Carnegie Mellon

Multimedia Analysis:

Marriage of Signal Processing and Machine Learning

Tsuhan Chen 陈祖翰

Professor, Carnegie Mellon University

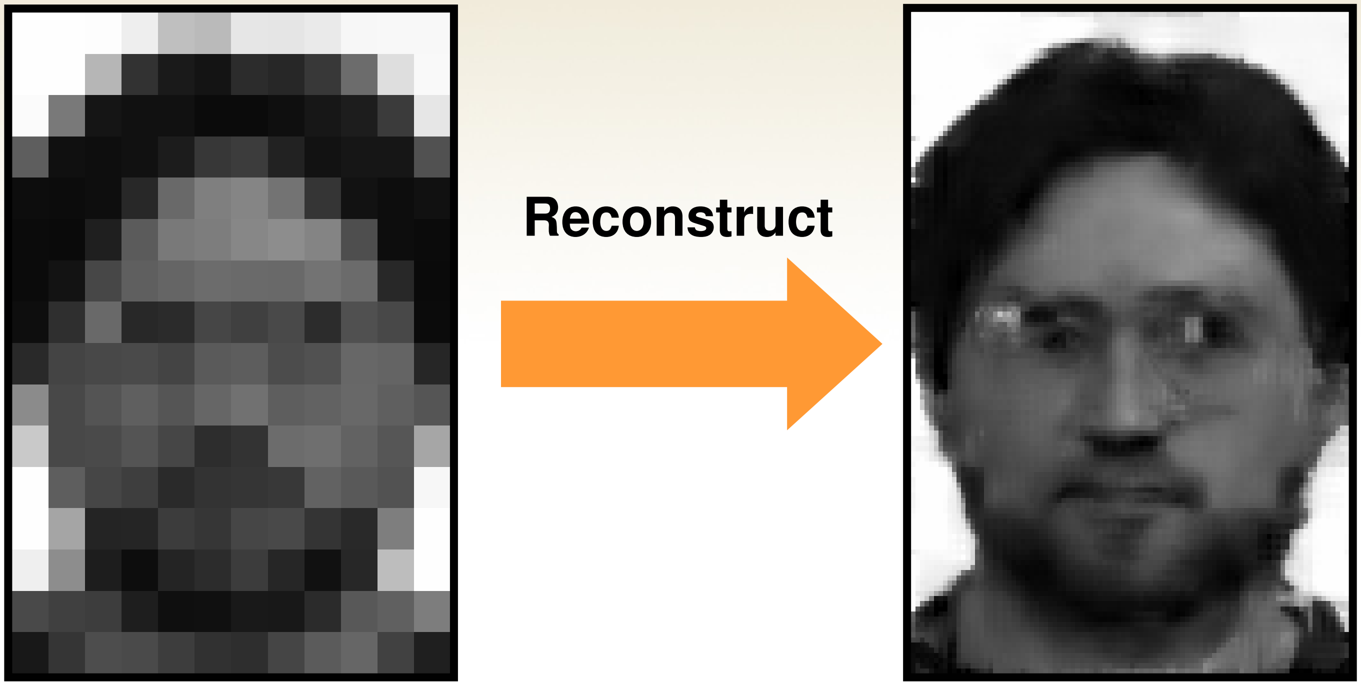
tsuhan@cmu.edu

Advanced Multimedia Processing Lab

A 10-Year Journey…

- IEEE Multimedia Signal Processing (MMSP) Technical Committee, 1996~
- MMSP Workshops
  - Princeton 1997, Los Angeles 1998, Copenhagen 1999, Cannes 2001, St. Thomas 2002, Siena 2004, Shanghai 2005, Victoria 2006
- International Conf on Multimedia (ICME)
  - New York 2000, Tokyo 2001, Lausanne 2002, Baltimore 2003, Taipei 2004, Amsterdam 2005, Toronto 2006, Beijing 2007
- *IEEE Transactions on Multimedia*, March 1999~
  - Special issues: networked multimedia 2001, multimedia database 2002, multimodal interface 2003, streaming media 2004, MPEG-21 2005
- SPS Distinguished Lecturer 2007

Multimedia Analysis: Bits vs. Content



[Baker/Kanade]

It's Bayesian in machine learning, a.k.a., *prior*…

$$\text{Number of all possible } 16\times12 \text{ images } = 2^{16\times12\times8}$$

$$>> 30\times60\times60\times24\times365\times\text{human history}\times\text{world population}$$

$$>> \text{number of all possible face images}$$

Thoughts

- "The most compelling shapes are those near to our hearts: people's faces, a gracefully moving body, a natural scene with rustling leaves and flowing water. Evolution has tuned us to these sights…"

  [Lengyel, 1998]

- Multimedia analysis… more than processing bits…

  it's all about the *content …*

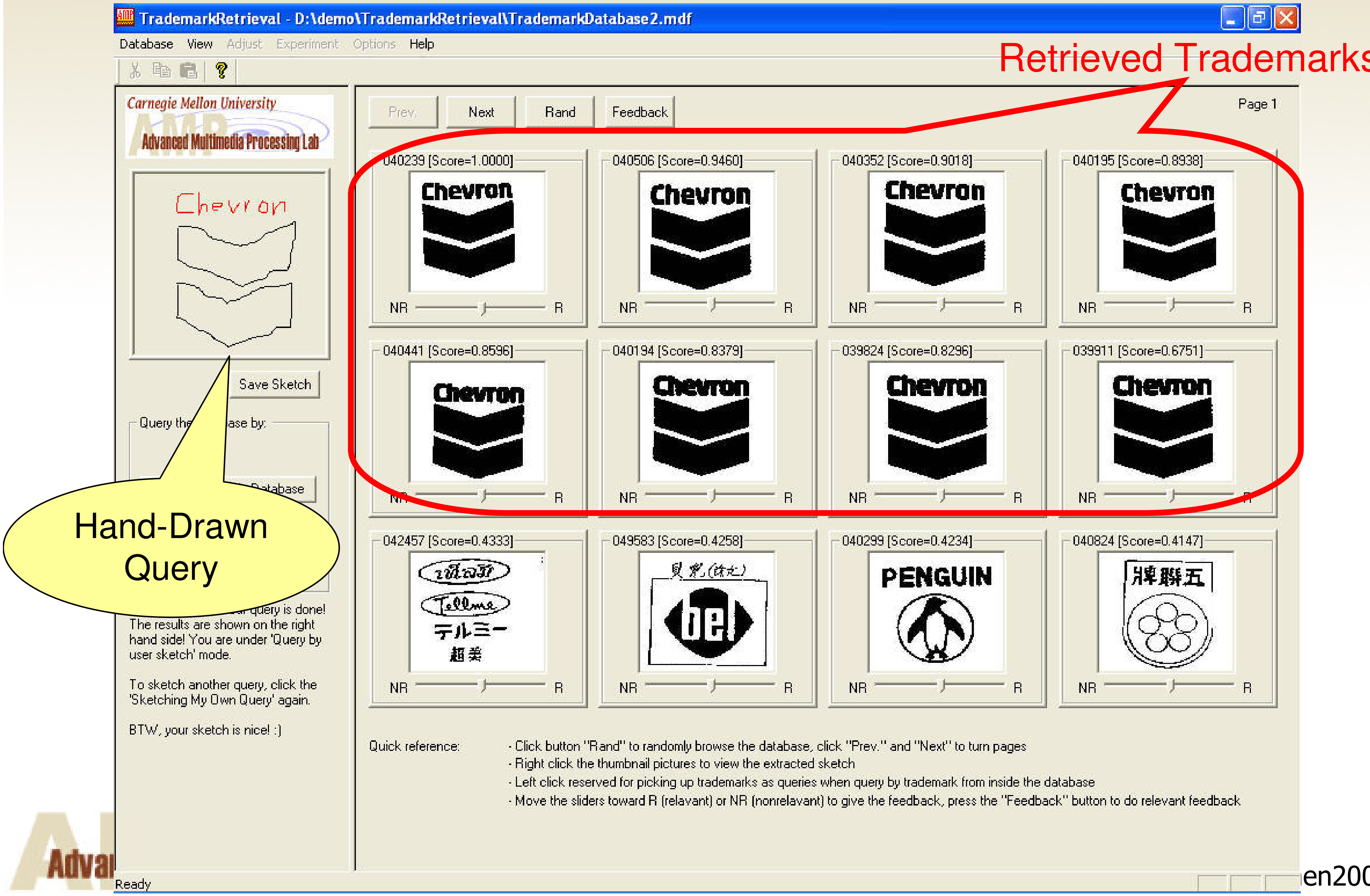
  it's signal processing + machine learning

  [Chen, 2006]

Content-Based Information Retrieval

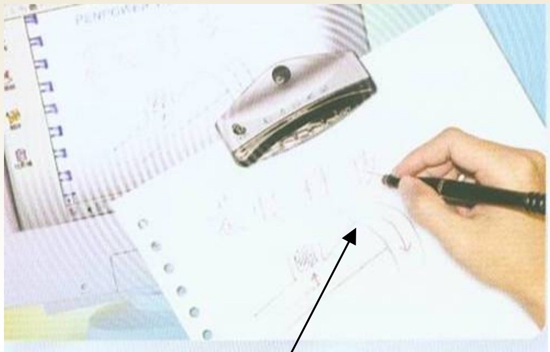
Many Interesting Applications…

Logos [Leung&Chen ICME'02]



Hand-Drawn Sketches

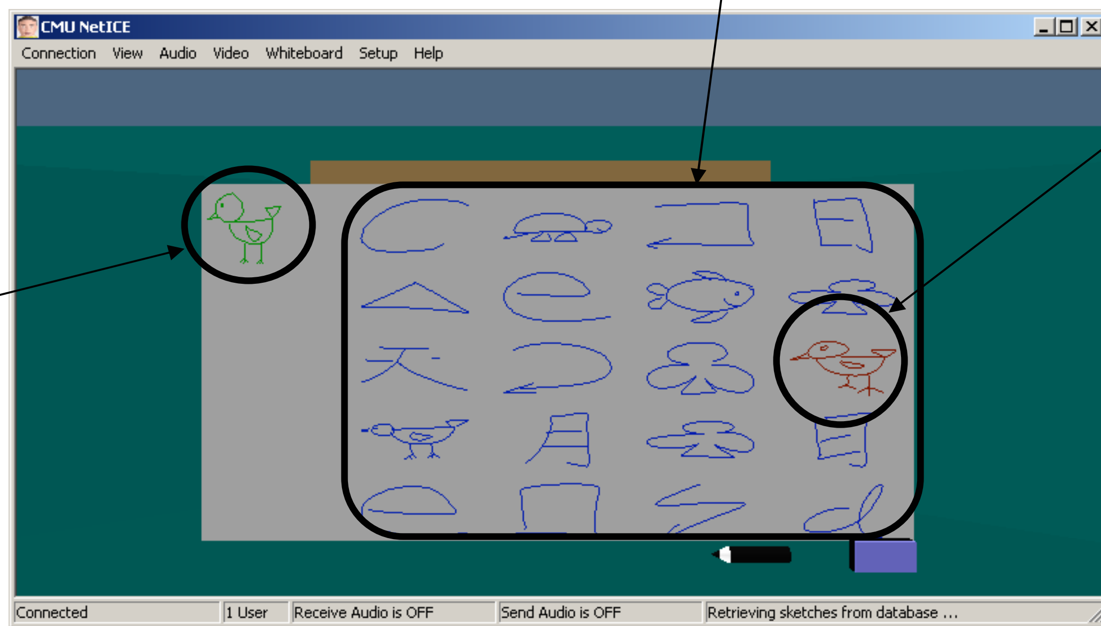
[Leung&Chen ICME'03]



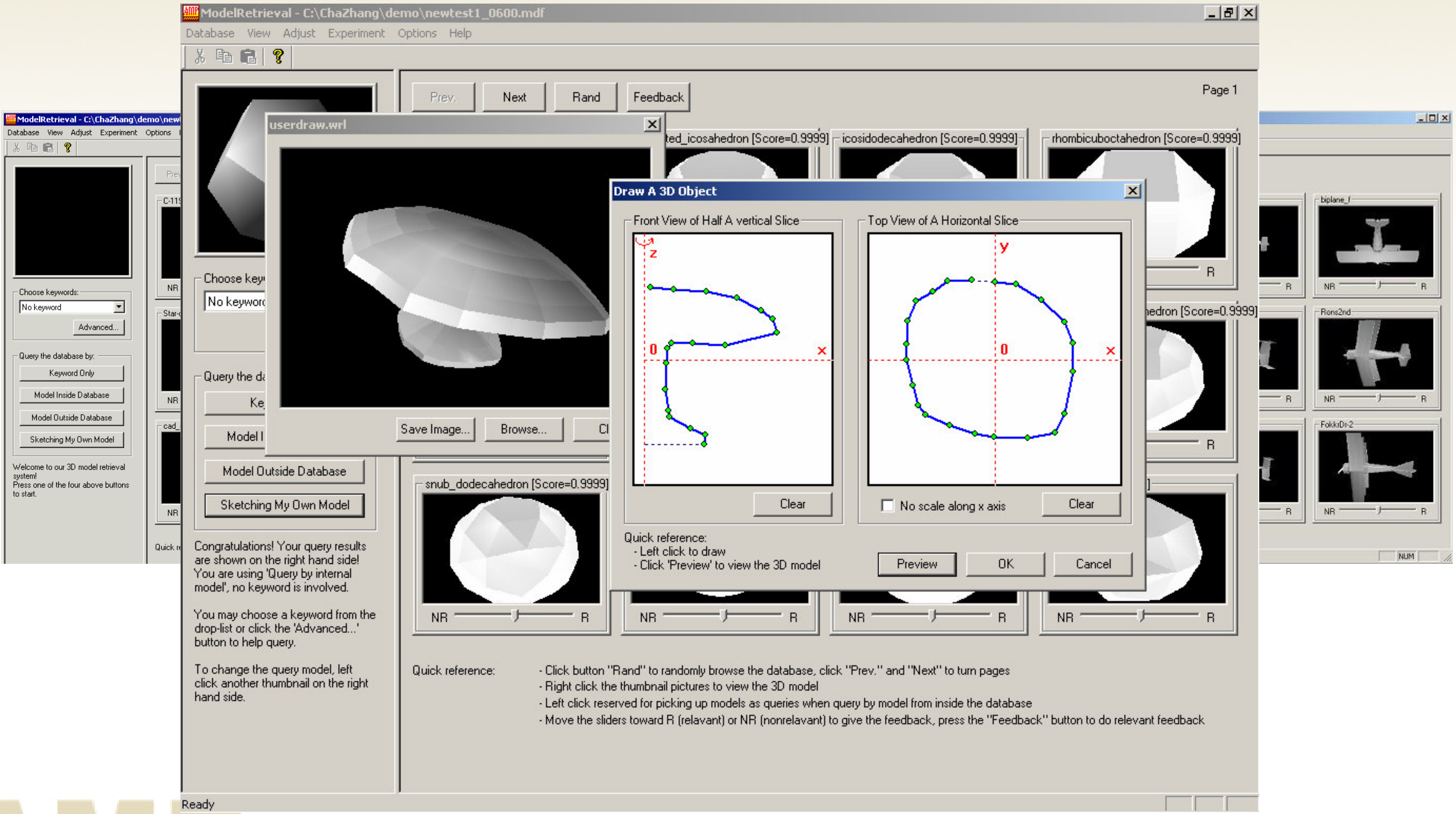
Page Stored in Database

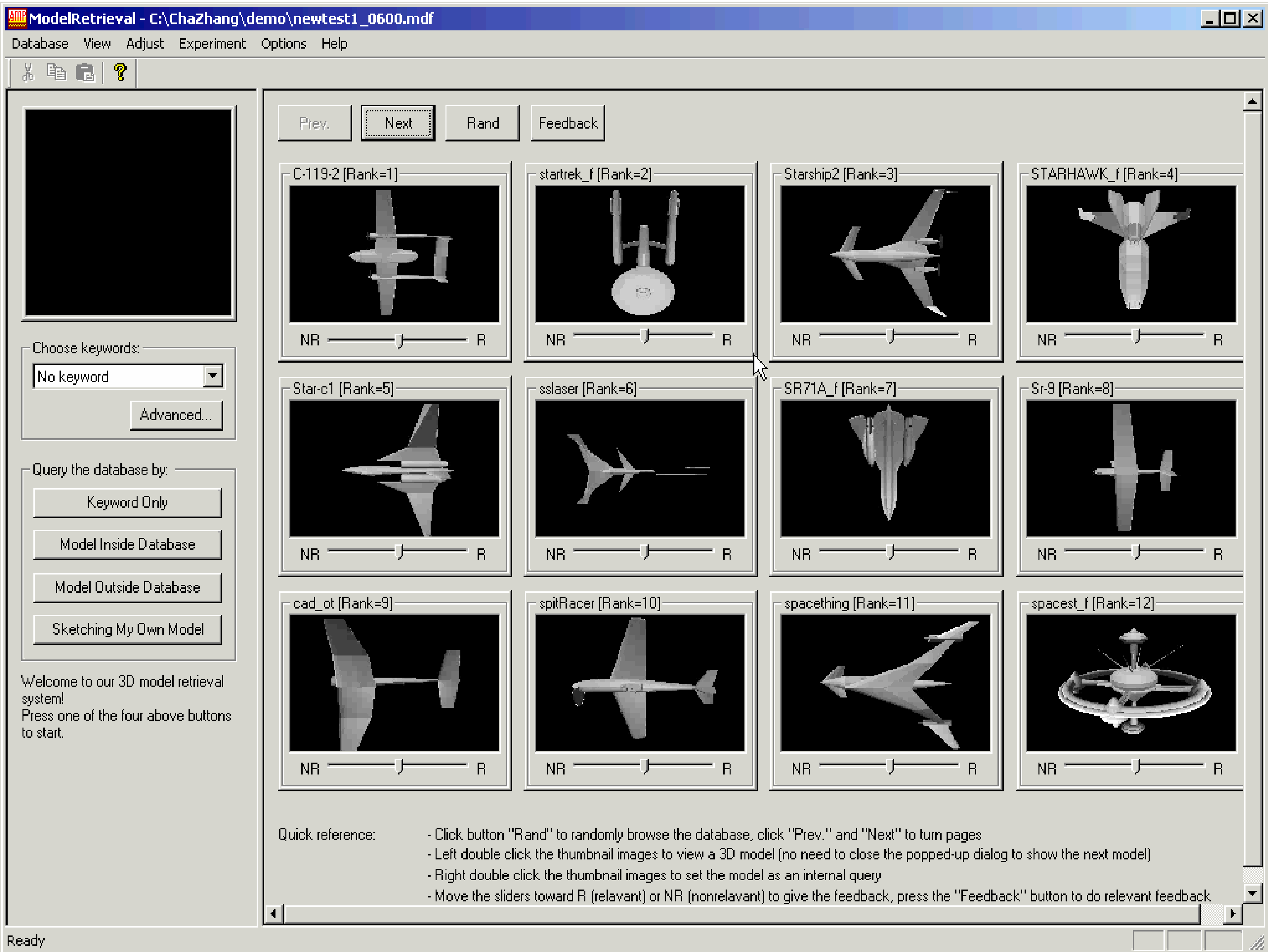
Query Sketch

Similar Sketch



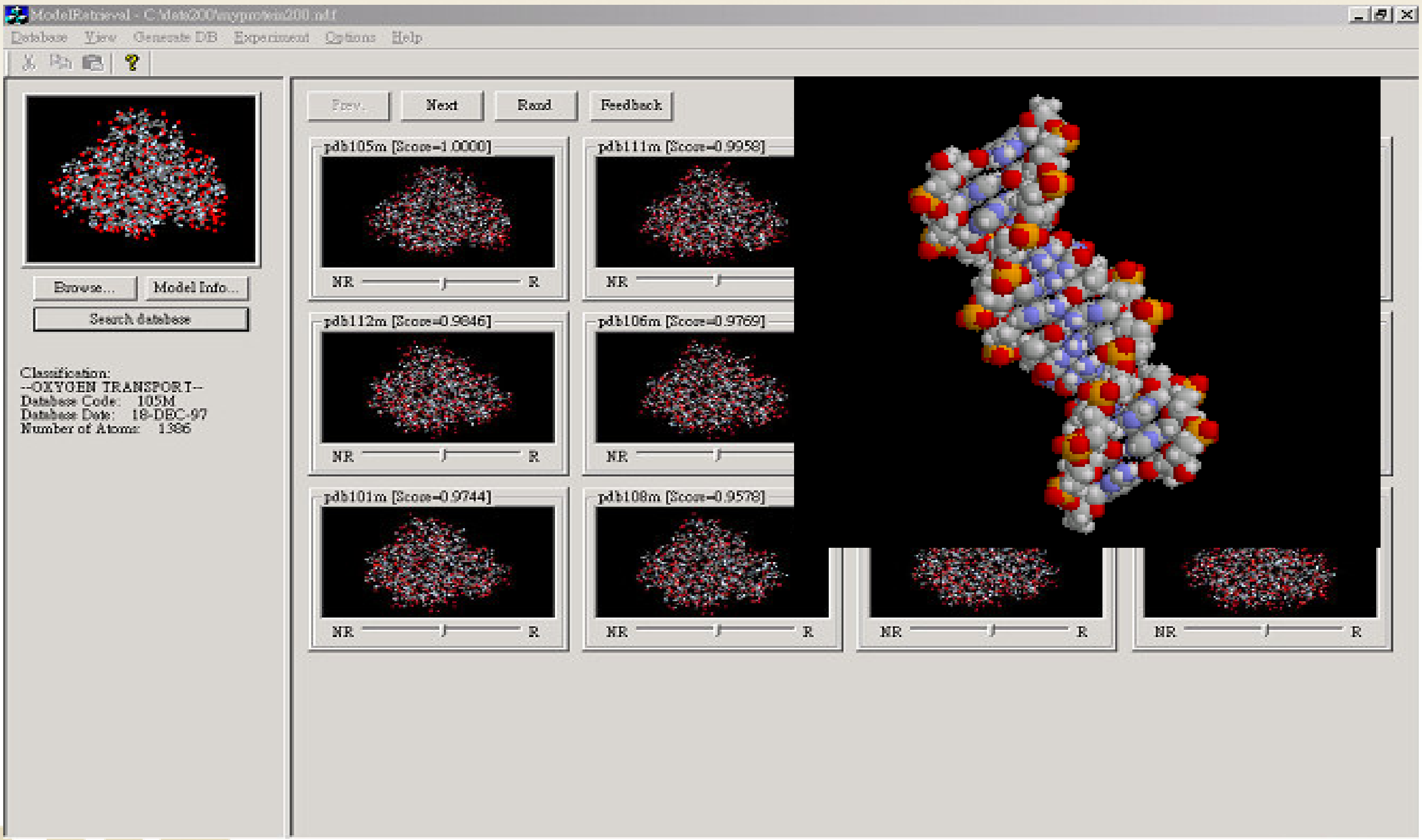
3D Objects [Zhang&Chen ACM MM'01]



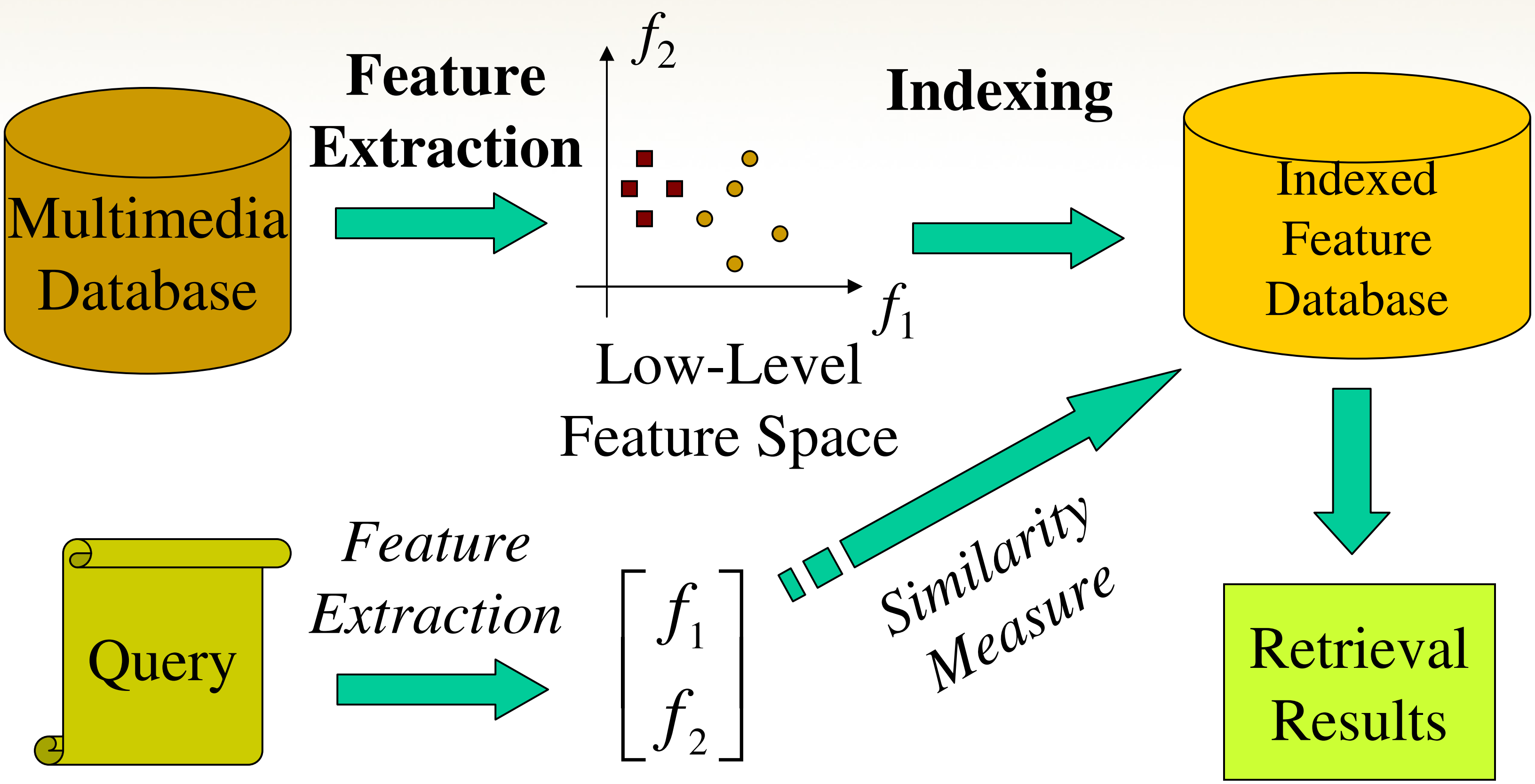

ModelRetrieval - C:\ChaZhang\demo\newtest1_0600.mdf
Database View Adjust Experiment Options Help
Prev.
Next
Rand
Feedback
C-119-2 [Rank=1]
startrek_f [Rank=2]
Starship2 [Rank=3]
STARHAWK_f [Rank=4]
Star-c1 [Rank=5]
sslaser [Rank=6]
SR71A_f [Rank=7]
Sr-9 [Rank=8]
cad_ot [Rank=9]
spitRacer [Rank=10]
spacething [Rank=11]
spacest_f [Rank=12]
NR
R
Choose keywords:
No keyword
Advanced...
Query the database by:
Keyword Only
Model Inside Database
Model Outside Database
Sketching My Own Model
Welcome to our 3D model retrieval system!
Press one of the four above buttons to start.
Quick reference:
- Click button "Rand" to randomly browse the database, click "Prev." and "Next" to turn pages
- Left double click the thumbnail images to view a 3D model (no need to close the popped-up dialog to show the next model)
- Right double click the thumbnail images to set the model as an internal query
- Move the sliders toward R (relavant) or NR (nonrelavant) to give the feedback, press the "Feedback" button to do relevant feedback
Ready

3D Protein Structures [Chen&Chen ICIP'02]

Carnegie Mellon



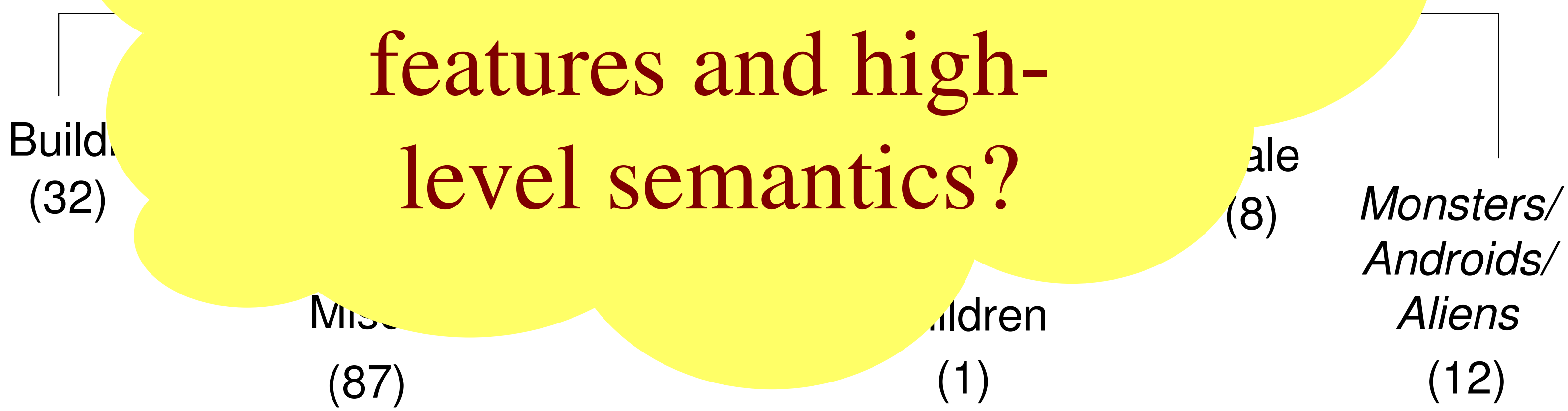
Content-Based Information Retrieval (CBIR)



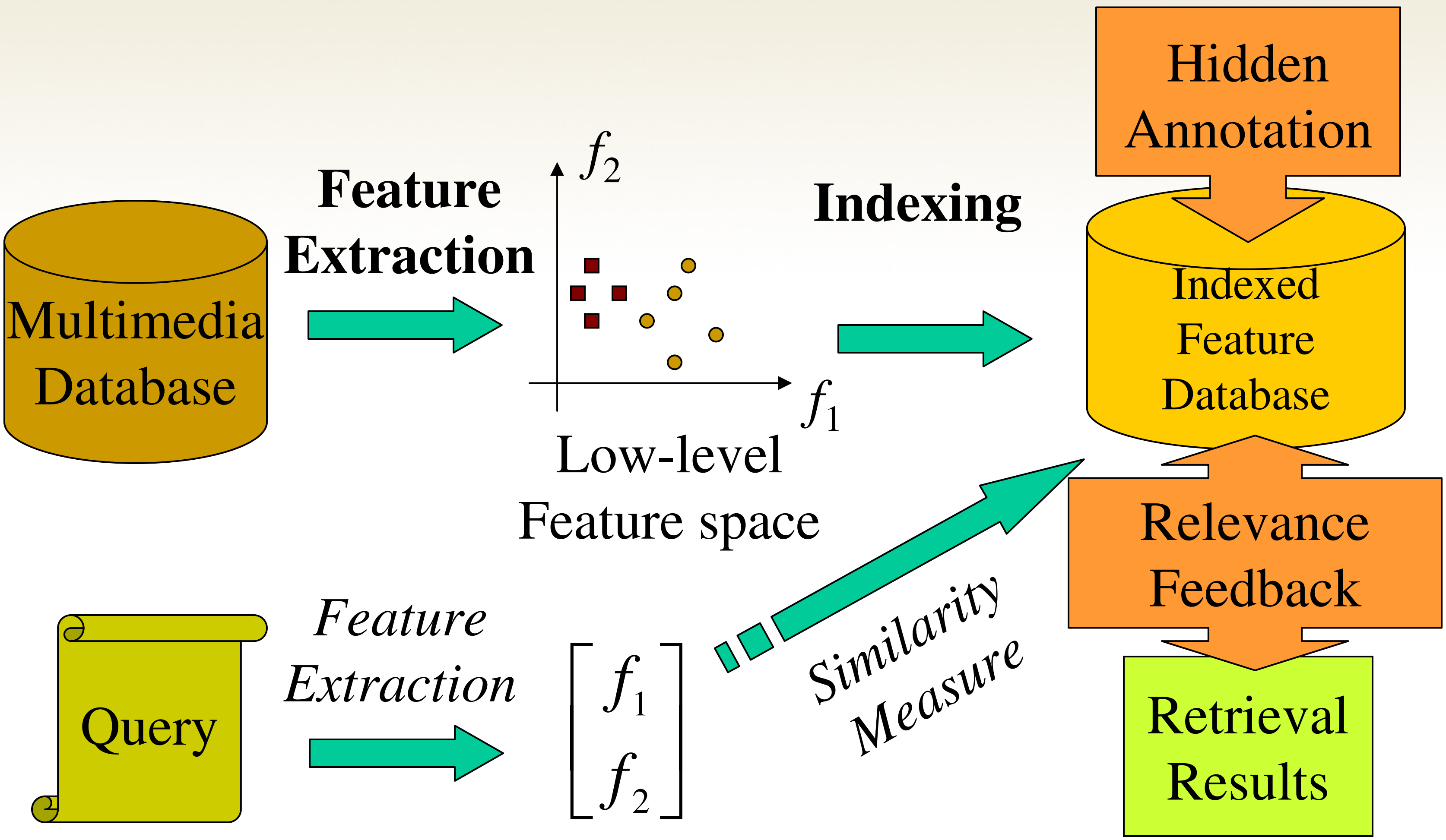
High-Level Semantics

- e.g. hierarchical

BIG CHALLENGE:
How to bridge the gap between low-level features and high-level semantics?



Possible Solutions



Semantic Information

- Hidden annotation
  - Object # n has Attribute # k → explicit
- Relevance feedback
  - Objects # m and # n are (not) similar → implicit

- Q: How to represent and propagate semantic information?
- Q: How to use explicit/implicit semantic information to improve retrieval?

Semantic Information as Probabilities

	Attribute 1	Attribute 2	Attribute 3	…	Attribute K
Object 1	p_{11}	p_{12}	p_{13}	…	p_{1K}
Object 2	p_{21}	p_{22}	p_{23}	…	p_{2K}
⋮	⋮	⋮	⋮		⋮
Object N	p_{N1}	p_{N2}	p_{N3}	…	p_{NK}

p_{nk} : Attribute Probabilities

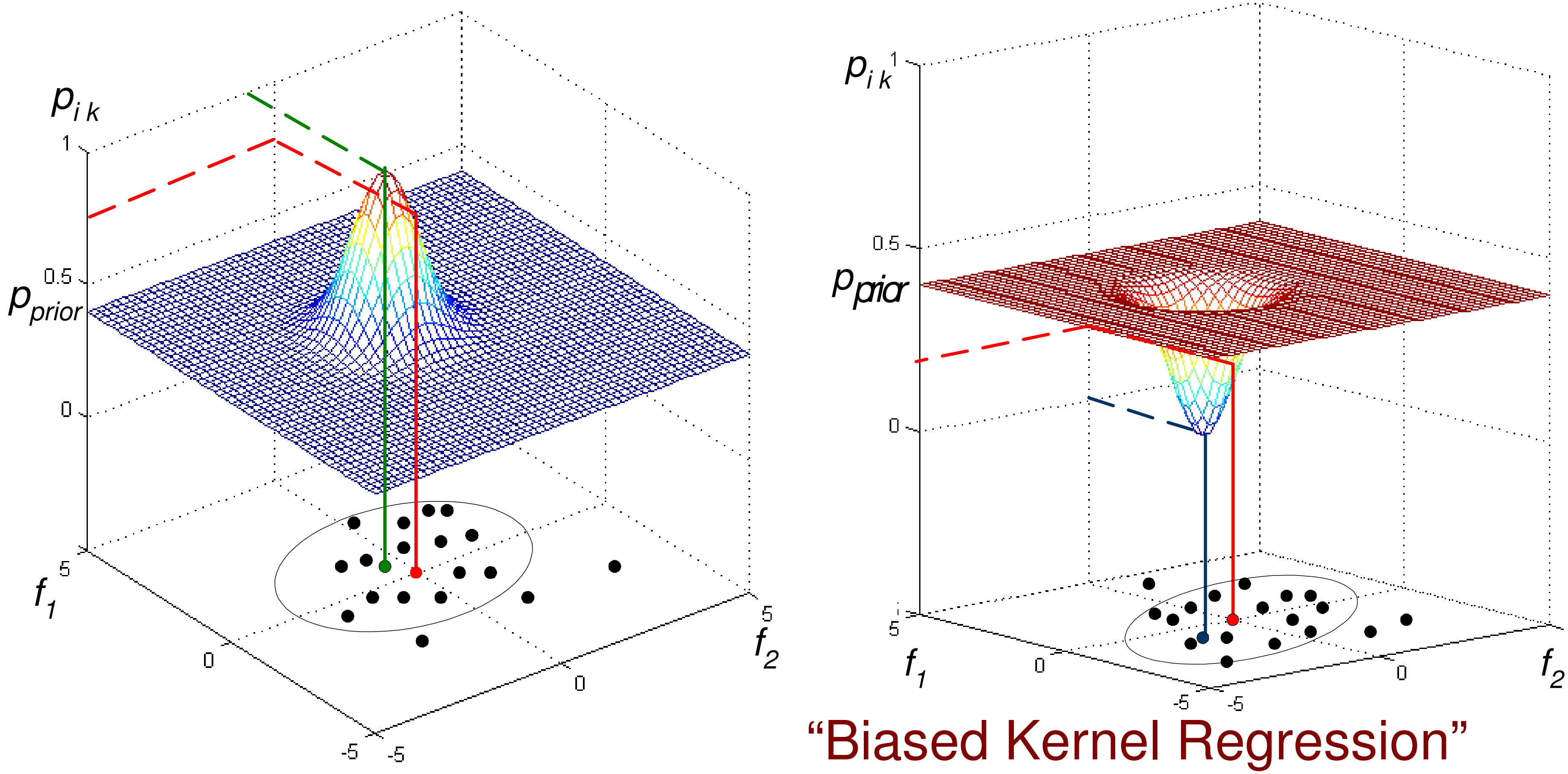
Annotate one object…

	Attribute 1	Attribute 2	Attribute 3	…	Attribute K
Object 1					
Object 2	**1**	**0**	**1**		**0**
⋮	⋮	⋮	⋮		⋮
Object N					

When an object is annotated, p_{nk} is set to 0/1

Q: How to propagate? A: Based on low level features

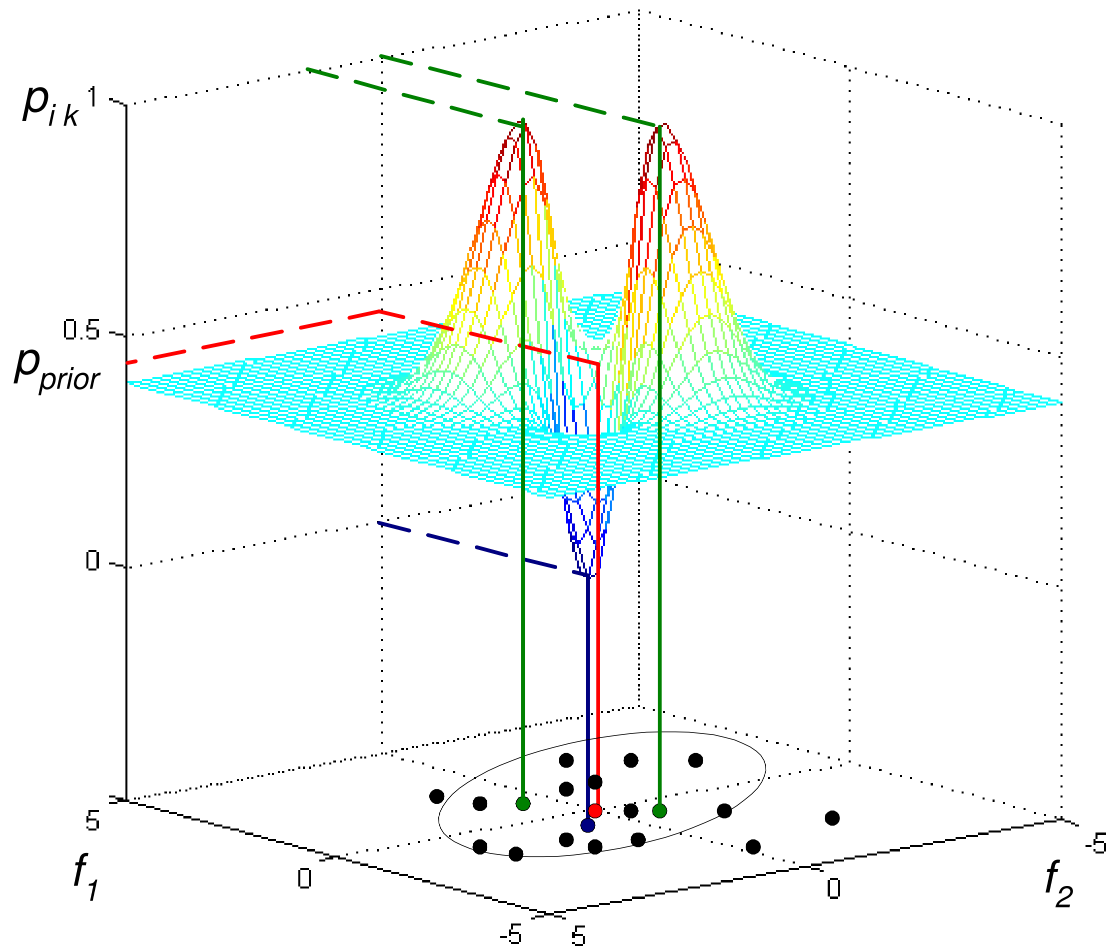
Semantic Propagation



"Biased Kernel Regression"

Chen2006

Semantic Propagation (cont.)



Chen2006

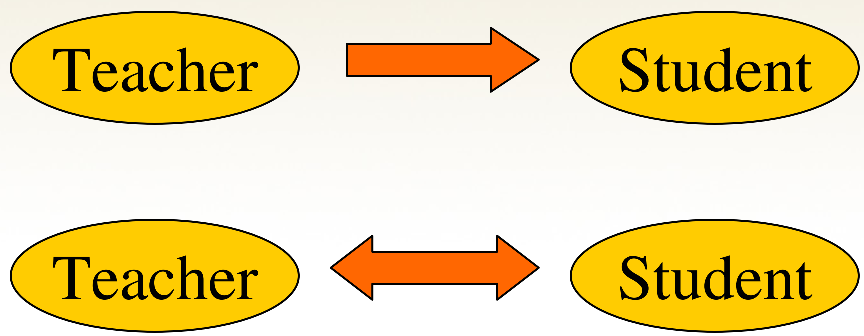
Semantic Propagation (cont.)

	Attribute 1	Attribute 2	Attribute 3	…	Attribute K
Object 1	p_{11}	p_{12}	p_{13}	…	p_{1K}
Object 2	**1**	**0**	**1**		**0**
⋮	⋮	⋮	⋮		⋮
Object N	p_{N1}	p_{N2}	p_{N3}	…	p_{NK}

Q: Which to annotate next?

Chen2006

Active Learning



Passive Learning

Active Learning

Annotator

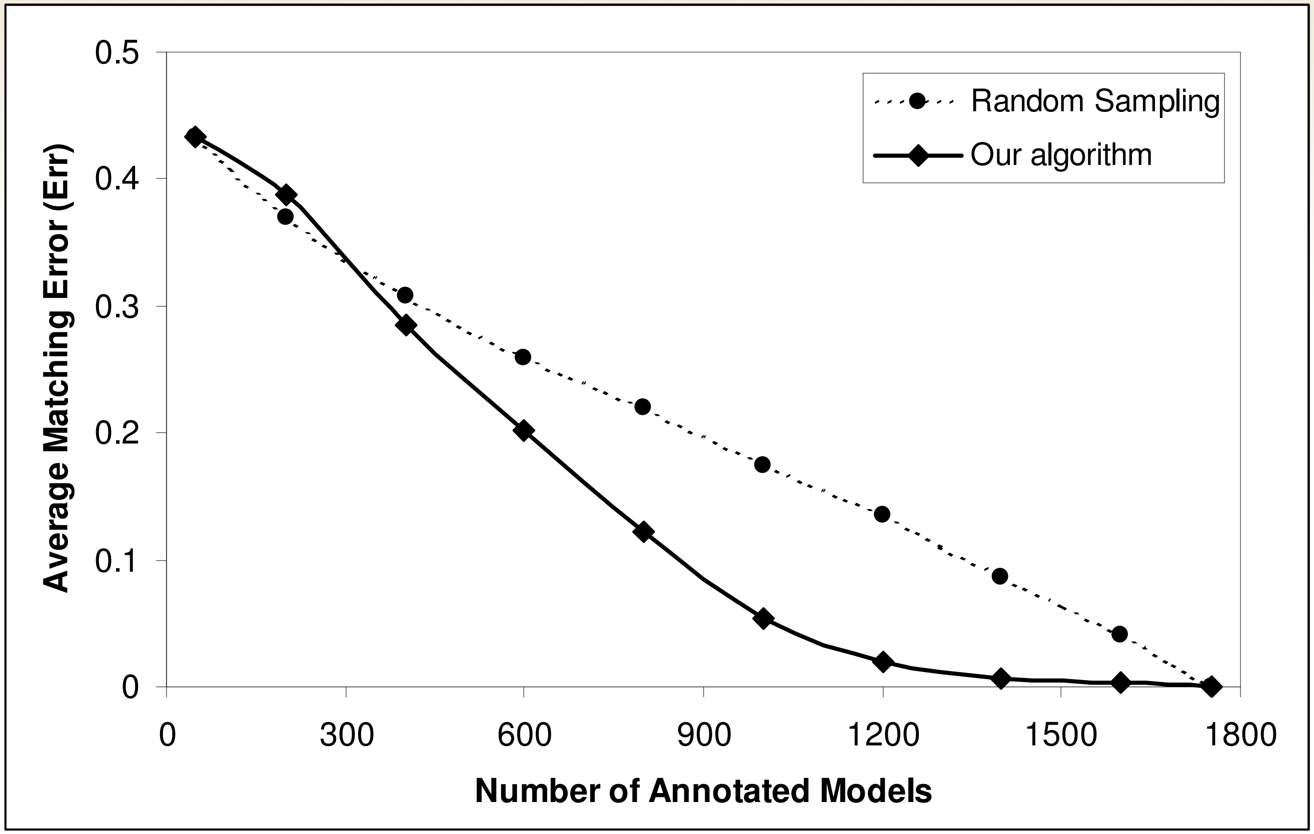
Retrieval System

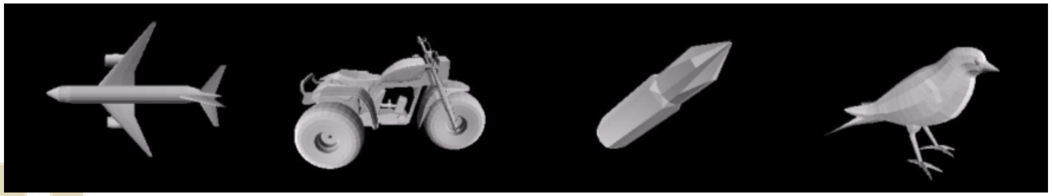
- Choose the most uncertain object to annotate
  - Uncertainty determined by the entropy of attribute probabilities
- "Selective sampling"
  - May want to consider density in feature space too

Recap…

- Maintain attribute probabilities of each object
- Set an attribute probability to 1/0 when annotated
- Propagate probabilities to non-annotated objects
- Choose the most uncertain object in the database to annotate next
  - Use probabilities to estimate uncertainty
- Use probabilities to measure semantic distance…

Result [Zhang&Chen T-MM'02]





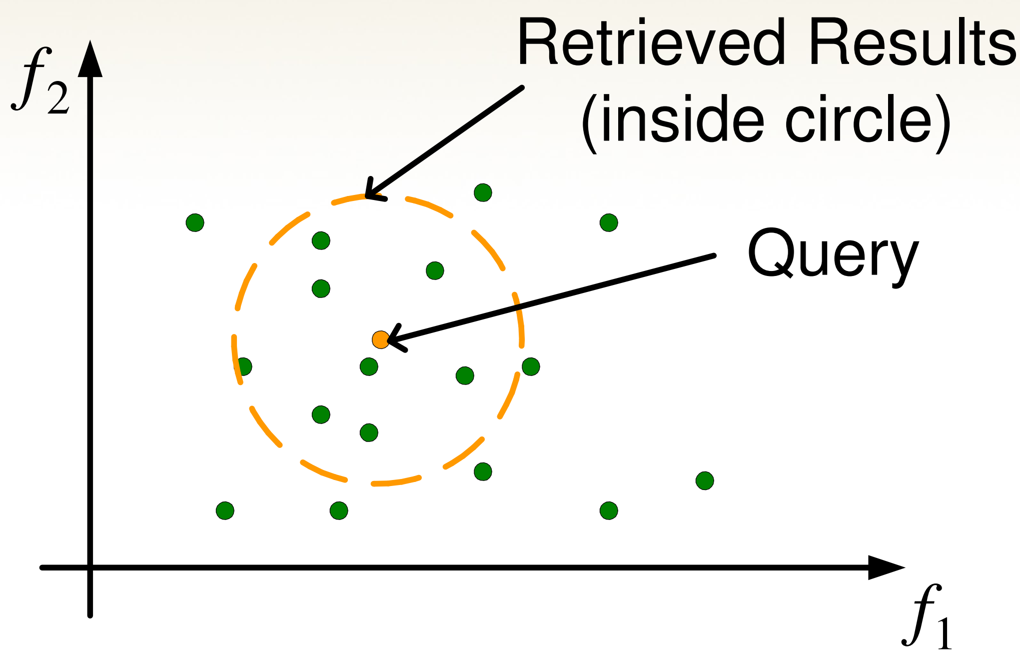
3D Objects
(1750 total)

Relevance Feedback

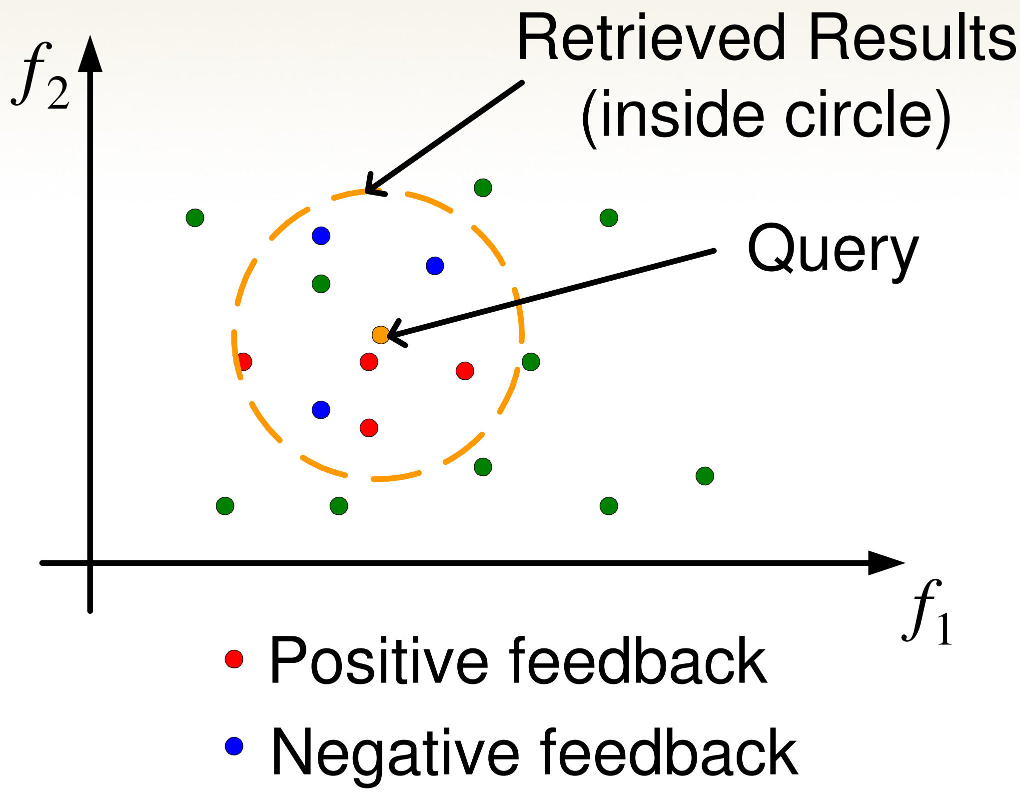
- Relevance feedback
  - Ask for user's feedback during the retrieval
    - "Object #*i* is (not) similar to the query"
    - "Objects #*m* and #*n* are (not) similar"
  - Implicit semantic information

- Use feedback to improve retrieval
  - Way 1: Move the query point
  - Way 2: Weigh the features
  - Way 3: "Warp" the feature space

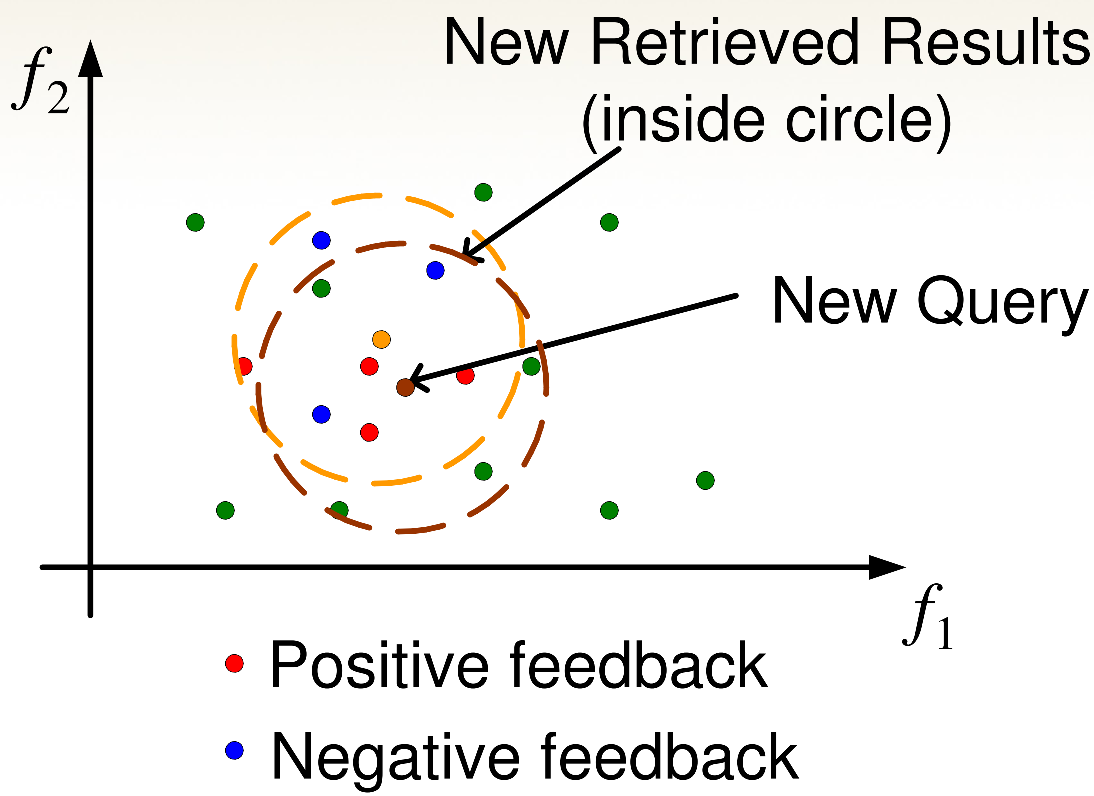
An Example



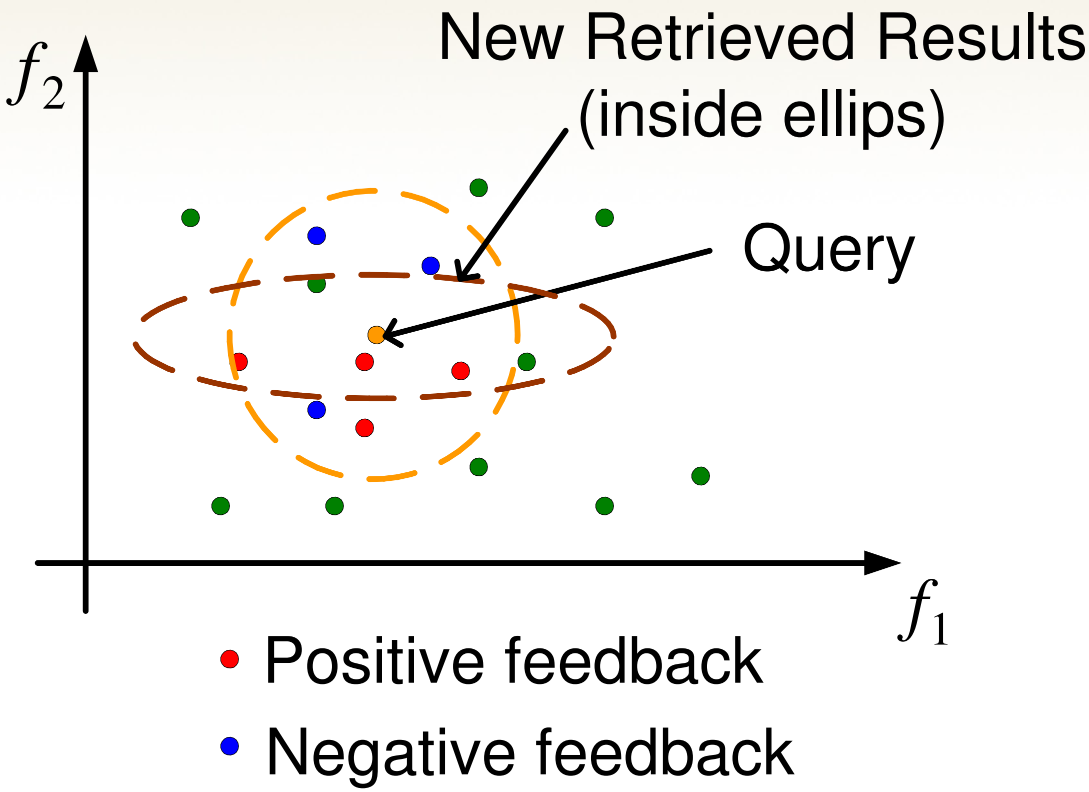
An Example


Retrieved Results
(inside circle)
Query
f_2
f_1
Positive feedback
Negative feedback

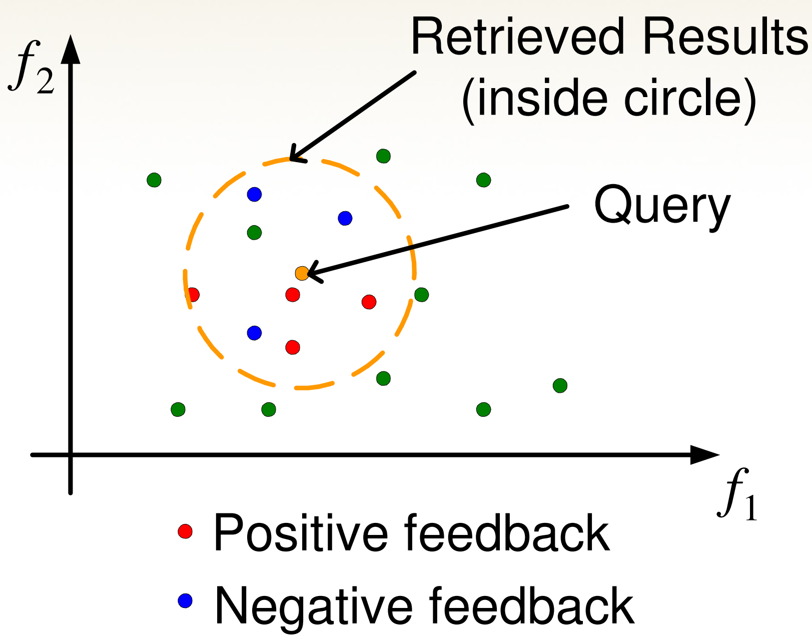
Move the Query Point



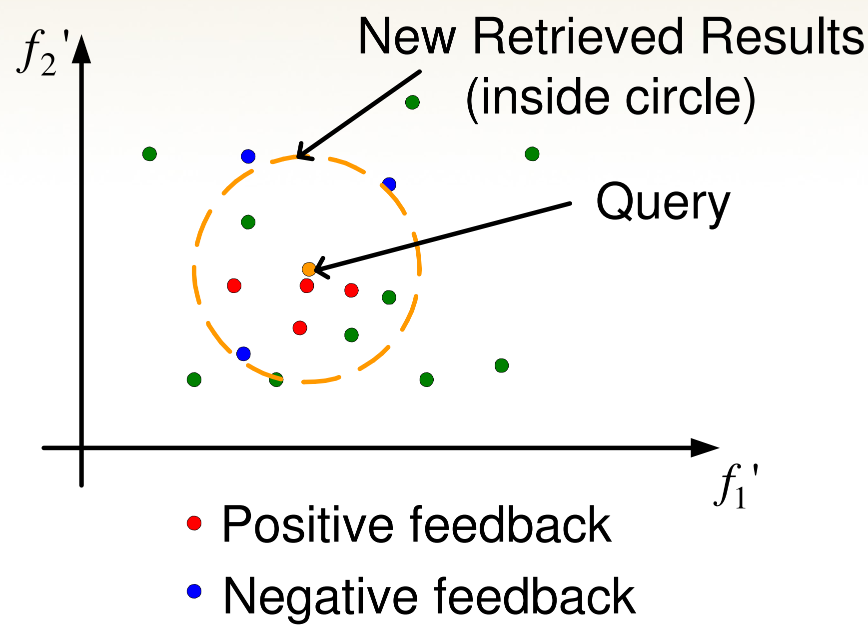
Feature Weighting



Feature Space Warping

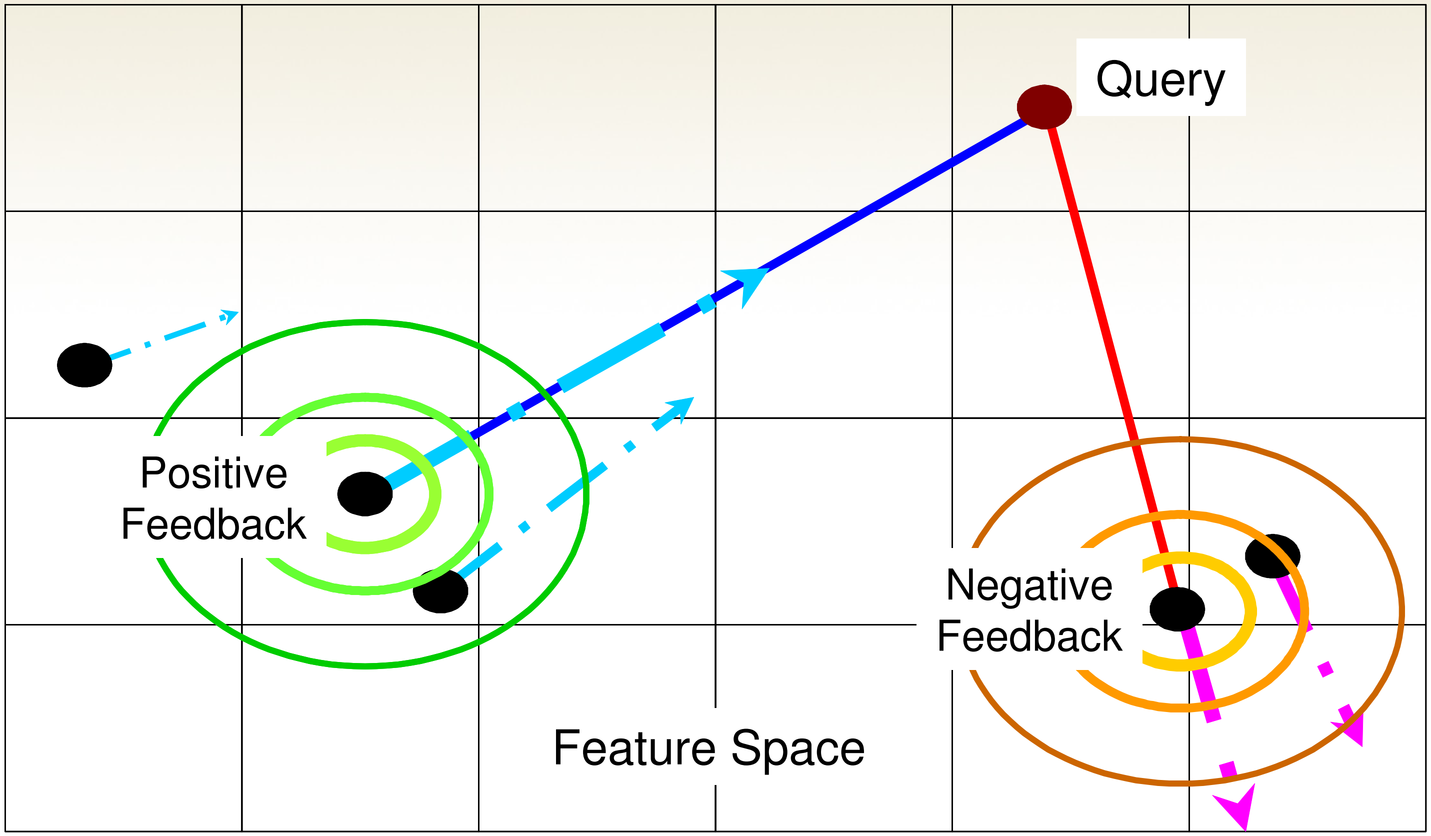


Before Warping



After Warping

Feature Space Warping



$$v_{pi} = \left[\gamma \sum_{j=1}^{M} u_i \exp\left(-c\left|v_{ij}\right|\right) \right] v_{iq}$$

This is also semantic propagation!!!

Experiment Result [Bang&Chen ICIP'02]

3D Objects
(1750 total)

Semantic Propagation is the Key

- Without semantic propagation, hidden annotation and relevance feedback are not very useful
- With enough relevance feedback, can we can accomplish information retrieval without low-level features at all?

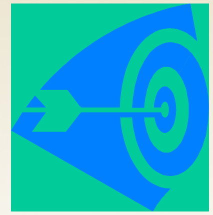
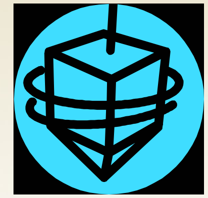
Pushing Content to Extreme

Content-Free Information Retrieval…

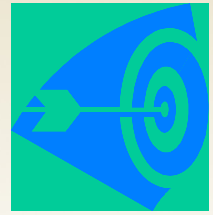
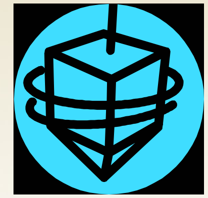
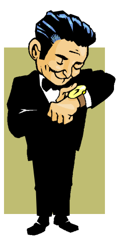
Content-Free Information Retrieval (CFIR)

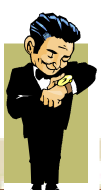
- With enough relevance feedback, retrieval is based more and more on feedback, less and less on features
- In the extreme case, retrieval based on feedback only
  - Retrieval based on user history
- e.g., Amazon.com

Example -- How CFIR Works

User History				
	1	0	0	1
	0	1	1	0
	0	1	1	?

Example -- How CFIR Works

User History				
	1	0	0	1
	0	1	1	0
	0	1	1	**0**

 and  are more similar than  and 

CBIR vs. CFIR

- Will user *U* like image *X*?
- Two different approaches:
  - Look at what *U* likes

    → Characterize images → Content-based IR
  - Look at which users like *X*

    → Characterize users → Content-free IR

Bayesian Framework

User History

Pair-wise conditional probability matrix

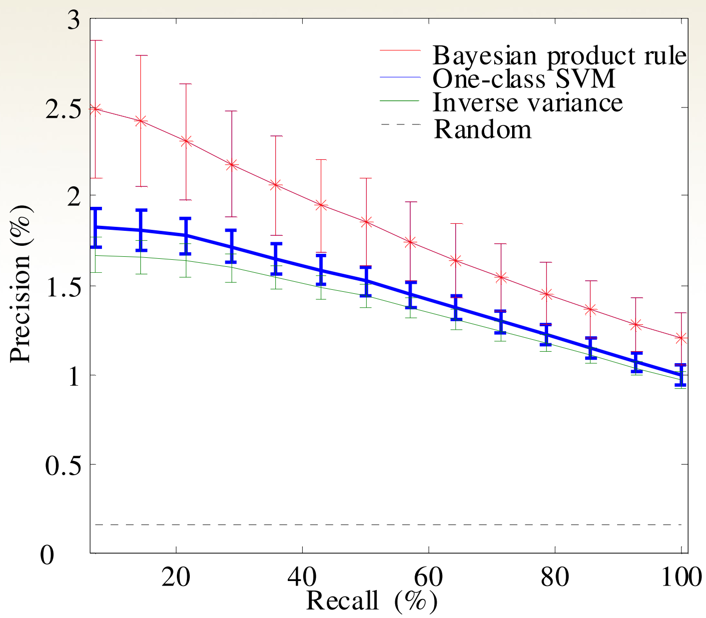
$$\tilde{P}(x_i = 1 \mid x_j = 1) \quad \forall i, j$$

Retrieval based on

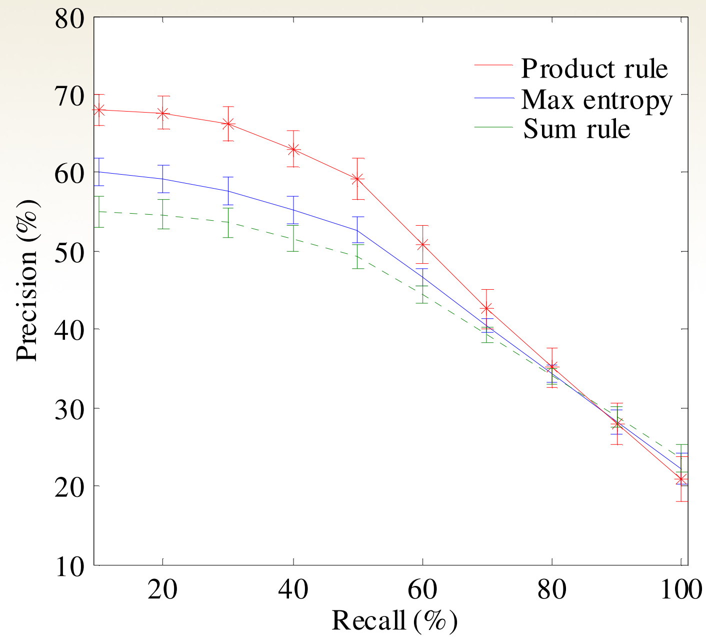
$$P(x_i = 1 \mid x_{j_1} = 1, \ldots, x_{j_F} = 1)$$

$$\propto \frac{\prod_{k=1}^{F} P(x_i = 1 \mid x_{j_k} = 1)}{P(x_i = 1)^{F-1}}$$

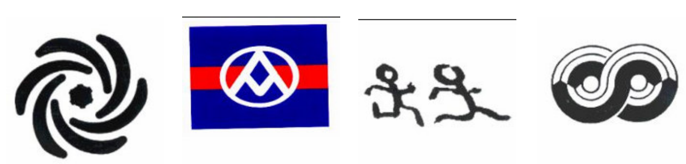
Experiment Results [Liu&Chen ICASSP'05]



CBIR Methods



CFIR Methods

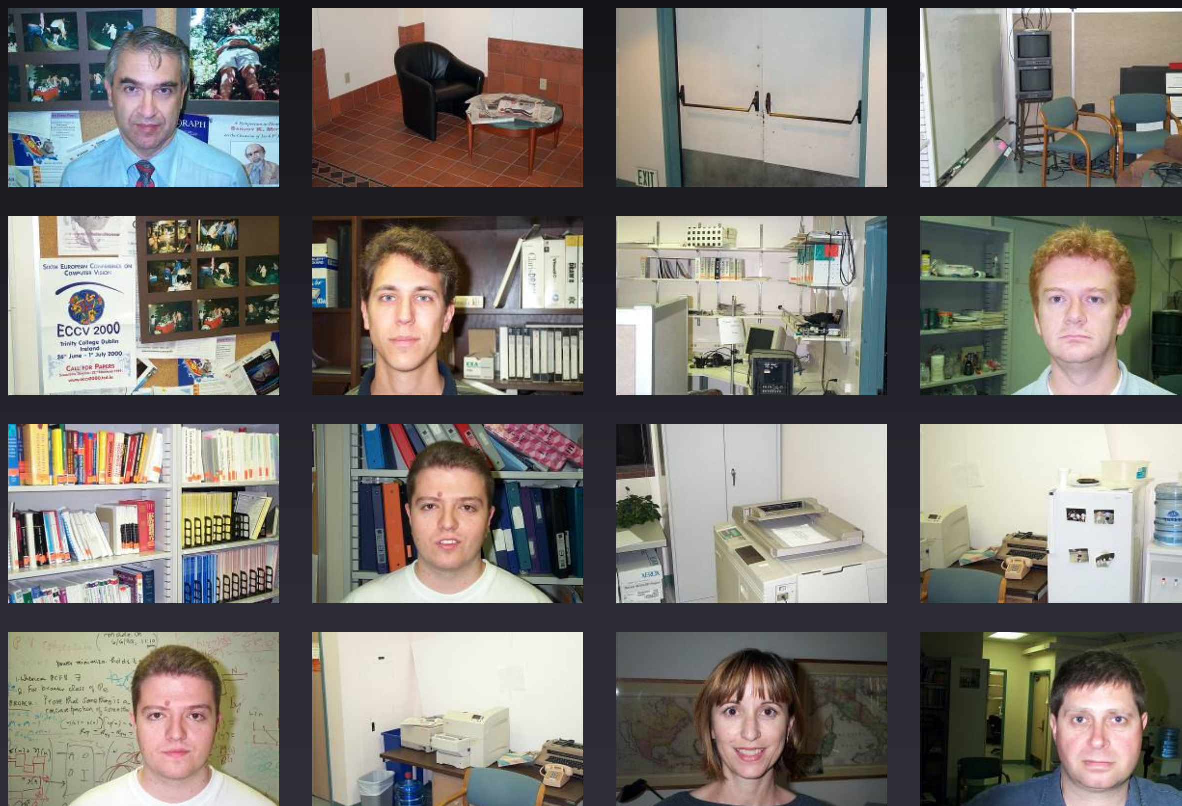


Sample Images

Content without User Feedback

Extracting content from nothing…

Unsupervised Image Categorization



[Caltech face + background dataset]

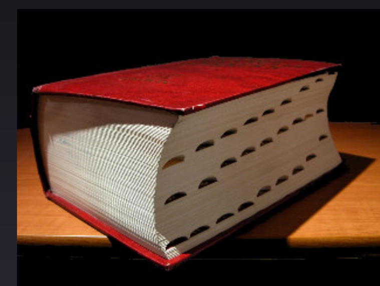
Unsupervised Image Categorization

[UIUC car dataset]

“Bag of Words” Representation

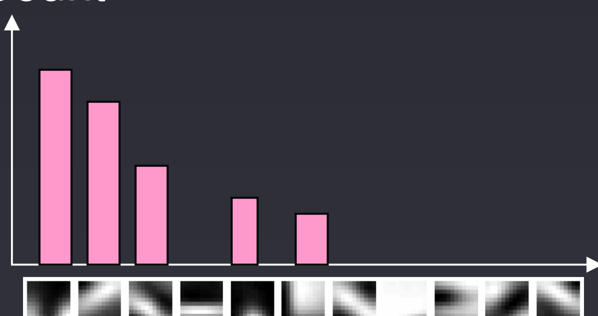
DoG interest point detector
+ SIFT descriptor [Lowe]

Vector
Quantization



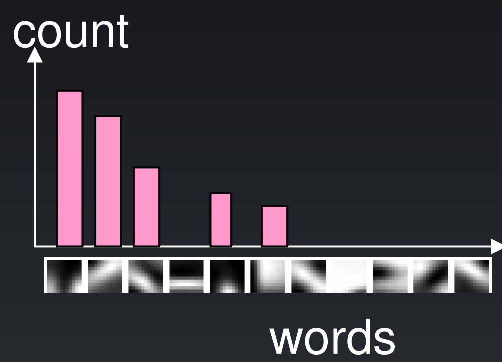
Codebook

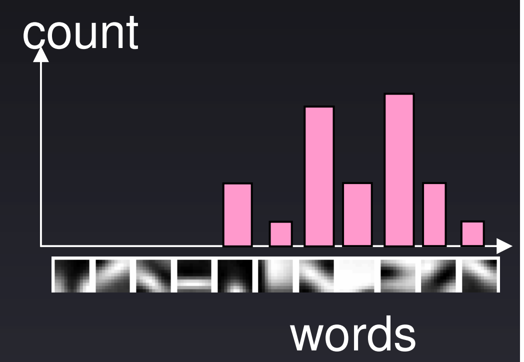
Count



Words

Graphical Model



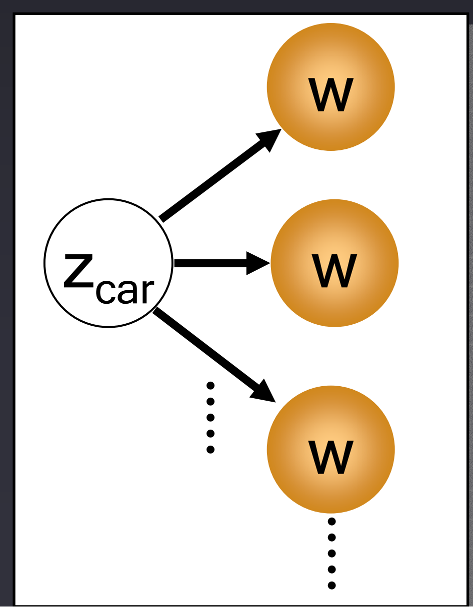


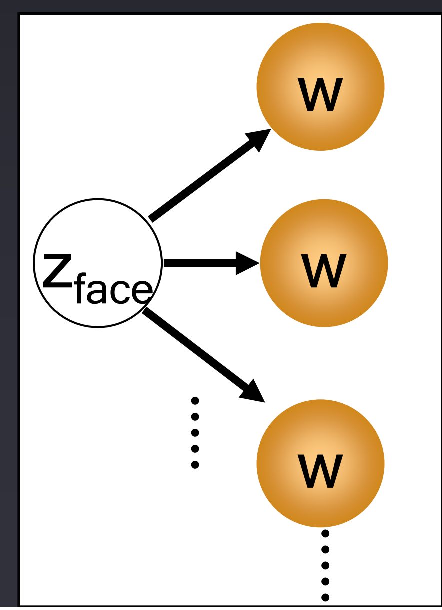
z : topic

w : word

$$P(z,w) = P(z)\boxed{P(w|z)}$$

Topic Appearance

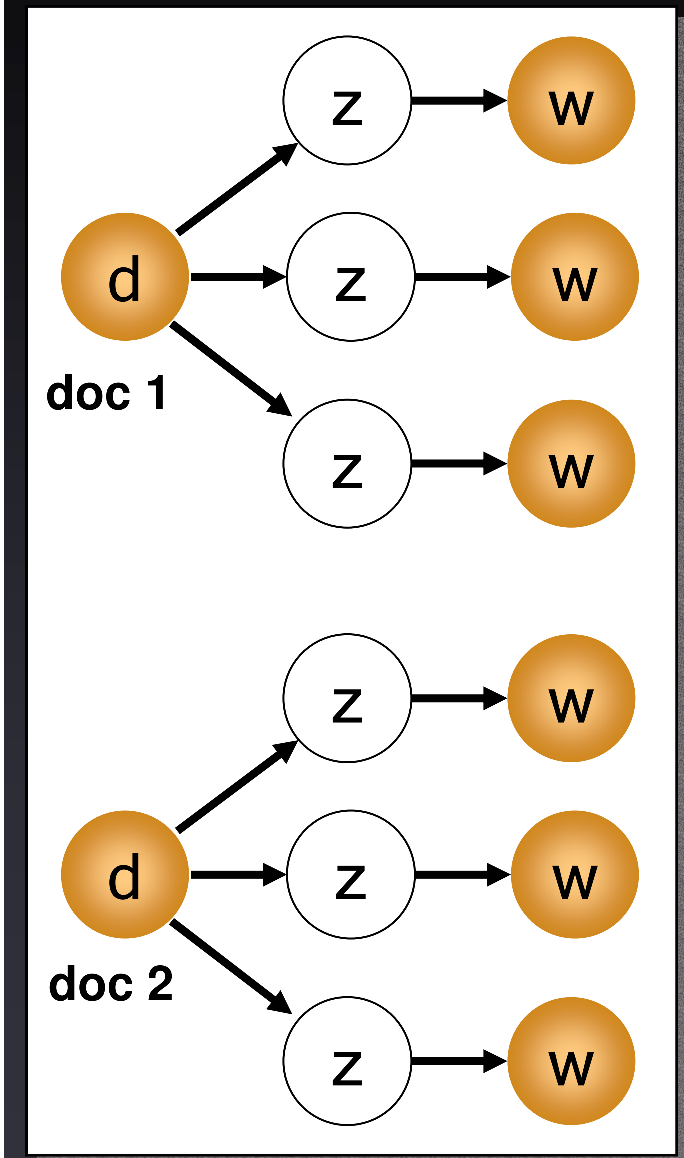




Need to handle background...

Solution: Add a hidden layer → PLSA

Probabilistic Latent Semantic Analysis (PLSA)



- Hofmann 01, Monay and Gatica-Perez 04, Sivic et al. 05, Quelhas et al. 05
- Can model complex scenes

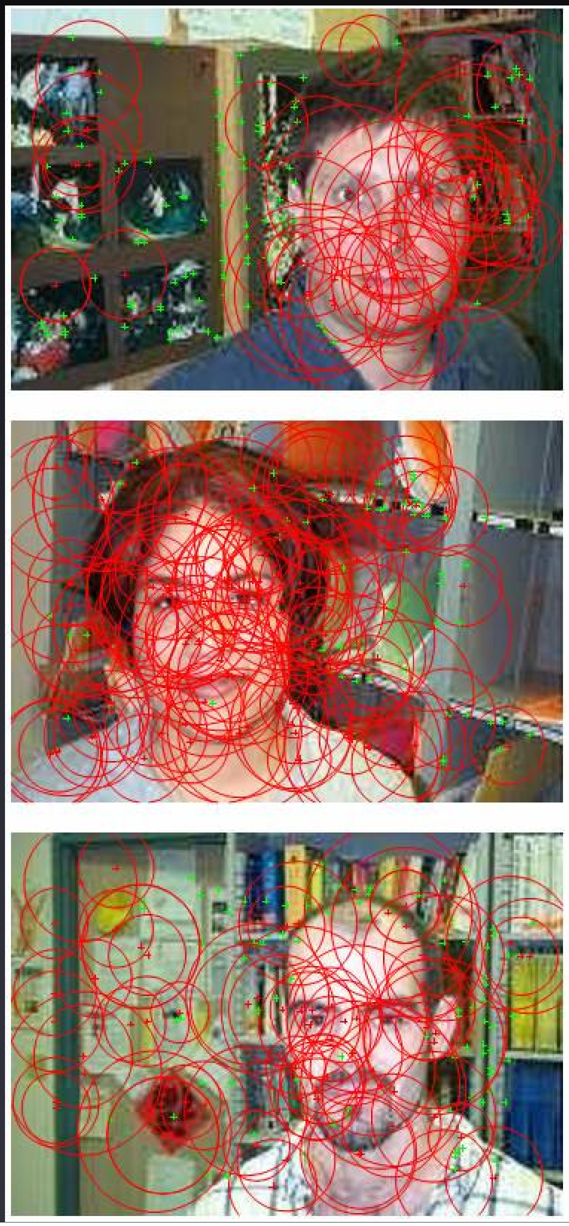
Document Characteristic *Topic Appearance*

$$P(d,z,w) = P(d)\boxed{P(z|d)}\boxed{P(w|z)}$$

- Inference:
  - Categorization $P(z|d)$
  - Segmentation $P(z|d,w)$
- Learning: Maximum likelihood estimation using EM algorithm

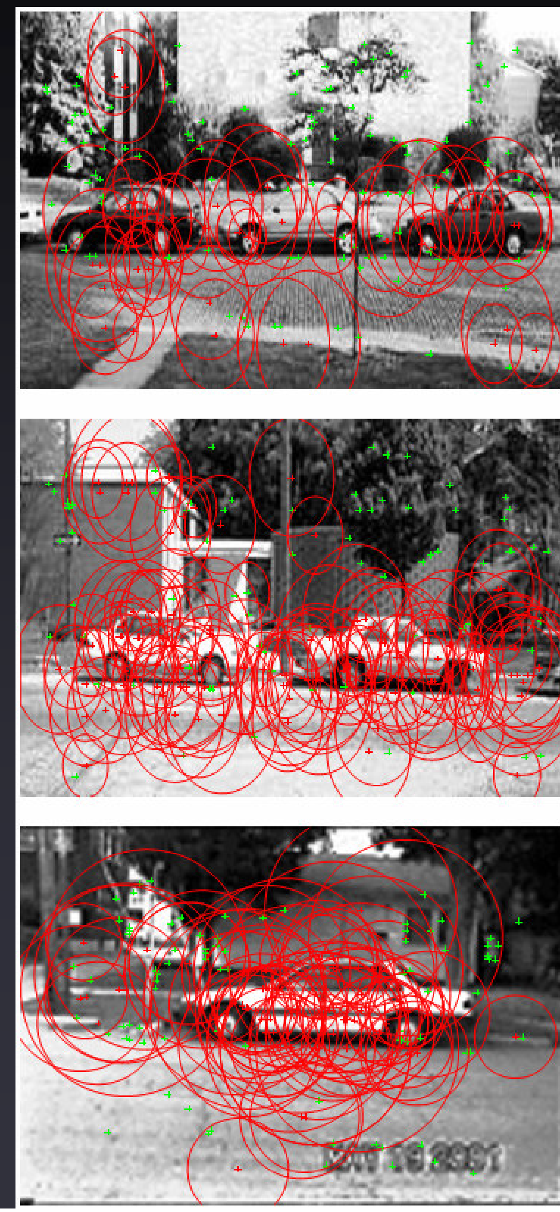
d : image
z : topic
w : word

Problem with PLSA



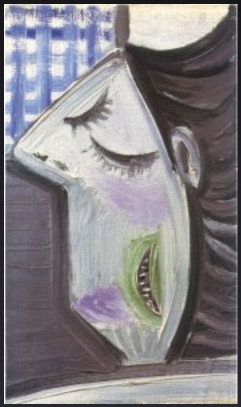
$$P(z|d,w) \geq 0.5$$

$$P(z|d,w) < 0.5$$

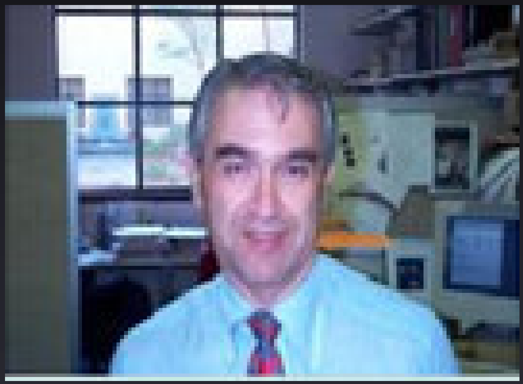


"Bag of Words" is the problem...

- As long as the parts are present, the exact position does not matter too much



Picasso, 1943



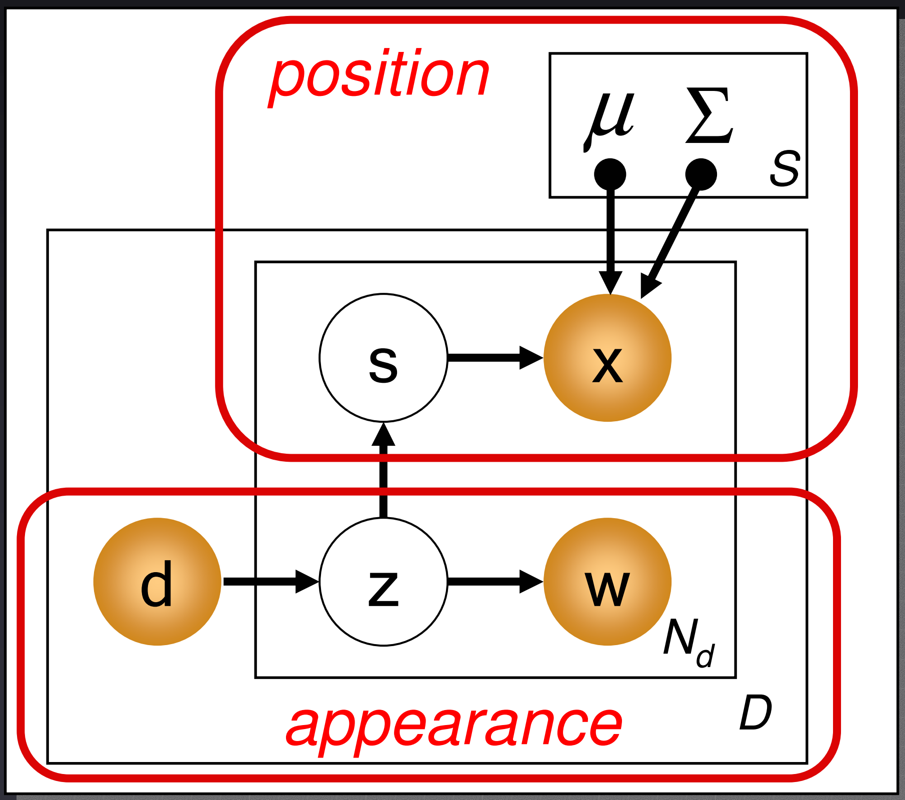
Dali, 1936

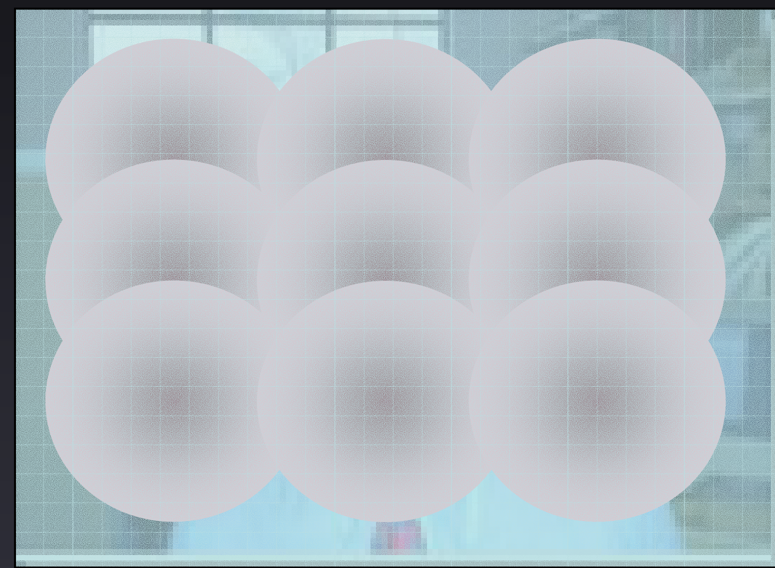
Not so for general objects!

Enforcing Clustering

[Liu&Chen ICIP'06]

- A number of S =10 *fixed* spatial distributions



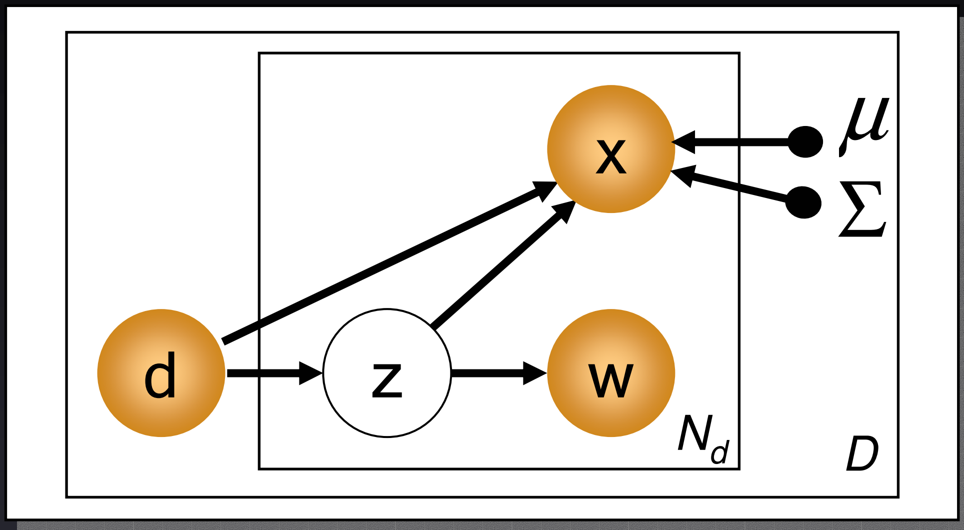


s_1 to s_9

s_{10}

Enforcing Clustering: Semantic-Shift

[Liu&Chen CVPR'06]

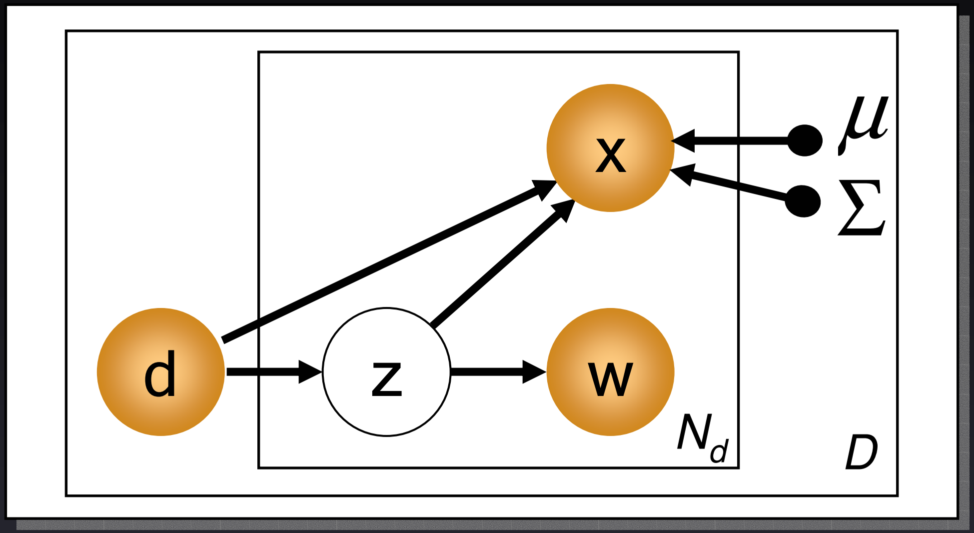


d : image
z : topic
w : word appearance
x : word position

Document Characteristic
Topic Appearance
Location Semantics

$$p(d,z,w,x) = P(d)\boxed{P(z|d)}\boxed{P(w|z)}\boxed{p(x|z,d)}$$

Representing Location Semantics



- Assume single foreground object
- Location semantics $p(x|z,d)$
  - Foreground:

$$p(x|z_{\mathrm{FG}}, d_i) \equiv \mathcal{N}(x|\mu_i, \Sigma_i)$$

  - Background:

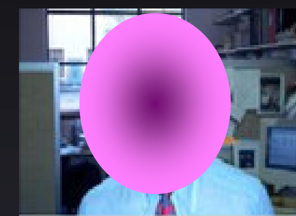
Complement of foreground distribution

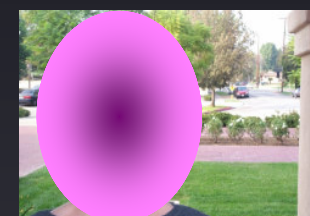
Learning in Semantic-Shift

$$P(z_k|d_i, w_j, x_p^{d_i}) \propto \boxed{p(x_p^{d_i}|z_k, d_i)}\boxed{P(z_k|d_i)P(w_j|z_k)}$$

posterior

$$\mathcal{N}(x|\mu_i, \Sigma_i)$$





Document Characteristic

$$P(z_k|d_i) \propto \sum_j \sum_p n_{ijp} P(z_k|d_i, w_j, x_p^{d_i})$$

Topic Appearance

$$P(w_j|z_k) \propto \sum_i \sum_p n_{ijp} P(z_k|d_i, w_j, x_p^{d_i})$$

$n_{ijp} \equiv n(d_i, w_j, x_p)$: bag of word position pairs

Learning in Semantic-Shift

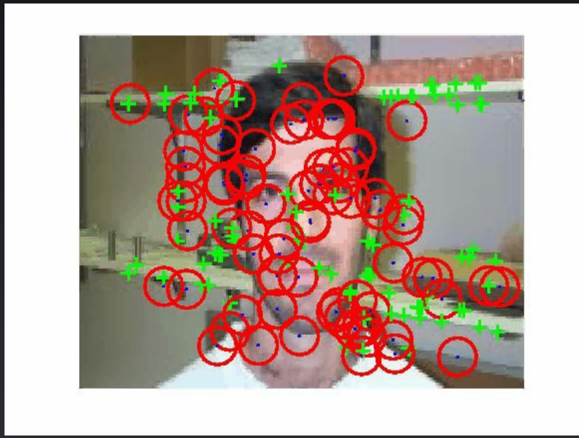
$p(x|z,d)$

Location semantics

This is why "semantic-shift"

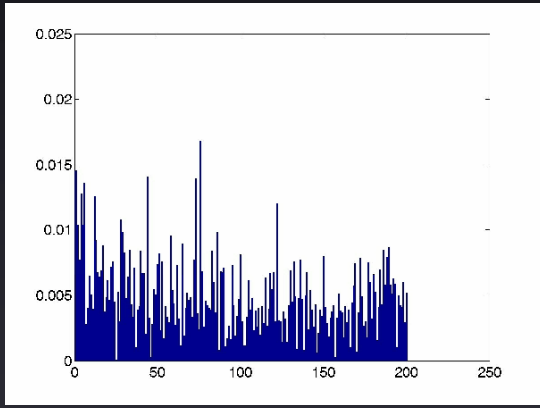
$P(z|d,w,x)$

posterior



$P(w|z)$

Topic appearance



Learning all 3 terms simultaneously…

Completely unsupervised…

Results

[Liu&Chen CVPR'06]

PLSA Semantic Shift

0.7 → 0.85

PLSA Semantic Shift

0.9 → 0.98

Future Work

- Intra-object modeling
- Video: spatial-temporal modeling
- Training the codewords in the loop
- Multiple ojects

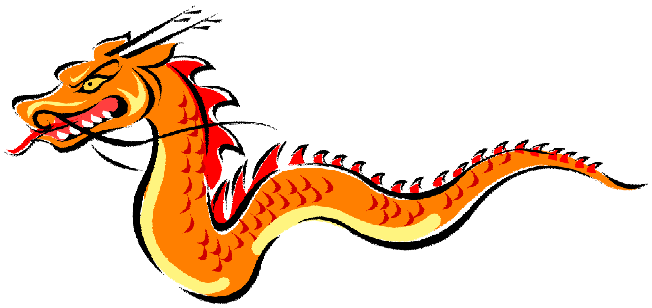
Conclusions

- Machine learning can bridge the gap between low-level features (bits) and high-level semantics (content)
- Hidden annotation and relevance feedback can help; semantic propagation is the key
  - Active learning
- “Content-free” information retrieval is possible
  - Bayesian framework
- Content extraction without user feedback is possible
  - Unsupervised learning; graphical models

Afterthoughts…

- Feng-Shui (风水)
  - Ancient Chinese room arrangement technique

- Way 1 (low-level):
  - Write down all the rules
  - Too many and do not generalize
- Way 2 (high-level):
  - Imagine how a dragon would move through the room to arrange it in a livable manner
  - Intuitive and creative
  - Done by some Feng-Shui masters



Advanced Multimedia Processing Lab

Please visit us at:

http://amp.ece.cmu.edu