person)
- "Show all pictures with a lot of green"

- They do so because these are the only kind of queries that can be answered by the methods used.

## Methodological Issues: Solutions in Search of a Problem - 2

- A classifier is trained on a finite set of classes of objects. The retrieval system is limited to such classes. It cannot deal with an open collection of images.
- Such training will need labeled samples, but if we label all images we do not need CBIR.
- (Using trained classifiers is fine for application specific CBIR where object categories may be well defined.)



## Methodological Issues: Underestimating the Image Space

- Methods are developed over too “small” a set of samples: We deal, in effect, with hashing schemes. No collisions in small sets but many collisions in large sets.
- The possible number of images for a given size is huge. The number of binary images on a 10x10 array is

  $2^{100} > 10^{30}$ > trillions of trillions.

  For a 6x6 array it is

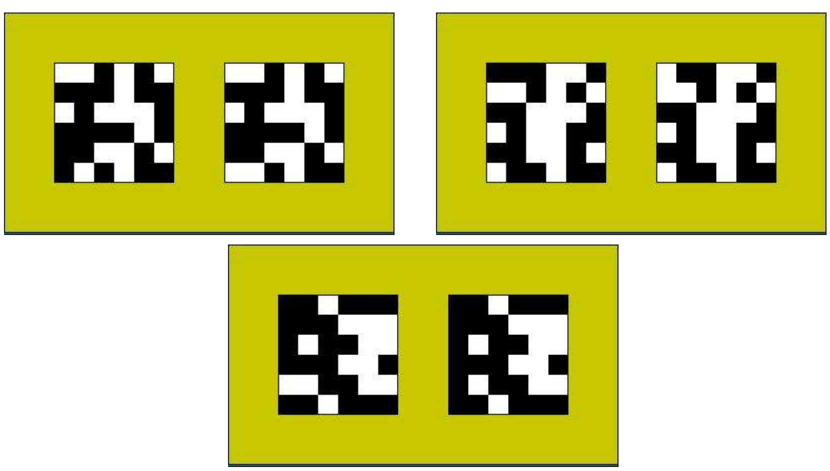
  $2^{36} > 64*10^{9}$ = 64 billions.
- Even “80 million images” (part of the title of a PAMI Nov. 2008 paper) cover too small a part of the space of all images.

## Human Discriminating Ability

- People seem to be able to discriminate 6x6 arrangements, so that 64 billion distinct images seems a realistic lower bound.
- In the following two slides we show three pairs of random patterns that differ in just one location. Because of the lack of order this is the most difficult case for discrimination.
- The second version of the patterns marks the differences.

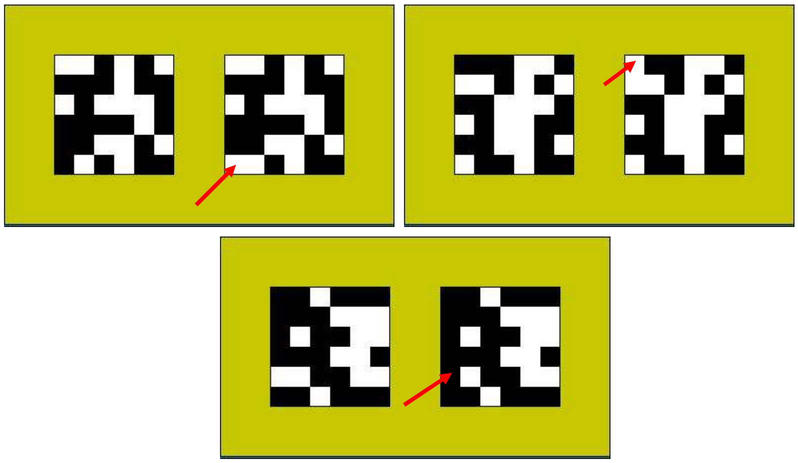


## Some 6x6 Boards



See the next slide for help in finding differences within each pair

# Some 6x6 Boards



Spring 2013 ITS102.23-J 17

## Methodological Issues: Confusion between Looking for Similar Images or Looking for Similar Objects

- Many papers (incl. some in the Nov. 2008 issue of PAMI) are vague on whether they search for similar 2D images or similar 3D objects contained in images.
- Difference in viewpoint/pose and illumination offer serious challenges to methods based on simple features.
- Even segmentation and techniques such as “salient” point matching cannot deal with viewpoint/pose issues, especially for “articulated” objects.

Spring 2013 ITS102.23-J 18

## What is Needed in the Real World

- Infamous example: “Find all pictures of president Clinton and Monica Lewinski.”
- The images in the data cannot possibly be tagged because ML became famous only during the impeachment proceedings.
- The need to perform such open queries requires that we have general image similarity measures that allow for detailed matching. (Possibly comparing parts of images.) **Scene similarity measures** would be even better.
  - Text search works by similarity measures.



## The big challenge: Perceptual versus Computational Similarity

- Two pictures may differ a lot in their pixel values but appear similar to an observer. (“They have the same meaning”.)
- Two pictures may differ in few pixels only but they have different meaning. (Face portraits of two different people in front of the same background.)
- *By the way:* Color is not particularly useful because of the isoluminance effect [Bach_02].

## Two Images with Nearly Identical Color Histograms

More examples in [Appendix C]

Spring 2013 ITS102.23-J 21

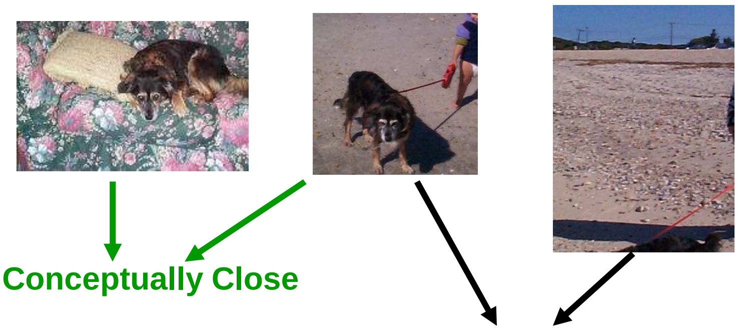
## Perceptual versus Computational Similarity – An Example

**Perceptually close**

**Computationally close (similar pixel statistics)**

Spring 2013 ITS102.23-J 22

# Conceptual Similarity is Even Harder to Deal With



**Conceptually Close**

**Computationally close**
(Large areas with similar local pixel statistics)



# In Summary:

- Pixel values have little to do with the human interpretation of an image.
  - Unfortunately, comparing images pixel-wise persists (papers in Nov. 2008 issue of PAMI)
- There is not just a semantic gap (to be patched with heuristics)
- There is a **SEMANTIC ABYSS** requiring new methodologies.