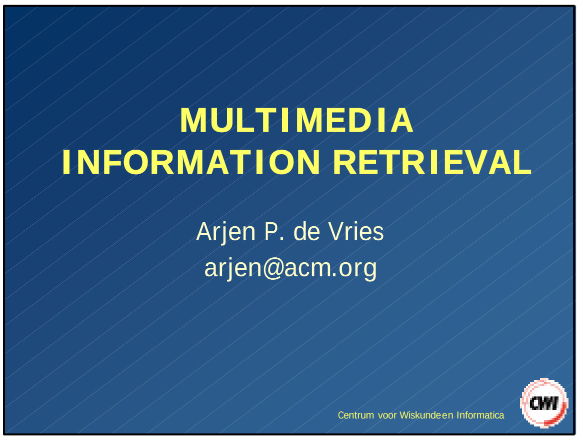
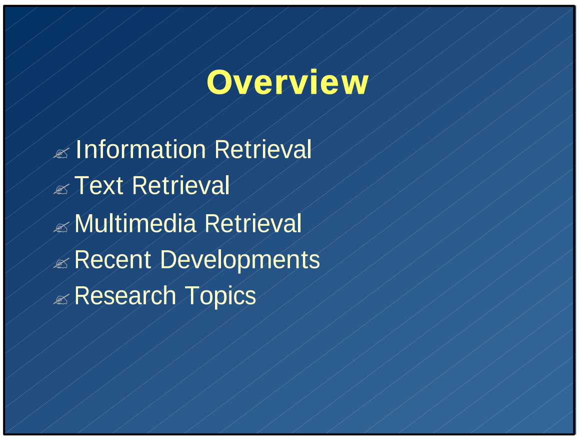


MULTIMEDIA INFORMATION RETRIEVAL

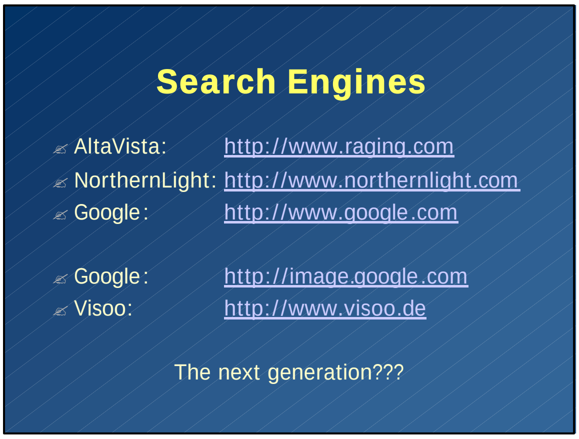
Arjen P. de Vries
arjen@acm.org

Centrum voor Wiskunde en Informatica



Overview

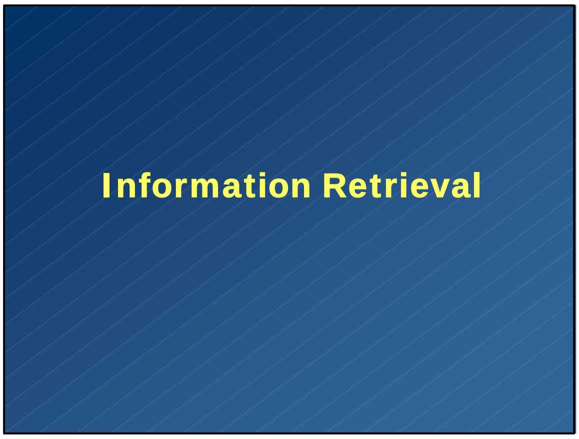
- Information Retrieval
- Text Retrieval
- Multimedia Retrieval
- Recent Developments
- Research Topics



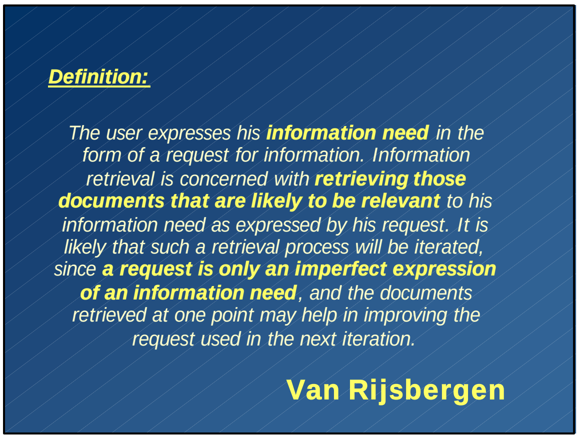
Search Engines

AltaVista:	http://www.raging.com
NorthernLight:	http://www.northernlight.com
Google:	http://www.google.com
Google:	http://image.google.com
Visoo:	http://www.visoo.de

The next generation???



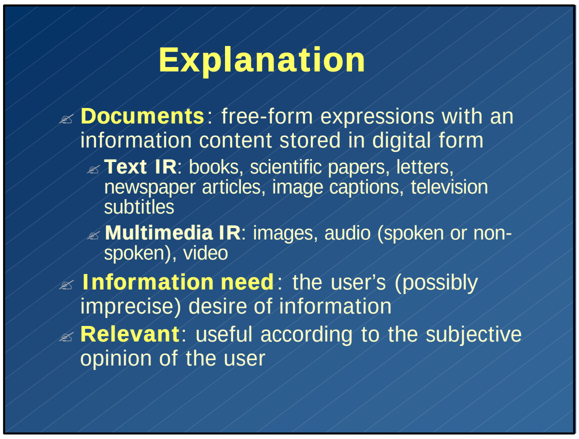
Information Retrieval



Definition:

*The user expresses his **information need** in the form of a request for information. Information retrieval is concerned with **retrieving those documents that are likely to be relevant** to his information need as expressed by his request. It is likely that such a retrieval process will be iterated, since **a request is only an imperfect expression of an information need**, and the documents retrieved at one point may help in improving the request used in the next iteration.*

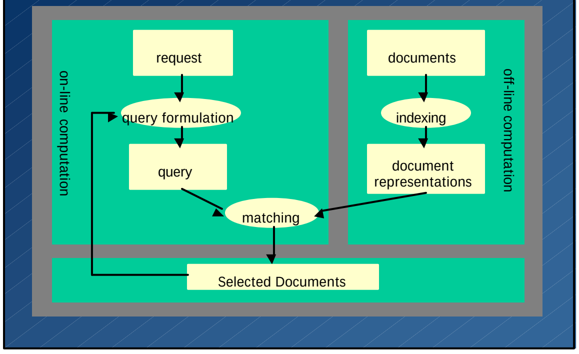
Van Rijsbergen



Explanation

- **Documents**: free-form expressions with an information content stored in digital form
  - **Text IR**: books, scientific papers, letters, newspaper articles, image captions, television subtitles
  - **Multimedia IR**: images, audio (spoken or non-spoken), video
- **Information need**: the user's (possibly imprecise) desire of information
- **Relevant**: useful according to the subjective opinion of the user

Canonical IR System



IR is about satisfying vague information needs provided by users, (imprecisely specified in ambiguous natural language) by satisfying them approximately against information provided by authors (specified in the same ambiguous natural language)

Smeaton

No 'Exact' Science!

- Evaluation is not done analytically, but experimentally
  - real users (specifying requests)
  - test collections (real document collections)
  - benchmarks (TREC: text retrieval conference)

  - Precision
  - Recall
  - ...

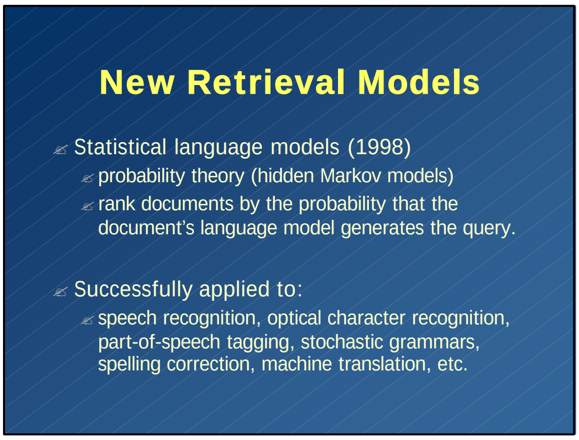
Text Retrieval

Full Text Retrieval

- Index based on uncontrolled (free) terms (as opposed to controlled terms)
- Every word in a document is a potential index term
- Terms may be linked to specific fragments in a text (title, abstract, etc.)

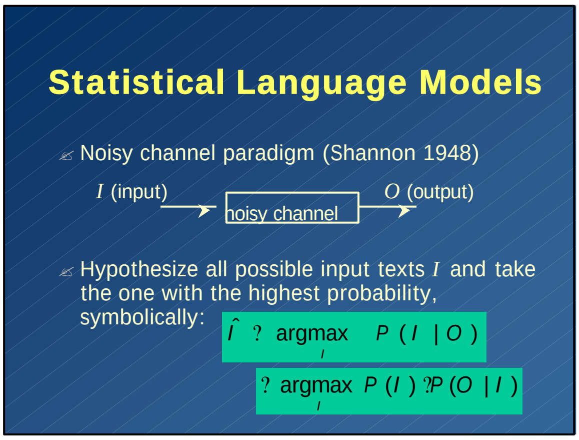
'Old' Retrieval Models

- Boolean model (±1965) 'exact matching'
  - Boolean logic / proposition logic
  - Term specifies a set of documents
- Vector space model (±1970) 'ranking'
  - Geometry
  - Term specifies a dimension in a vector space
- Probabilistic model (1976) 'ranking'
  - Probability theory
  - A term specifies a set of documents
  - Probability of relevance



New Retrieval Models

- Statistical language models (1998)
  - probability theory (hidden Markov models)
  - rank documents by the probability that the document's language model generates the query.
- Successfully applied to:
  - speech recognition, optical character recognition, part-of-speech tagging, stochastic grammars, spelling correction, machine translation, etc.



Statistical Language Models

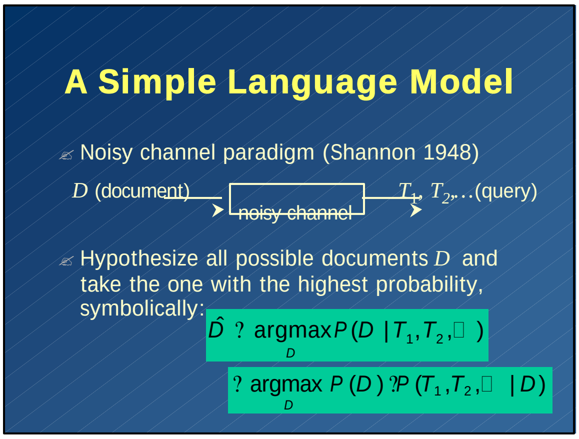
- Noisy channel paradigm (Shannon 1948)

I (input) → noisy channel → O (output)

- Hypothesize all possible input texts I and take the one with the highest probability, symbolically:

$$\hat{I} \ ? \ \underset{I}{\text{argmax}} \ P\ (I \ | \ O\ )$$

$$? \ \underset{I}{\text{argmax}} \ P\ (I\ ) \ ?P\ (O \ | \ I\ )$$



A Simple Language Model

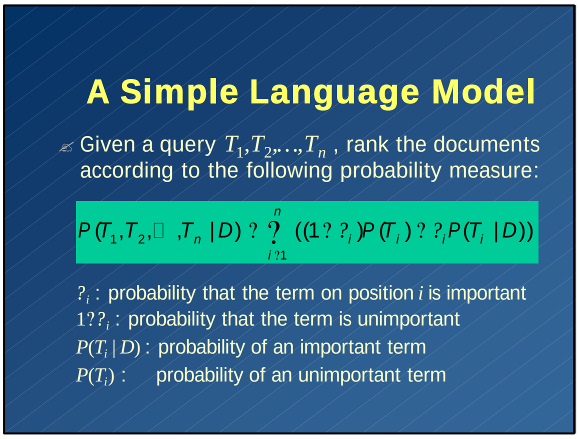
- Noisy channel paradigm (Shannon 1948)

D (document) → noisy channel → $T_1, T_2,\ldots$(query)

- Hypothesize all possible documents D and take the one with the highest probability, symbolically:

$$\hat{D} \ ? \ \underset{D}{\text{argmax}} P\,(D \ | T_1, T_2, \square \ )$$

$$? \ \underset{D}{\text{argmax}} \ P\ (D\ ) \ ?P\ (T_1, T_2, \square \ \ | D)$$



A Simple Language Model

- Given a query $T_1,T_2,\ldots,T_n$, rank the documents according to the following probability measure:

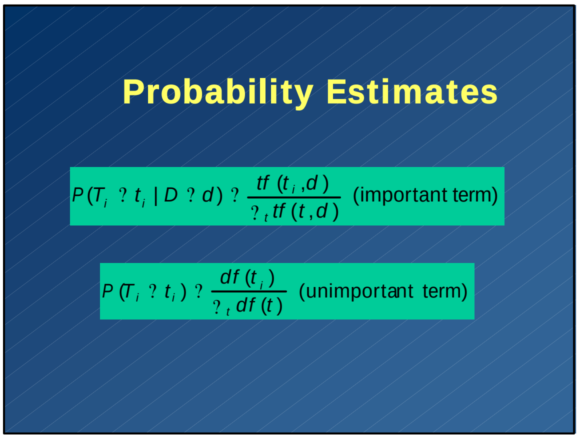
$$P\,(T_1, T_2, \square \ \ , T_n \ | D) \ ? \ \overset{n}{\underset{i\,?1}{?}} \ ((1\,?\ ?_i\,)P\,(T_i\,) \ ? \ ?_i P\,(T_i \ | D))$$

$?_i$: probability that the term on position i is important

$1??_i$: probability that the term is unimportant

$P(T_i\,|\,D)$: probability of an important term

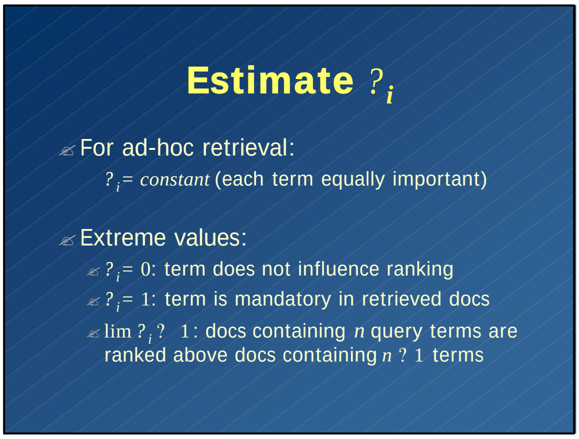
$P(T_i)$: probability of an unimportant term



Probability Estimates

$$P\,(T_i \ \ ? \ t_i \ | \ D \ \ ? \ d\,) \ ? \ \frac{tf\ (t_i\,, d\ )}{?_t\, tf\ (t\,, d\ )} \ \ \text{(important term)}$$

$$P\ (T_i \ \ ? \ t_i\ ) \ ? \ \frac{df\ (t_i\ )}{?_t\ df\ (t\,)} \ \ \text{(unimportant term)}$$



Estimate $?_i$

- For ad-hoc retrieval:

  $?_i$= *constant* (each term equally important)

- Extreme values:
  - $?_i$= 0: term does not influence ranking
  - $?_i$= 1: term is mandatory in retrieved docs
  - $\lim ?_i ?$ 1: docs containing n query terms are ranked above docs containing n ? 1 terms

Relevance Feedback

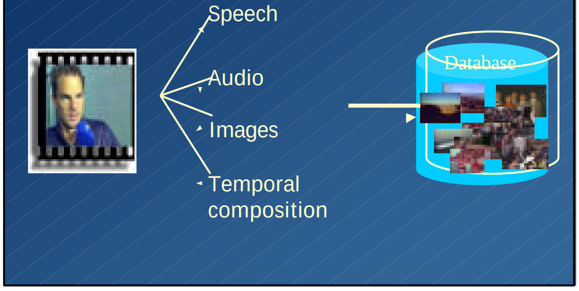
- Re-estimate the value of $?_i$ from relevant documents
  - Expectation Maximisation algorithm
  - Different value of $?_i$ for each term (i.e. different importance of each term.)

Multimedia Retrieval

Indexing Multimedia

- Manually added descriptions
  - 'Metadata'
- Analysis of associated data
  - Speech, captions, …
- Content-based retrieval
  - Approximate retrieval
  - Domain-specific techniques

A Wealth of Information



Speech

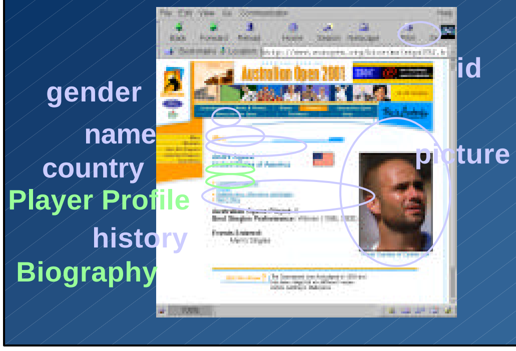
Audio

Images

Temporal composition

Database

Associated Information



gender

name

country

Player Profile

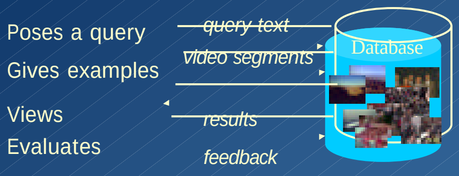
history

Biography

id

picture

User Interaction



Poses a query

Gives examples

Views

Evaluates

query text

video segments

results

feedback

Database

Limitations of Metadata

- Vocabulary problem
  - Dark vs. somber
- Different people describe different aspects
  - Dark vs. evening

Limitations of Metadata

- Encoding Specificity Problem
  - A single person describes different aspects in different situations
- Many aspects of multimedia simply cannot be expressed unambiguously
  - Processes in left (analytic, verbal) vs. right brain (aesthetics, synthetic, nonverbal)

Approximate Retrieval

- Based on **similarity**
  - Find all objects that are similar to this one
  - Distance function
  - Representations capture some (syntactic) meaning of the object
- 'Query by Example' paradigm

Collaborative Filtering

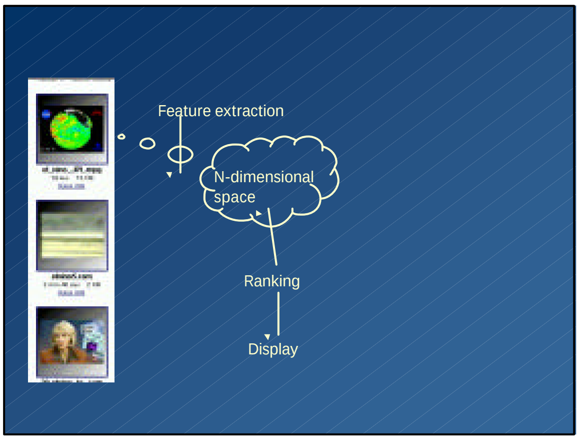
- Also: **social information filtering**
  - Compare user judgments
  - Recommend differences between similar users
- People's tastes are not randomly distributed
- You are what you buy (Amazon)

Collaborative Filtering

- Benefits over content-based approach
  - Overcomes problems with finding suitable features to represent e.g. art, music
  - Serendipity
  - Implicit mechanism for qualitative aspects like style
- Problems: large groups, broad domains

Content-based Retrieval

Query image



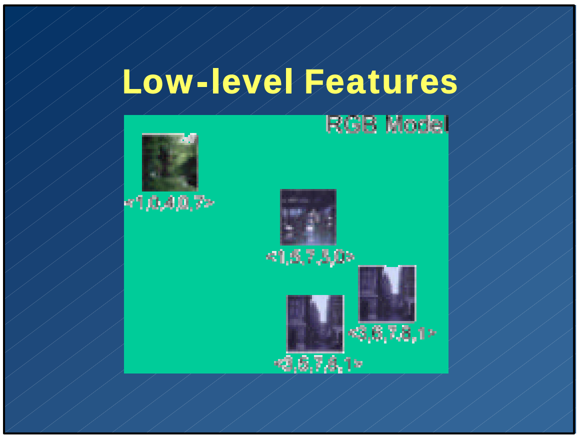
Feature extraction

N-dimensional space

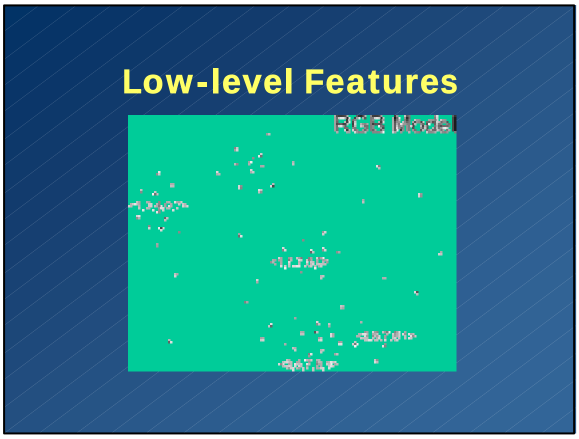
Ranking

Display

Low-level Features



Low-level Features



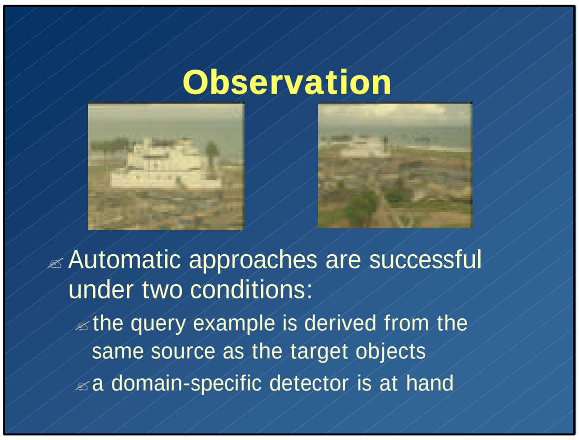
Complicating Factors

- What are Good Feature Models?
- What are Good Ranking Functions?
- Queries are Subjective!

So… is this ever gonna work?!

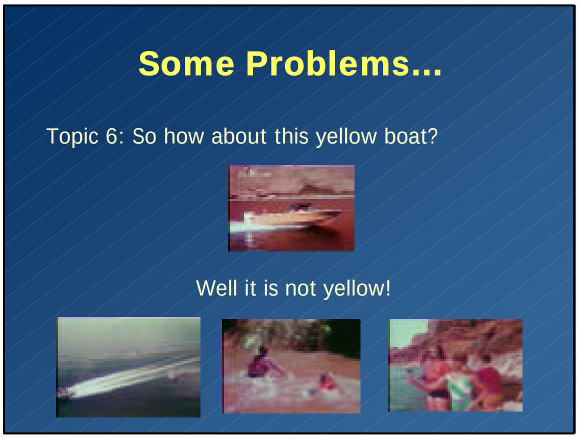
Application to Video

…



Observation

- Automatic approaches are successful under two conditions:
  - the query example is derived from the same source as the target objects
  - a domain-specific detector is at hand



Some Problems…

Topic 6: So how about this yellow boat?

Well it is not yellow!

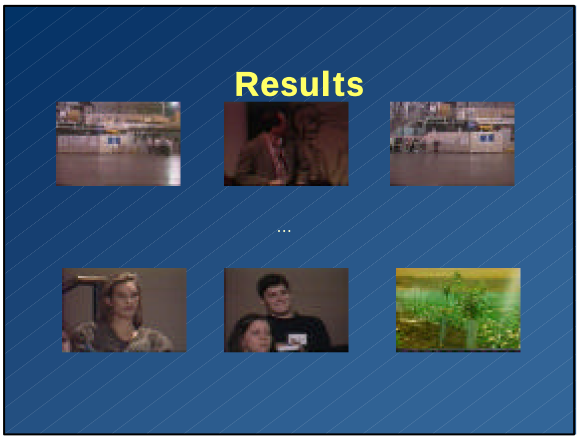
Known Item

Query

Results

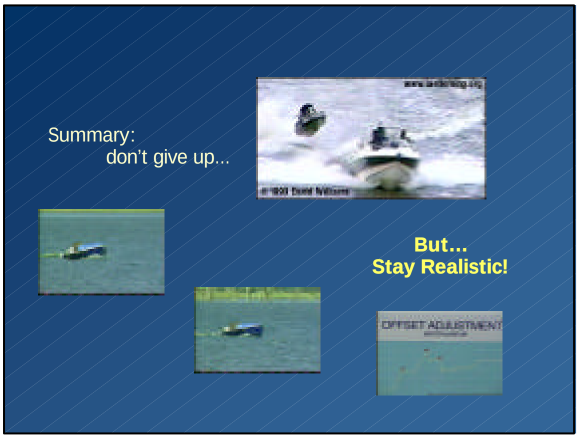
…

Query



Results

...

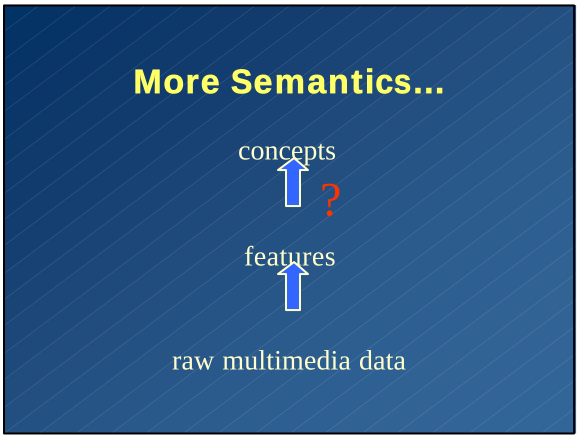


Summary:
don't give up...

But...
Stay Realistic!

OFFSET ADJUSTMENT

Recent Developments



More Semantics...

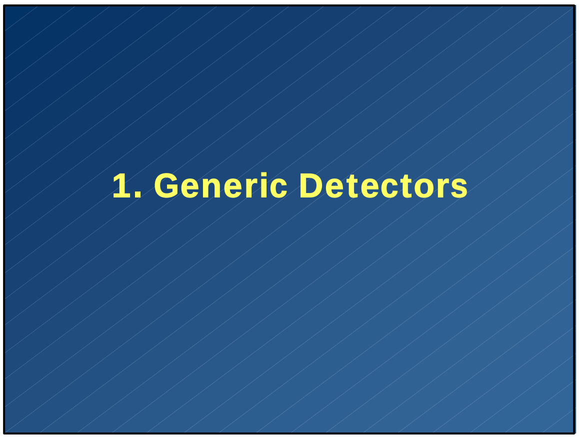
concepts

↑ ?

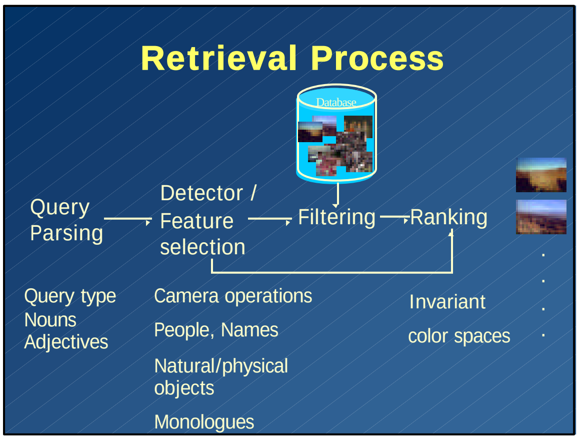
features

↑

raw multimedia data



1. Generic Detectors



Retrieval Process

Query Parsing → Detector / Feature selection → Filtering → Ranking

Query type
Nouns
Adjectives

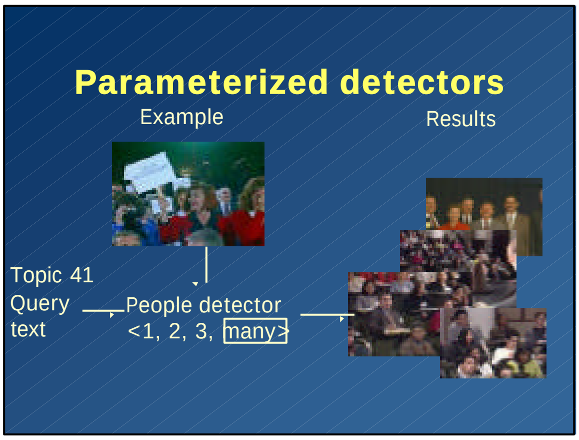
Camera operations

People, Names

Natural/physical objects

Monologues

Invariant color spaces

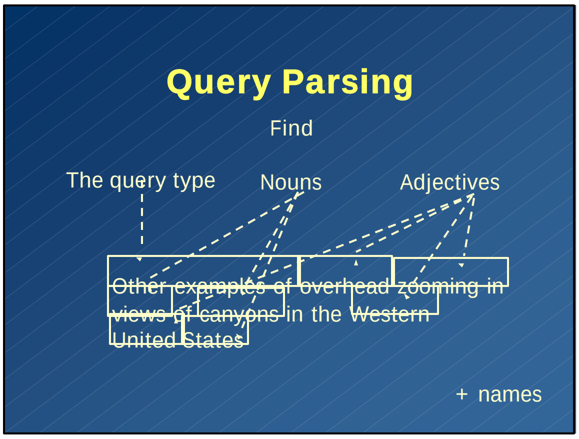


Parameterized detectors

Example

Results

Topic 41
Query text → People detector
<1, 2, 3, many>



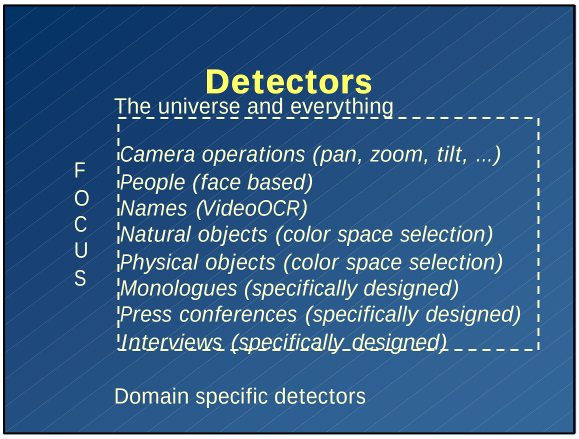
Query Parsing

Find

The query type Nouns Adjectives

Other examples of overhead zooming in views of canyons in the Western United States

+ names



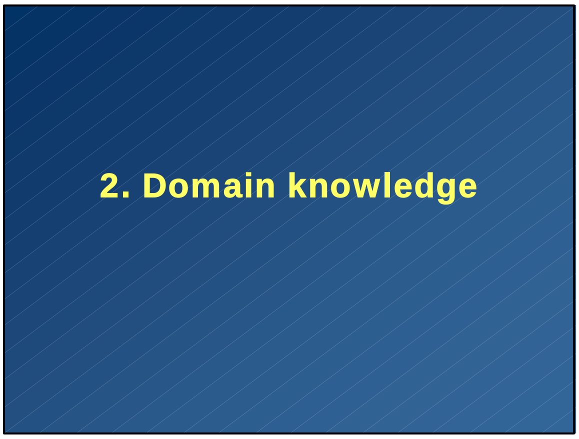
Detectors

The universe and everything

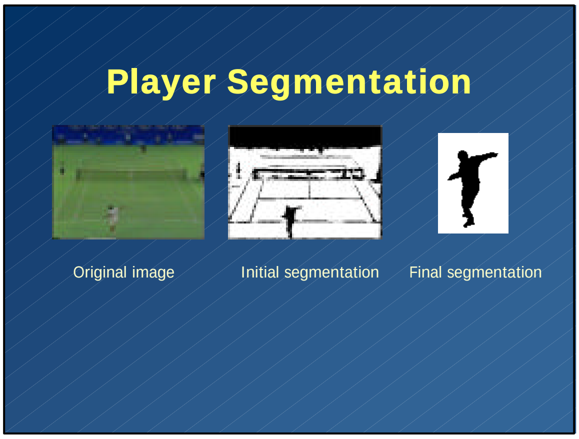
F O C U S

Camera operations (pan, zoom, tilt, ...)
People (face based)
Names (VideoOCR)
Natural objects (color space selection)
Physical objects (color space selection)
Monologues (specifically designed)
Press conferences (specifically designed)
Interviews (specifically designed)

Domain specific detectors



2. Domain knowledge



Player Segmentation

Original image

Initial segmentation

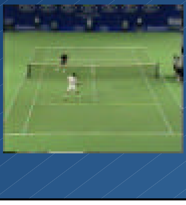
Final segmentation

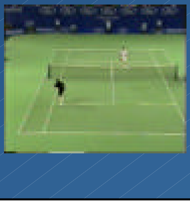
Advanced Queries

Show clips from tennis matches,

starring Sampras,

playing close to the net;

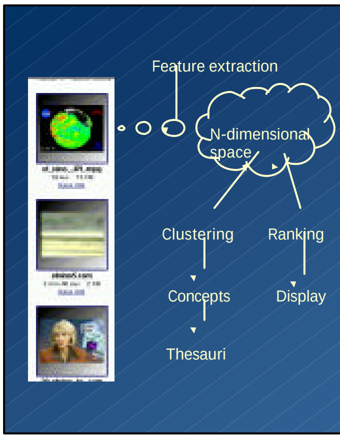




3. Get to know your users

Mirror Approach

- Gather User's Knowledge
  - Introduce semi-automatic processes for selection and combination of feature models
- Local Information
  - Relevance feedback from *a* user
- Global Information
  - Thesauri constructed from *all* users



Feature extraction

N-dimensional space

Clustering

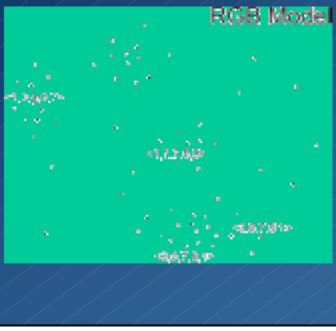
Ranking

Concepts

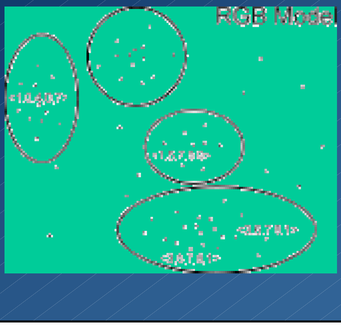
Display

Thesauri

Low-level Features



Identify Groups



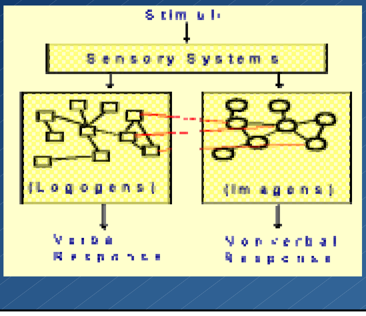
Representation

- Groups of feature vectors are *conceptually* equivalent to words in text retrieval
- So, techniques from text retrieval can now be applied to multimedia data as if these were text!

‘Explaining’ the Results

- Paivio’s dual coding theory conjectures that the human brain processes textual terms (logogens) as well as image terms (imagens)
- Also matches similar music: grunge, house, ...
- Even works for predicting avalanches!

Paivio’s Dual Coding Theory



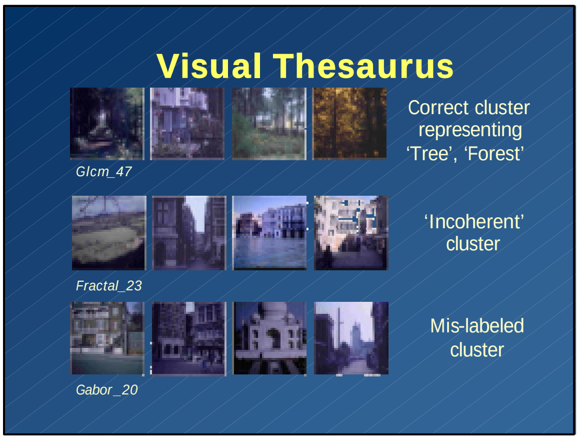
Query Formulation

- Clusters are *internal* representations, not suited for user interaction
- Use automatic query formulation based on *global* information (thesaurus) and *local* information (user feedback)

Interactive Query Process

- Select relevant clusters from thesaurus
- Search collection
- Improve results by adapting query
  - Remove clusters occuring in irrelevant images
  - Add clusters occuring in relevant images

Assign Semantics



Visual Thesaurus

Glcm_47

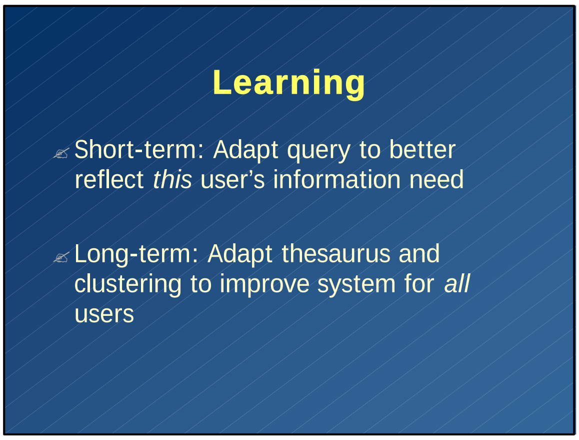
Correct cluster representing 'Tree', 'Forest'

Fractal_23

'Incoherent' cluster

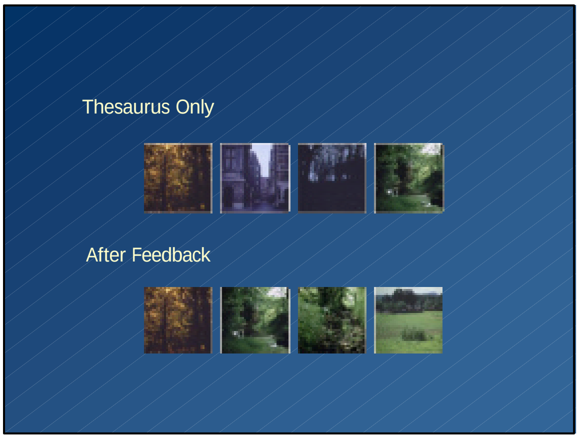
Gabor_20

Mis-labeled cluster



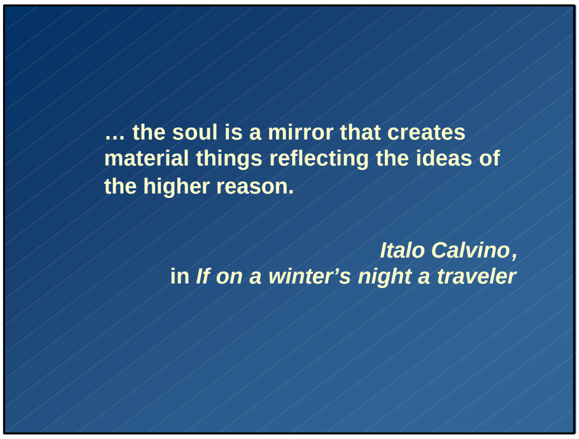
Learning

- Short-term: Adapt query to better reflect *this* user's information need
- Long-term: Adapt thesaurus and clustering to improve system for *all* users



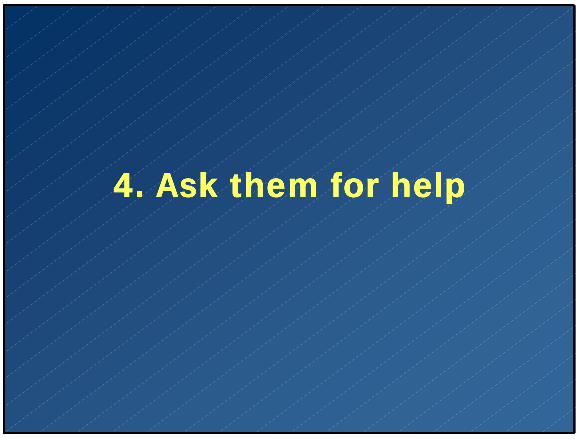
Thesaurus Only

After Feedback

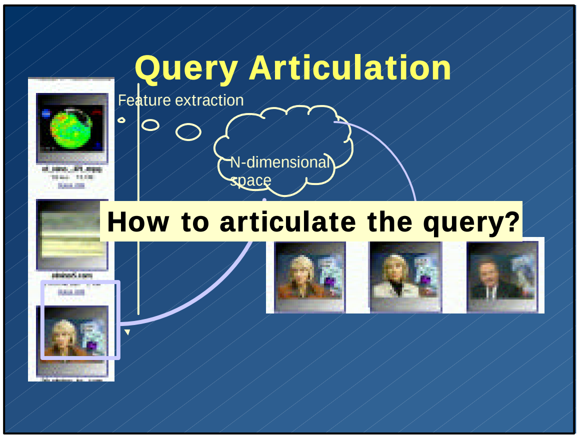


… the soul is a mirror that creates material things reflecting the ideas of the higher reason.

***Italo Calvino*,**
in *If on a winter's night a traveler*



4. Ask them for help

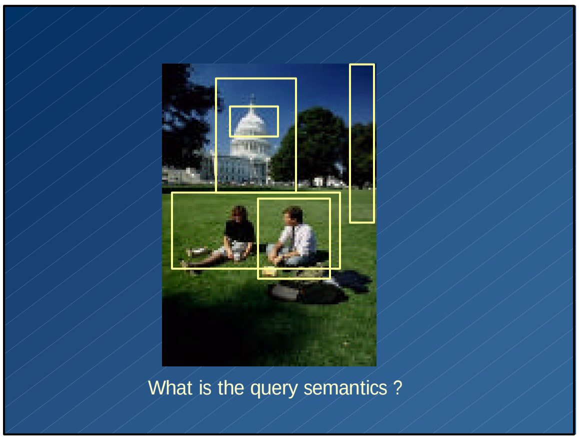


Query Articulation

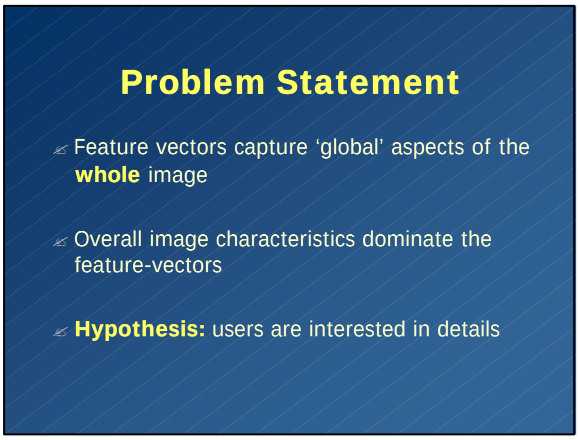
Feature extraction

N-dimensional space

How to articulate the query?

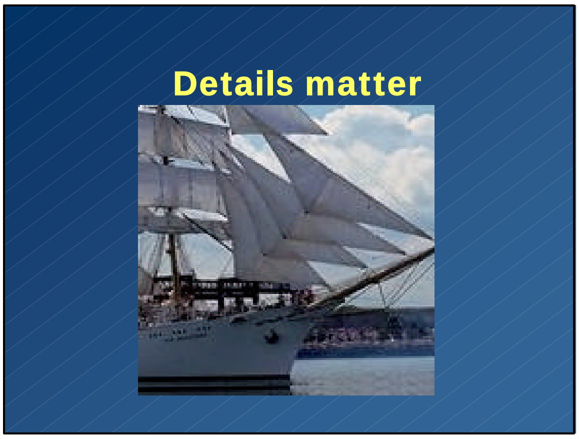


What is the query semantics ?

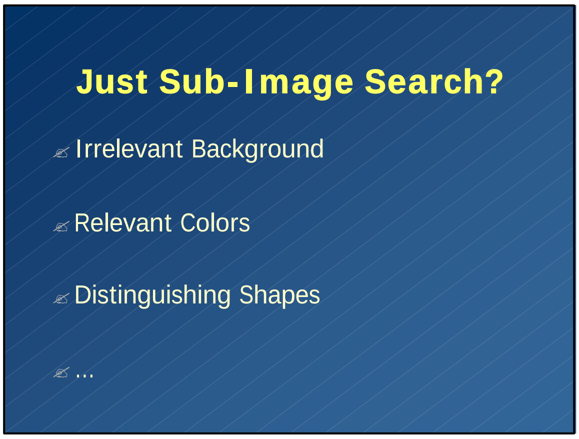


Problem Statement

- Feature vectors capture ‘global’ aspects of the **whole** image
- Overall image characteristics dominate the feature-vectors
- **Hypothesis:** users are interested in details

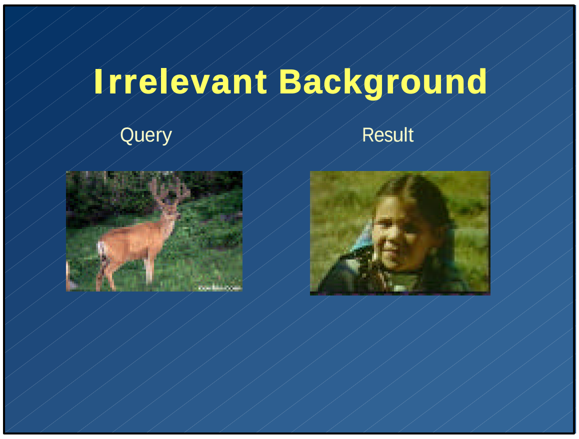


Details matter



Just Sub-Image Search?

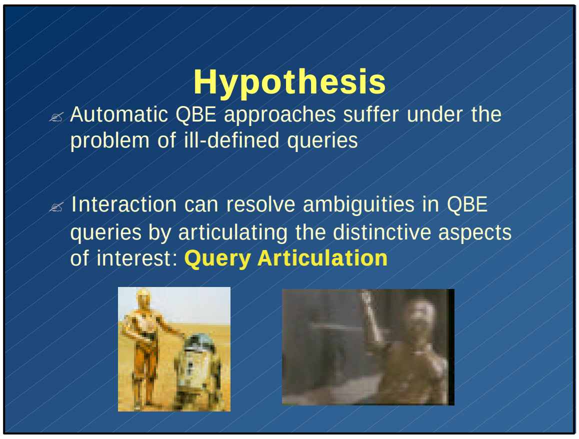
- Irrelevant Background
- Relevant Colors
- Distinguishing Shapes
- ...



Irrelevant Background

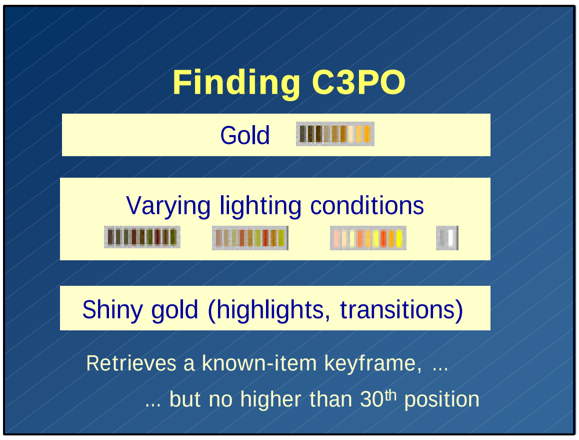
Query

Result



Hypothesis

- Automatic QBE approaches suffer under the problem of ill-defined queries
- Interaction can resolve ambiguities in QBE queries by articulating the distinctive aspects of interest: **Query Articulation**



Finding C3PO

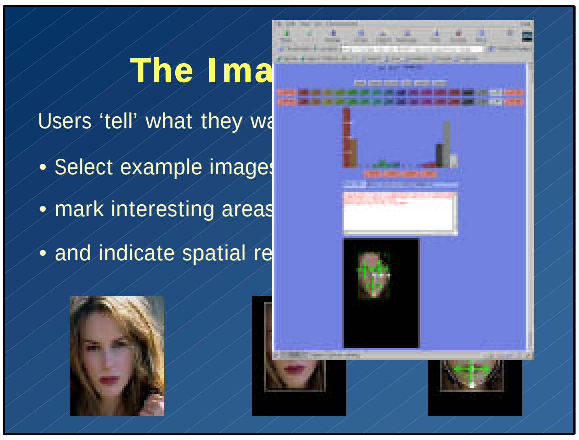
Gold

Varying lighting conditions

Shiny gold (highlights, transitions)

Retrieves a known-item keyframe, …

… but no higher than 30th position



The Ima

Users ‘tell’ what they wa

- Select example images
- mark interesting areas
- and indicate spatial re

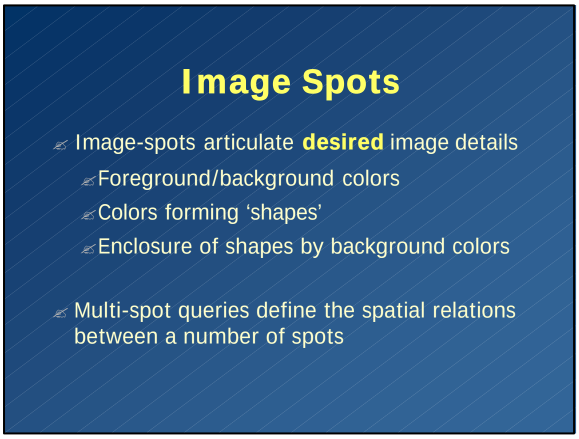
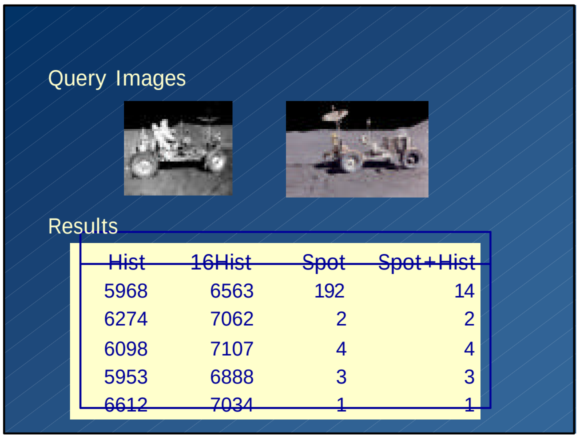


Image Spots

- Image-spots articulate **desired** image details
  - Foreground/background colors
  - Colors forming ‘shapes’
  - Enclosure of shapes by background colors
- Multi-spot queries define the spatial relations between a number of spots



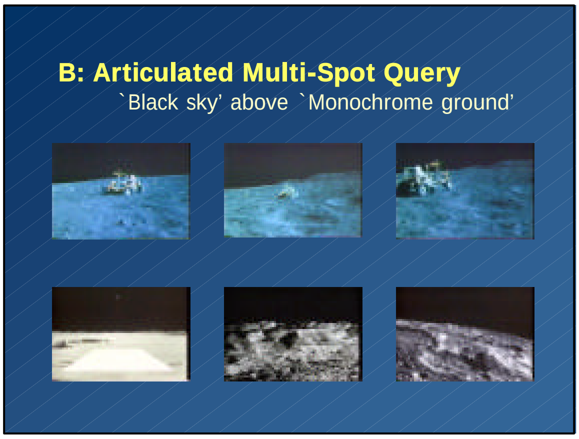
Query Images

Results

Hist	16Hist	Spot	Spot+Hist
5968	6563	192	14
6274	7062	2	2
6098	7107	4	4
5953	6888	3	3
6612	7034	1	1

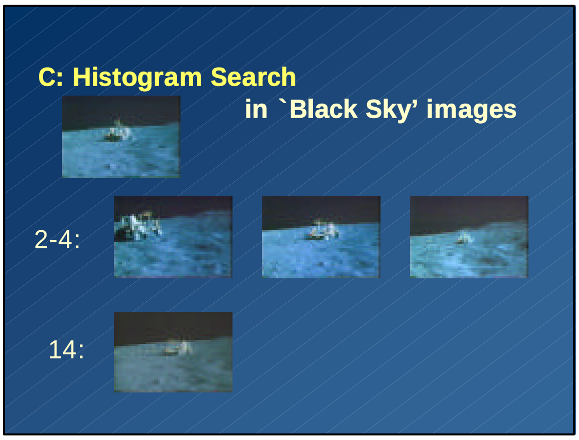
A: Simple Spot Query

\`Black sky’



B: Articulated Multi-Spot Query

\`Black sky’ above \`Monochrome ground’

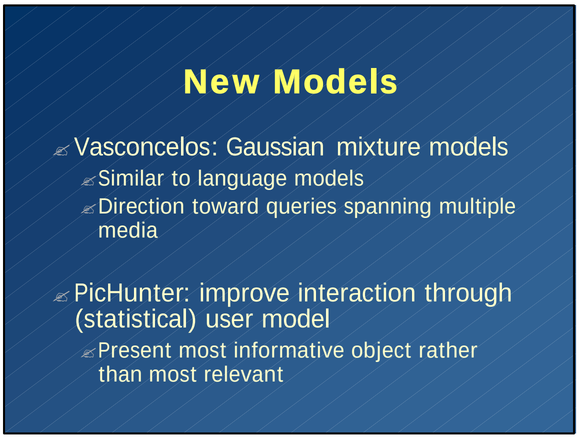


C: Histogram Search in `Black Sky’ images

2-4:

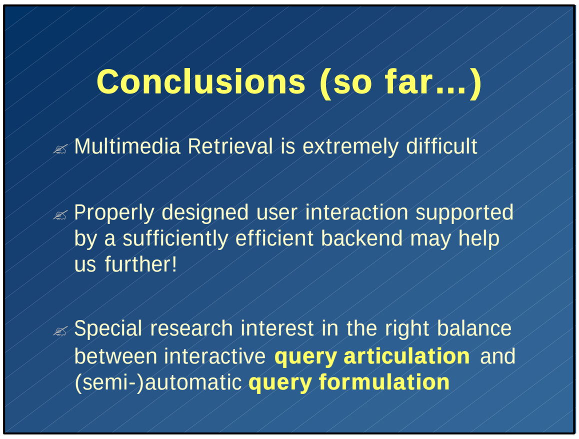
14:

5. Develop Better Models



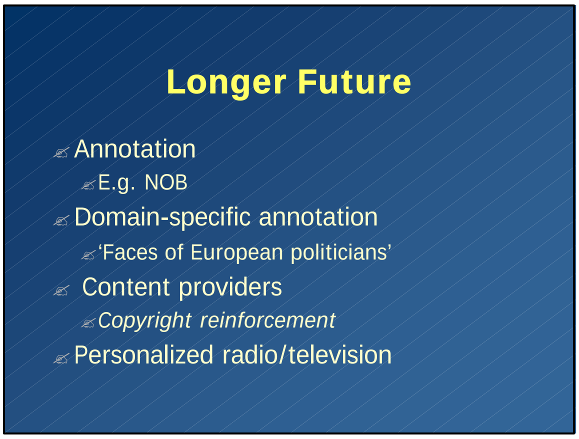
New Models

- Vasconcelos: Gaussian mixture models
  - Similar to language models
  - Direction toward queries spanning multiple media
- PicHunter: improve interaction through (statistical) user model
  - Present most informative object rather than most relevant



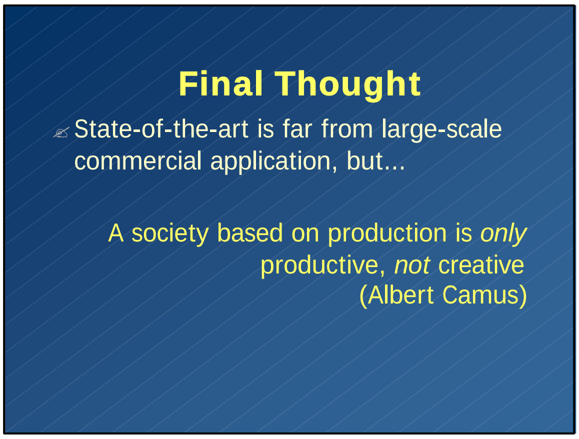
Conclusions (so far…)

- Multimedia Retrieval is extremely difficult
- Properly designed user interaction supported by a sufficiently efficient backend may help us further!
- Special research interest in the right balance between interactive **query articulation** and (semi-)automatic **query formulation**



Longer Future

- Annotation
  - E.g. NOB
- Domain-specific annotation
  - ‘Faces of European politicians’
- Content providers
  - *Copyright reinforcement*
- Personalized radio/television



Final Thought

- State-of-the-art is far from large-scale commercial application, but…

A society based on production is *only*
productive, *not* creative
(Albert Camus)