From Pixels to Semantics –

Mining Digital Imagery Data for Automatic Linguistic Indexing of Pictures

James Z. Wang

Assistant Professor, endowed PNC Professorship

School of Information Sciences and Technology

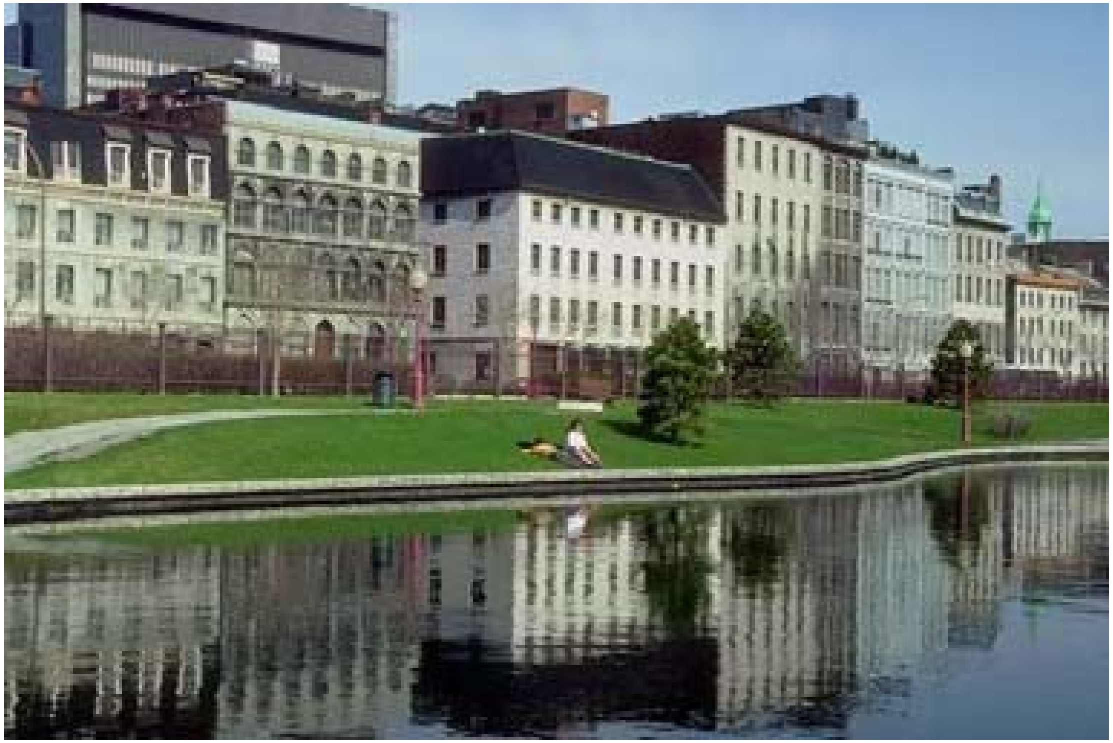
Jia Li

Assistant Professor, Statistics

The Pennsylvania State University

http://wang.ist.psu.edu

Poll: Can a computer do this?



- "Building, sky, lake, landscape, Europe, tree"

Outline

- Introduction
- Our related SIMPLIcity work
- ALIP: Automatic modeling and learning of concepts
- Conclusions and future work

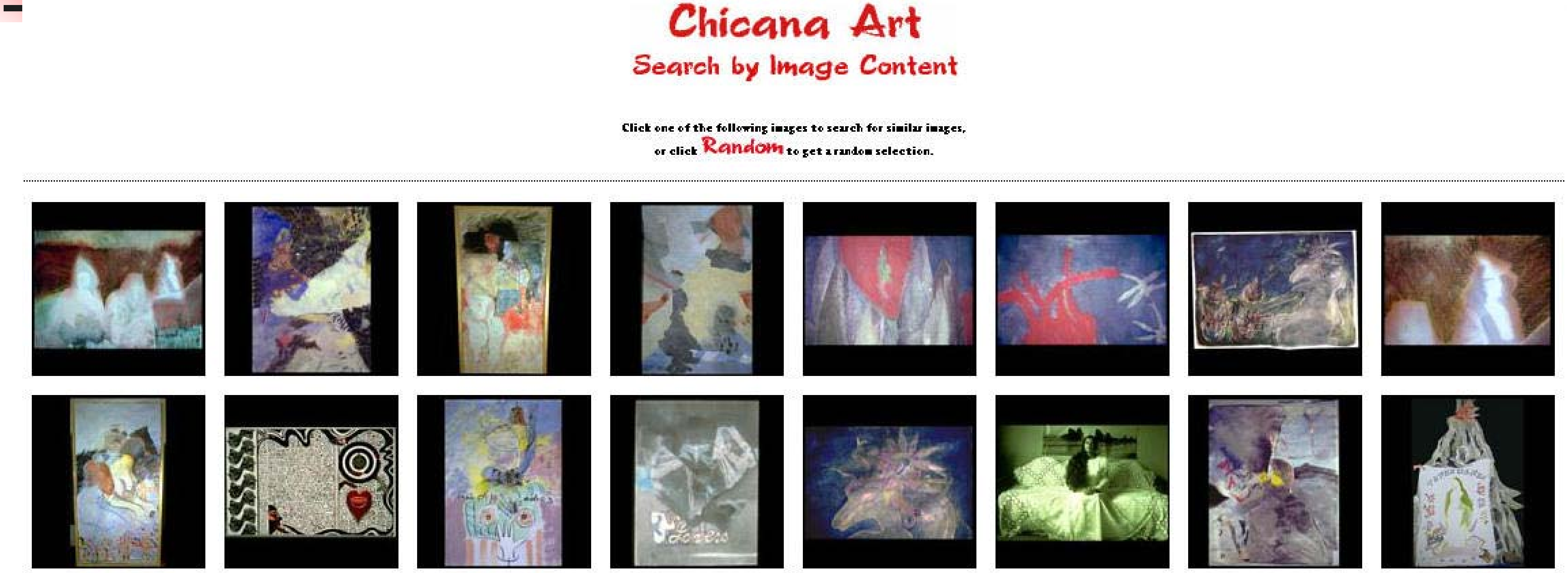
Automatic Linguistic Indexing of Pictures (ALIP)

- A new research direction for data miners
- Differences from computer vision
  - ALIP: deal with a large number of concepts
  - ALIP: rarely find enough number of "good" (diversified/3D?) training images
  - ALIP: build knowledge bases **automatically** for real-time linguistic indexing (generic method)
  - ALIP: highly interdisciplinary (AI, statistics, mining, imaging, applied math, domain knowledge, ……)
- Applications: biomedicine, homeland security, law enforcement, NASA, defense, commercial, cultural, education, entertainment, Web, ……

Related field: Image Retrieval

- The retrieval of relevant images from an image database on the basis of automatically-derived image features
- Our approach:
  - Wavelets
  - Statistical modeling
  - Supervised and unsupervised learning…

Chicana Art Project, 1995



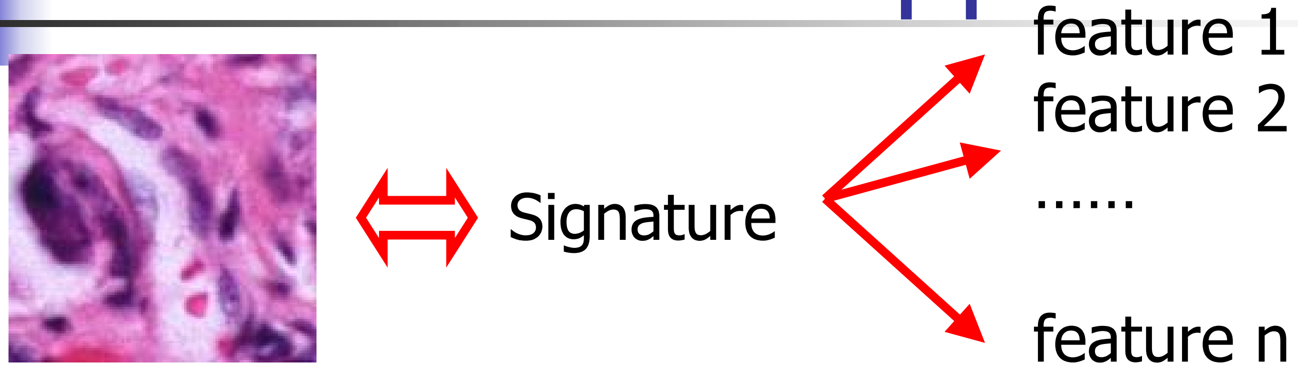
Chicana Art

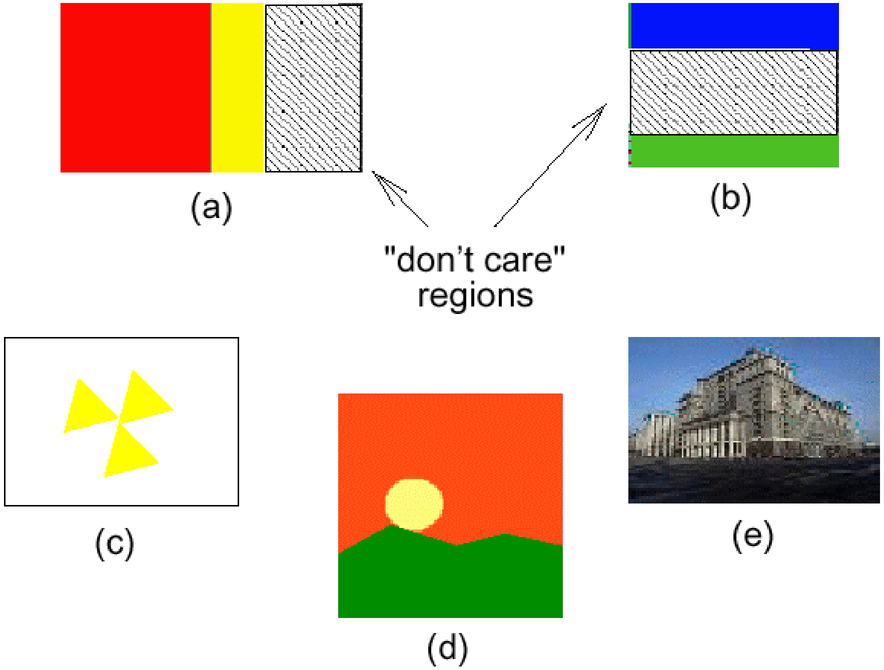
Search by Image Content

Click one of the following images to search for similar images,
or click Random to get a random selection.

- 1000+ high quality paintings of Stanford Art Library
- Goal: help students and researchers to find visually related paintings
- Used wavelet-based features [Wang+,1997]

Feature-based Approach





+ Handles low-level semantic queries

+ Many features can be extracted

-- Cannot handle higher-level queries (e.g.,objects)

Region-based Approach

- Extract objects from images first

+ Handles object-based queries

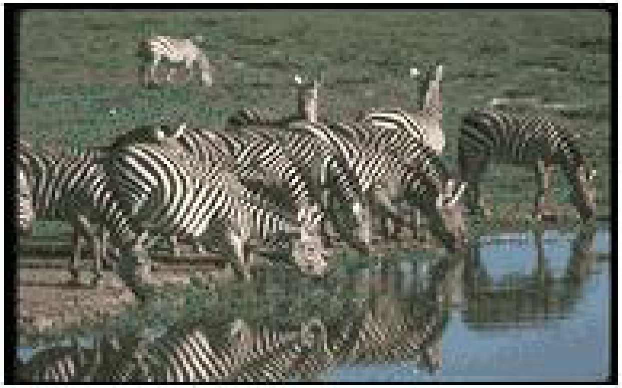
  e.g., find images with objects that are similar to some given objects

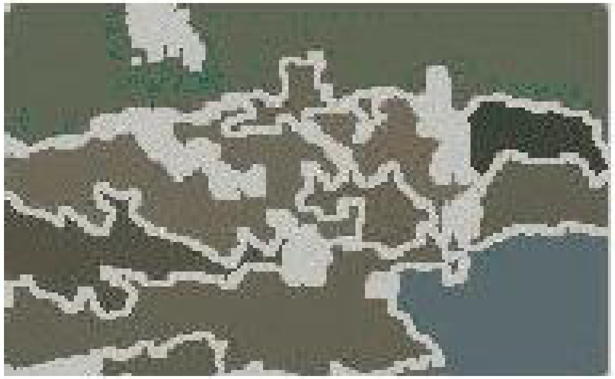
+ Reduce feature storage adaptively

-- Object segmentation is very difficult

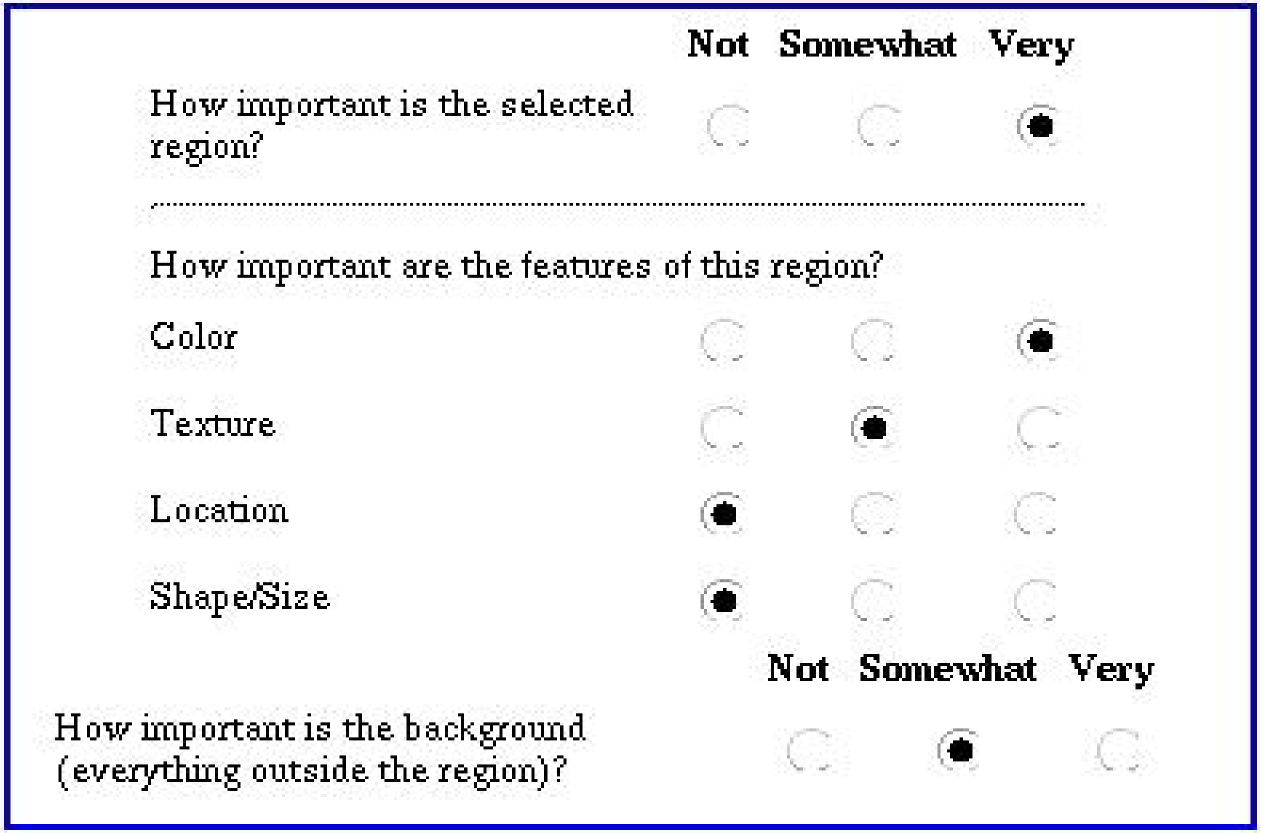
-- User interface: region marking, feature combination

1. Select up to two regions





2. Fill out this form for each region



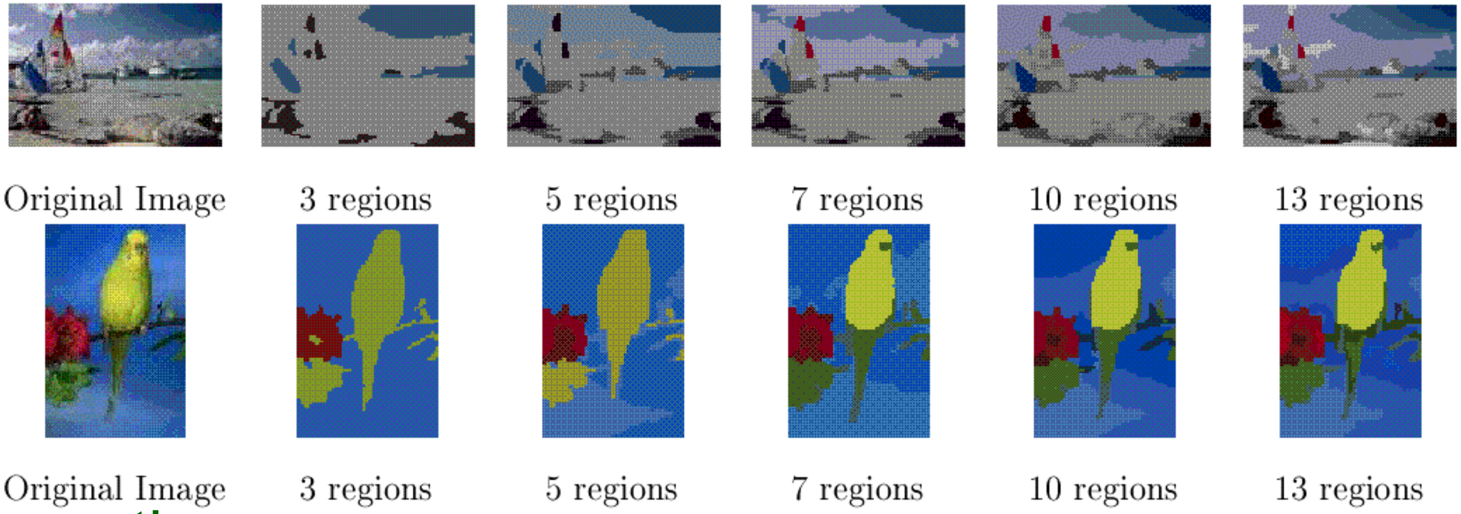
	Not	Somewhat	Very
How important is the selected region?	○	○	●
How important are the features of this region?			
Color	○	○	●
Texture	○	●	○
Location	●	○	○
Shape/Size	●	○	○
	Not	Somewhat	Very
How important is the background (everything outside the region)?	○	●	○

UCB Blobworld [Carson+, 1999]

Outline

- Introduction
- Our related SIMPLIcity work
- ALIP: Automatic modeling and learning of concepts
- Conclusions and future work

Motivations



Original Image 3 regions 5 regions 7 regions 10 regions 13 regions

Original Image 3 regions 5 regions 7 regions 10 regions 13 regions

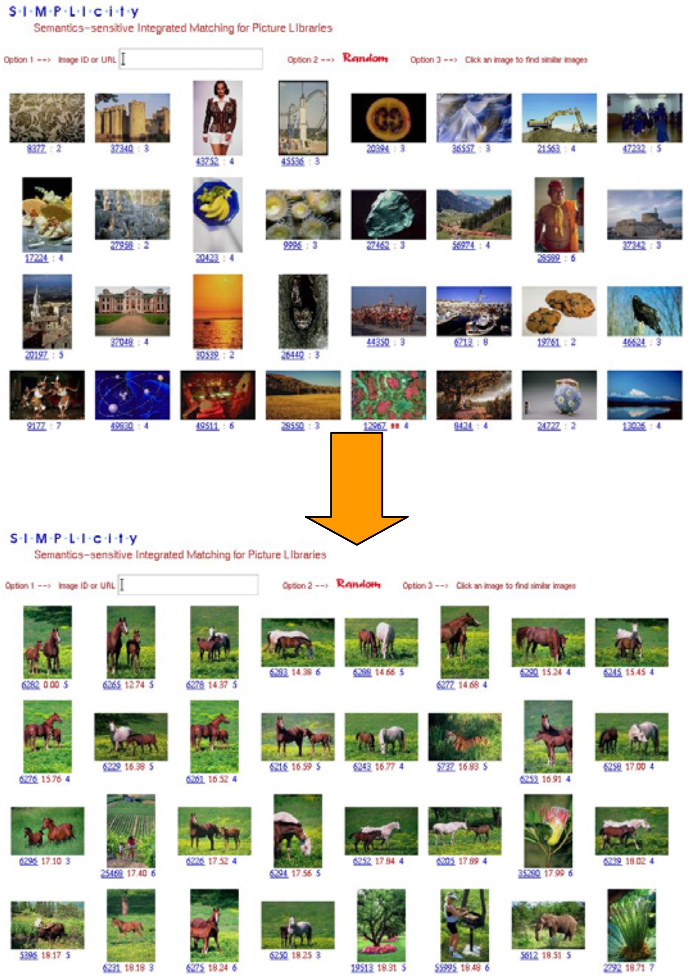
- Observations:
  - Human object segmentation relies on knowledge
  - Precise computer image segmentation is a very difficult open problem

- Hypothesis: It is possible to build robust computer matching algorithms without first segmenting the images accurately

Our SIMPLIcity Work

[PAMI, 2001(1)] [PAMI, 2001(9)][PAMI, 2002(9)]

- **S**emantics-sensitive **I**ntegrated **M**atching for **P**icture **LI**braries
- Major features
  - Sensitive to semantics: combine statistical semantic classification with image retrieval
  - Efficient processing: wavelet-based feature extraction
  - Reduced sensitivity to inaccurate segmentation and simple user interface: Integrated Region Matching (IRM)



S·I·M·P·L·I·c·i·t·y

Semantics-sensitive Integrated Matching for Picture LIbraries

Option 1 --> Image ID or URL

Option 2 --> Random

Option 3 --> Click an image to find similar images

S·I·M·P·L·I·c·i·t·y

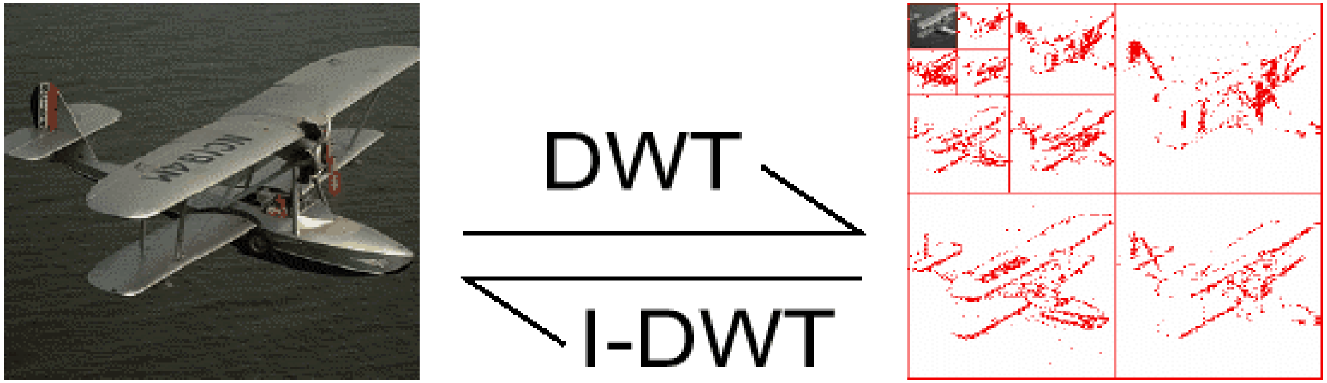
Semantics-sensitive Integrated Matching for Picture LIbraries

Option 1 --> Image ID or URL

Option 2 --> Random

Option 3 --> Click an image to find similar images

Wavelets



original image

3-level transform

- Wavelet: decomposes a 2-D image into trend (low-frequency) and fluctuation (high-frequency) bands in different scales
- Image applications: processing (denoising, enhancement), analysis/classification, compression
- Lossless inverse transform
- Daubechies' wavelets

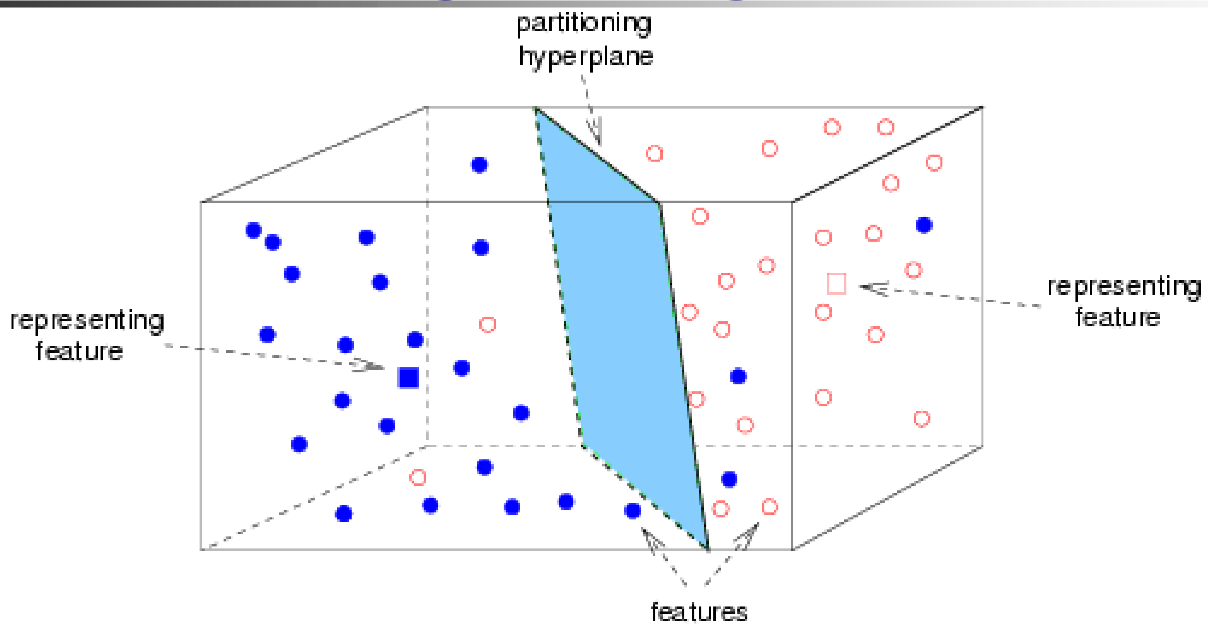
$$\phi_{r,j,k}(x) = 2^{j/2}\phi_r(2^j x - k),\ \ j,k \in \mathbb{Z}$$

$$< u, v >= \int_0^1 u(x)\overline{v(x)}dx$$

$$f_j(x) = \sum_k < f, \phi_{r,j,k} > \phi_{r,j,k}(x)$$

$$d_j(x) = f_{j+1}(x) - f_j(x)$$

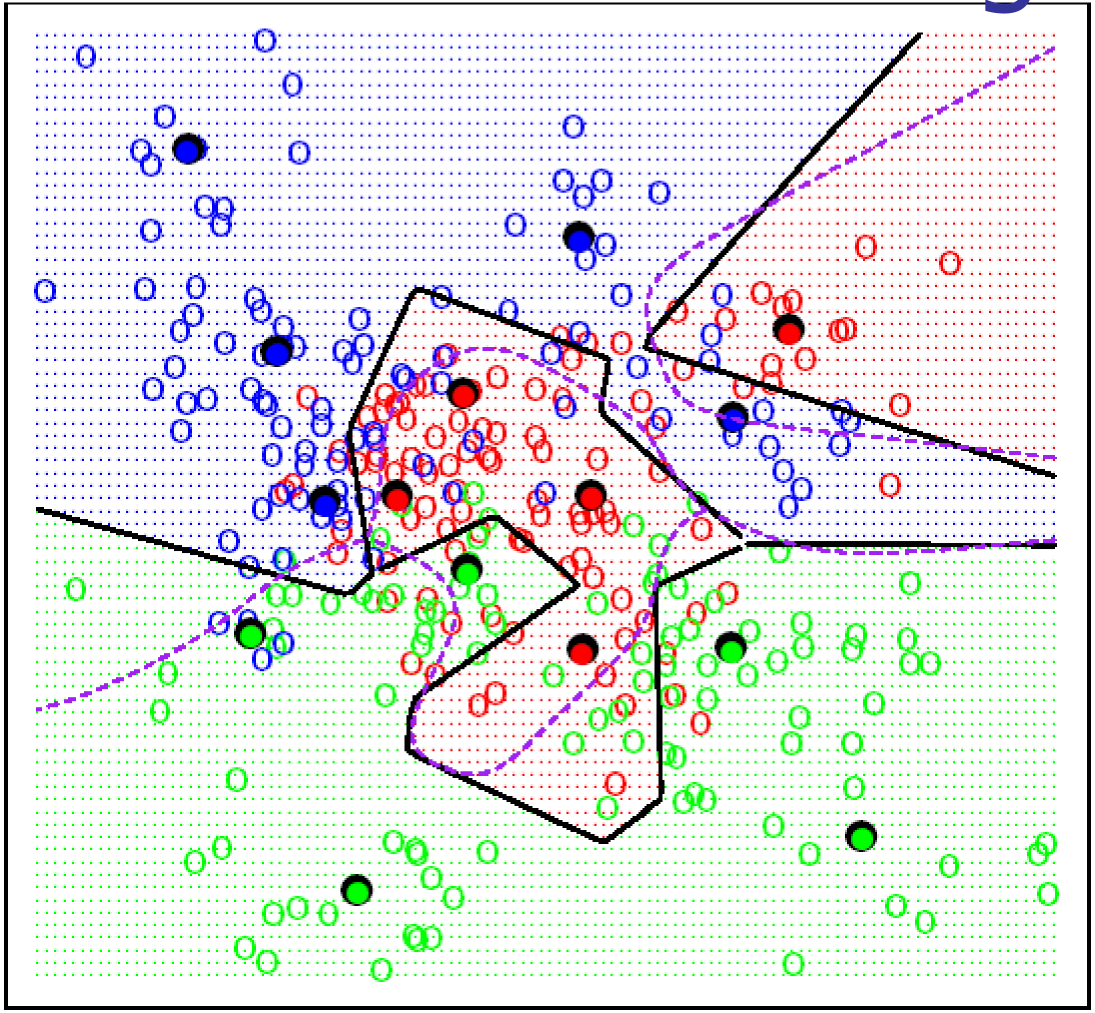
Fast Image Segmentation



- Partition an image into 4×4 blocks
- Extract wavelet-based features from each block
- Use *k*-means algorithm to cluster feature vectors into 'regions'
- Compute the shape feature by normalized inertia

K-means Statistical Clustering

- Some segmentation algorithms: 8 minute CPU time per image
- Our approach: use unsupervised statistical learning method to analyze the feature space
- Goal: minimize the mean squared error between the training samples and their representative prototypes
- Learning VQ



[Hastie+, Elements of Statistical Learning, 2001]

IRM: Integrated Region Matching

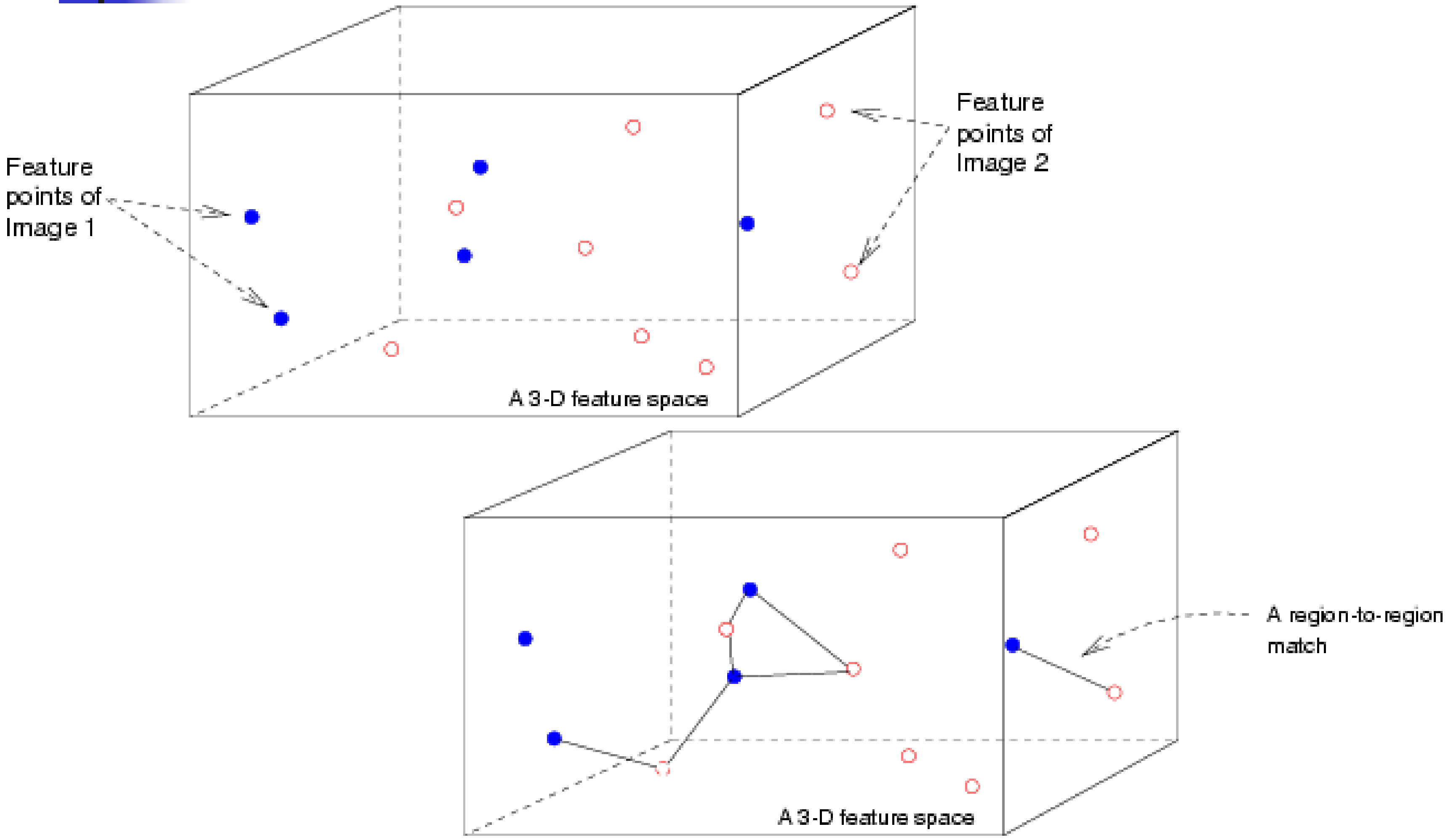
- IRM defines an image-to-image distance as a weighted sum of region-to-region distances

$$d_{IRM}(R_1, R_2) = \sum_{i,j} s_{i,j} d_{i,j}$$

- Weighting matrix is determined based on significance constrains and a ‘MSHP’ greedy algorithm

$$\begin{aligned} \sum_{j=1}^{n} s_{i,j} &= p_i, \; i = 1, ..., m \\ \sum_{i=1}^{m} s_{i,j} &= p_j', \; j = 1, ..., n \end{aligned}$$

A 3-D Example for IRM



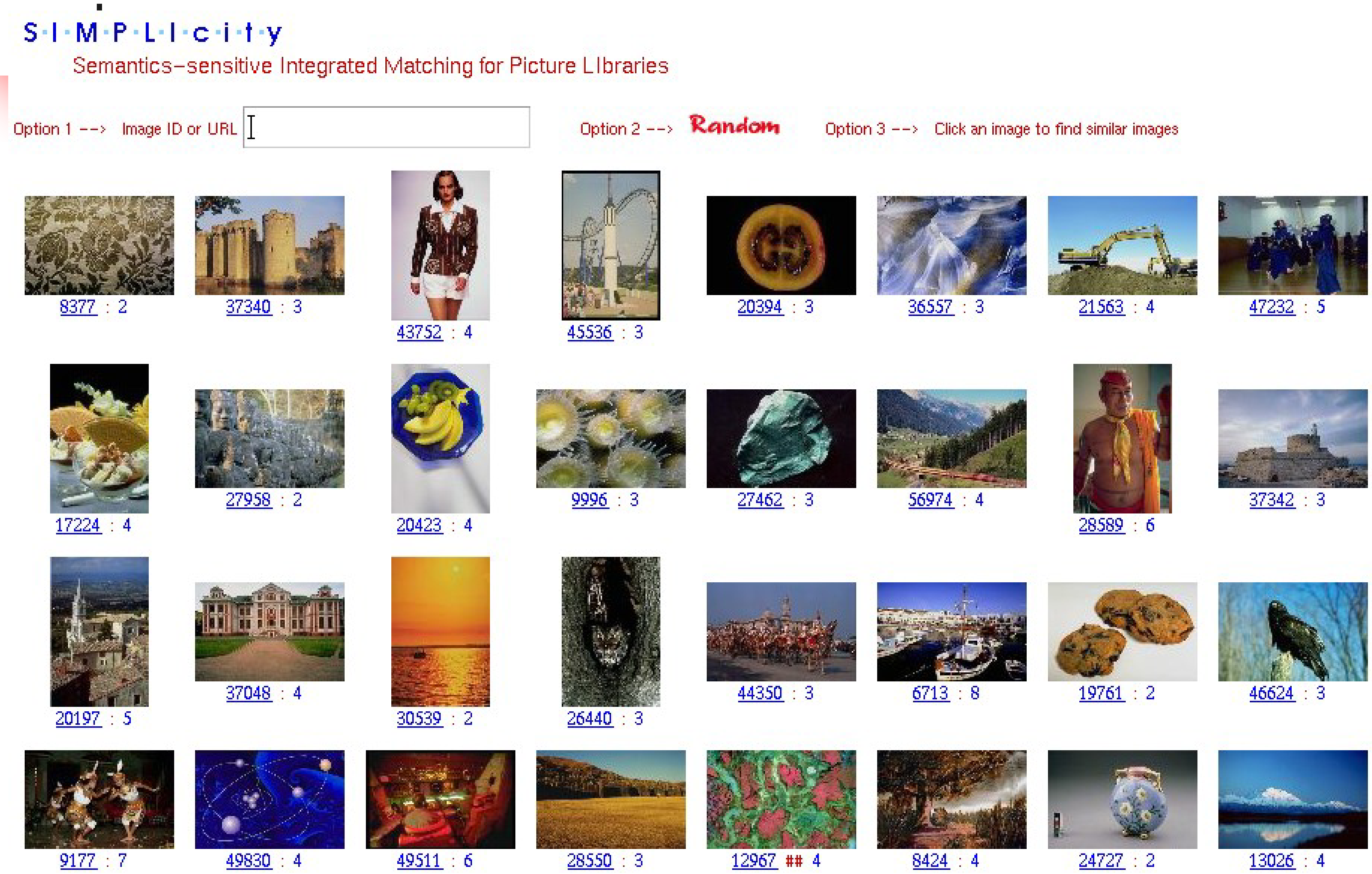
IRM: Major Advantages

1. Reduces the influence of inaccurate segmentation
2. Helps to clarify the semantics of a particular region given its neighbors
3. Provides the user with a simple interface

Experiments and Results

- Speed
  - 800 MHz Pentium PC with LINUX OS
  - Databases: 200,000 general-purpose image DB (60,000 photographs + 140,000 hand-drawn arts) 70,000 pathology image segments
  - Image indexing time: one second per image
  - Image retrieval time:
    - Without the scalable IRM, 1.5 seconds/query CPU time
    - With the scalable IRM, 0.15 second/query CPU time
  - External query: one extra second CPU time

RANDOM SELECTION



S·I·M·P·L·I·c·i·t·y

Semantics–sensitive Integrated Matching for Picture LIbraries

Option 1 --> Image ID or URL

Option 2 --> Random

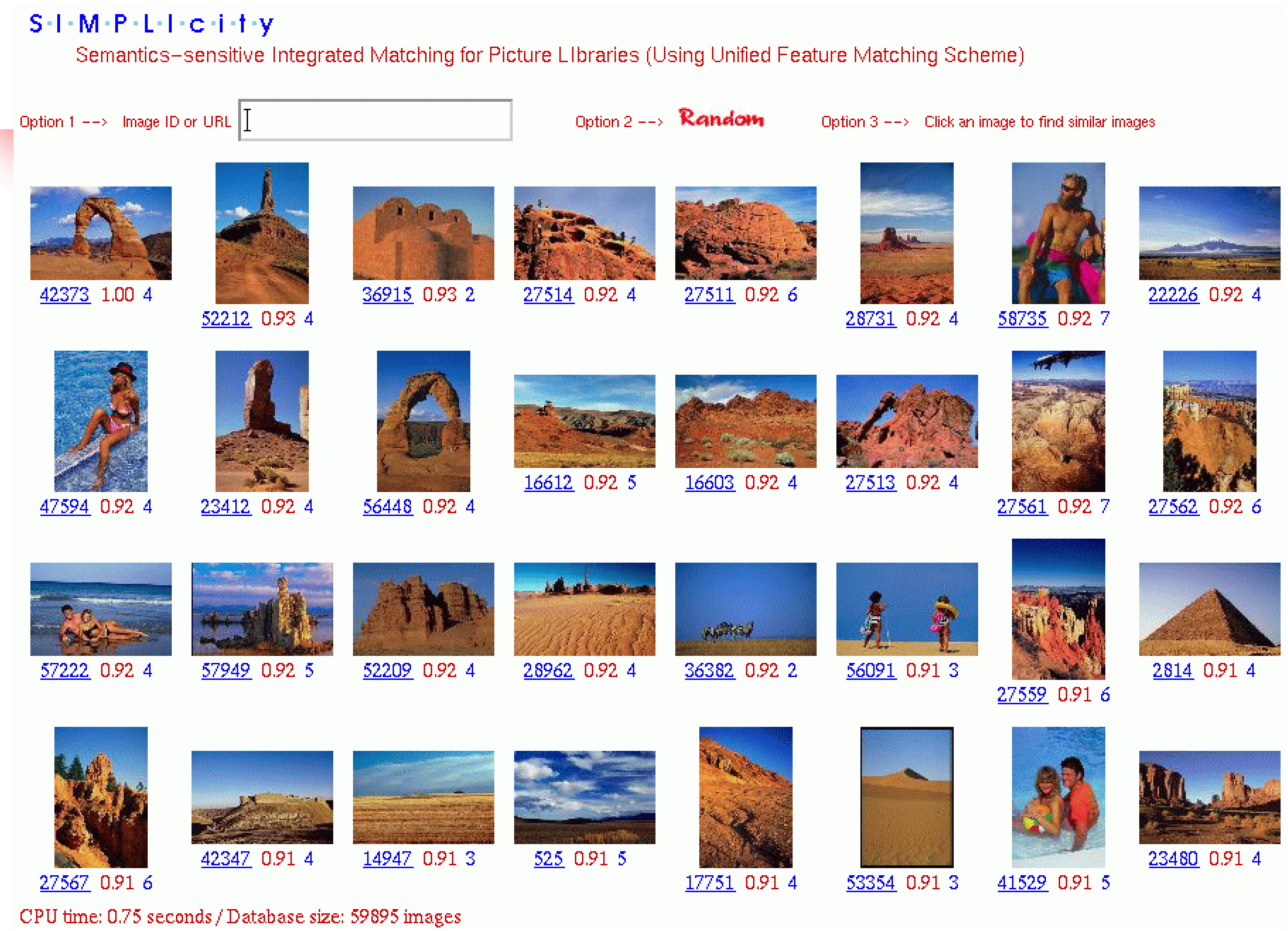
Option 3 --> Click an image to find similar images

8377 : 2
37340 : 3
43752 : 4
45536 : 3
20394 : 3
36557 : 3
21563 : 4
47232 : 5

17224 : 4
27958 : 2
20423 : 4
9996 : 3
27462 : 3
56974 : 4
28589 : 6
37342 : 3

20197 : 5
37048 : 4
30539 : 2
26440 : 3
44350 : 3
6713 : 8
19761 : 2
46624 : 3

9177 : 7
49830 : 4
49511 : 6
28550 : 3
12967 ## 4
8424 : 4
24727 : 2
13026 : 4



S·I·M·P·L·I·c·i·t·y

Semantics–sensitive Integrated Matching for Picture LIbraries (Using Unified Feature Matching Scheme)

Option 1 --> Image ID or URL

Option 2 --> Random

Option 3 --> Click an image to find similar images

42373 1.00 4
52212 0.93 4
36915 0.93 2
27514 0.92 4
27511 0.92 6
28731 0.92 4
58735 0.92 7
22226 0.92 4

47594 0.92 4
23412 0.92 4
56448 0.92 4
16612 0.92 5
16603 0.92 4
27513 0.92 4
27561 0.92 7
27562 0.92 6

57222 0.92 4
57949 0.92 5
52209 0.92 4
28962 0.92 4
36382 0.92 2
56091 0.91 3
27559 0.91 6
2814 0.91 4

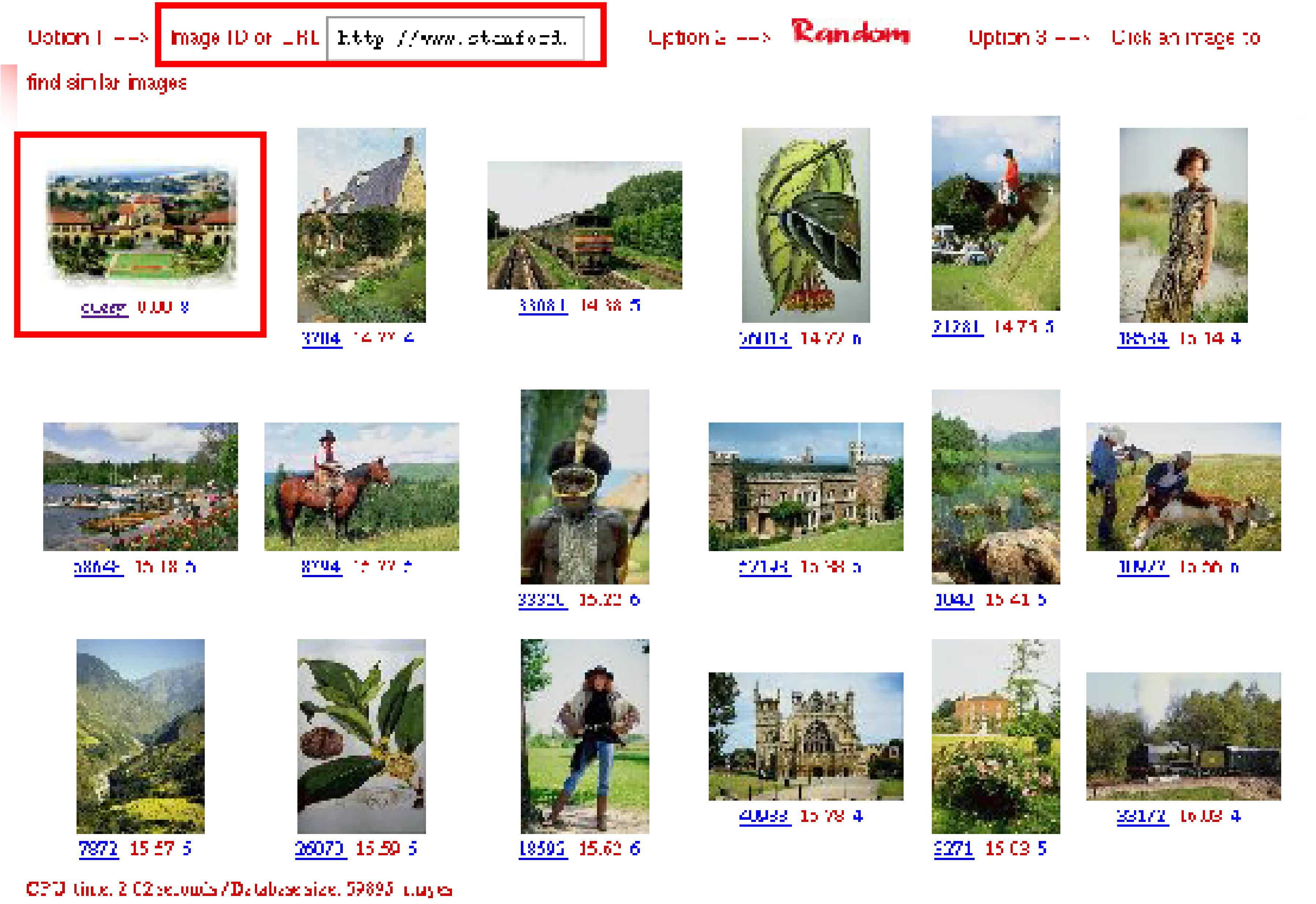
27567 0.91 6
42347 0.91 4
14947 0.91 3
525 0.91 5
17751 0.91 4
53354 0.91 3
41529 0.91 5
23480 0.91 4

CPU time: 0.75 seconds / Database size: 59895 images

External Query

S·I·M·P·L·I·c·i·t·y

Semantics-sensitive Integrated Matching for Picture Libraries



Option 1 --> Image ID or URL http://www.stanford.

Option 2 --> Random

Option 3 --> Click an image to find similar images

query 0.00 8

3704 14.99 4

33081 14.38 5

26013 14.77 6

21281 14.71 5

18534 15.14 4

38645 15.18 5

8794 15.20 5

33320 15.22 6

42143 15.38 5

3040 15.41 5

10922 15.46 6

7872 15.57 5

26073 15.59 5

18595 15.62 6

40088 15.78 4

9271 15.83 5

33172 16.08 4

CPU time: 2.02 seconds / Database size: 59895 images

Robustness to Image Alterations

- 10% brighten on average
- 8% darken
- Blurring with a 15x15 Gaussian filter
- 70% sharpen
- 20% more saturation
- 10% less saturation
- Shape distortions
- Cropping, shifting, rotation

Status of SIMPLIcity

- Researchers from more than 40 institutions/government agencies requested and obtained SIMPLIcity
- Where to find it -- do a google search of “image retrieval”
- We applied SIMPLicity to:
  - Automatic Web classification
  - Searching of pathological and biomedical images
  - Searching of art and cultural images

Outline

- Introduction
- Our related SIMPLIcity work
- ALIP: Automatic modeling and learning of concepts
- Conclusions and future work

Why ALIP?

- Size
  - 1 million images
- Understandability & Vision
  - "meaning" depend on the *point-of-view*
  - Can we translate contents and structure into linguistic terms

dogs

Kyoto

(cont.)

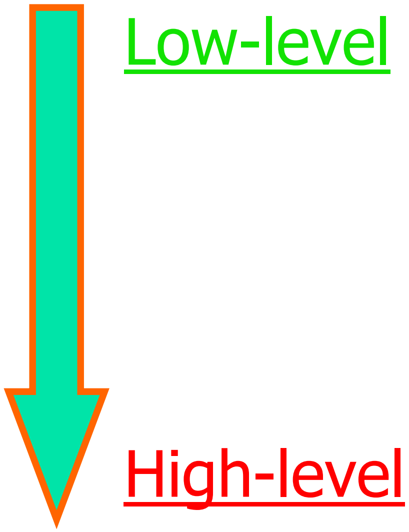
- Query formulation
  - SIMILARITY: look similar to a given picture
  - OBJECT: contains an explosive device
  - OBJECT RELATIONSHIP: contains a weapon and a person; find all nuclear facilities from a satellite picture
  - MOOD: a sad picture
  - TIME/PLACE: sunset near the Capital

Automatic Modeling and Learning of Concepts for Image Indexing

- Observations:
  - Human beings are able to build models about objects or concepts by mining visual scenes
  - The learned models are stored in the brain and used in the recognition process
- Hypothesis: It is achievable for computers to mine and learn a large collection of concepts by 2D or 3D image-based training
- [Wang+Li, ACM Multimedia, 2002][PAMI revision]

Concepts to be Trained

- **Concepts**: Basic building blocks in determining the semantic meanings of images
- Training concepts can be categorized as:
    - **Basic Object**: flower, beach
    - **Object composition**: building+grass+sky+tree
    - **Location**: Asia, Venice
    - **Time**: night sky, winter frost
    - **Abstract**: sports, sadness

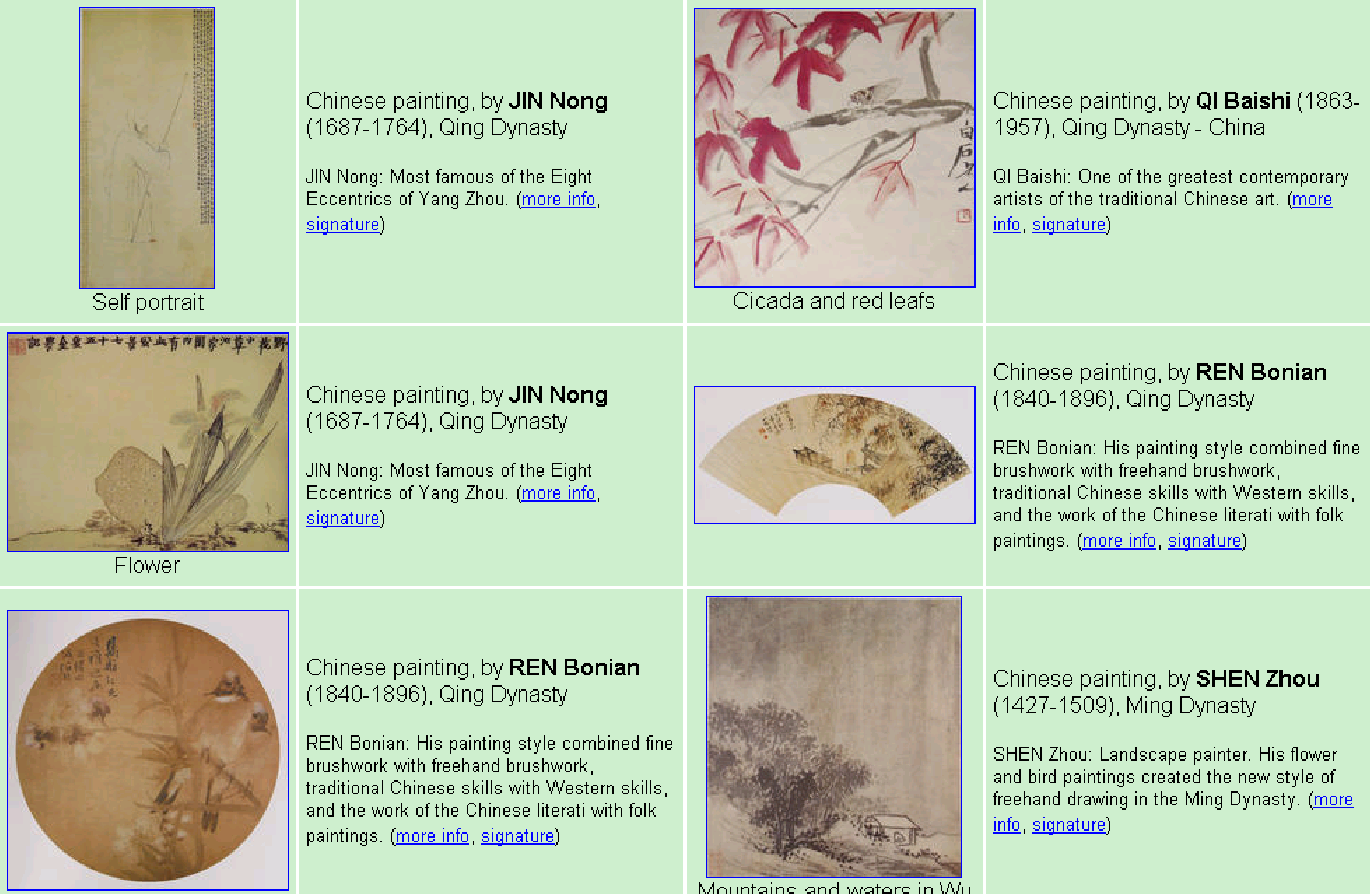


Low-level

High-level

Modeling of Artist's Handwriting (NSF ITR)

- Each artist has consistent as well as unique strokes, equivalent of a signature
  - Rembrandt: swift, accurate brush
  - Degas: deft line, controlled scribble
  - Van Gogh: turbulent, swirling strokes, rich of textures
  - Asian painting arts (focus of ITR, started 8/2002)
- Potential queries
  - Find paintings with brush strokes similar to those of van Gogh's
  - Find paintings with similar artist intentions

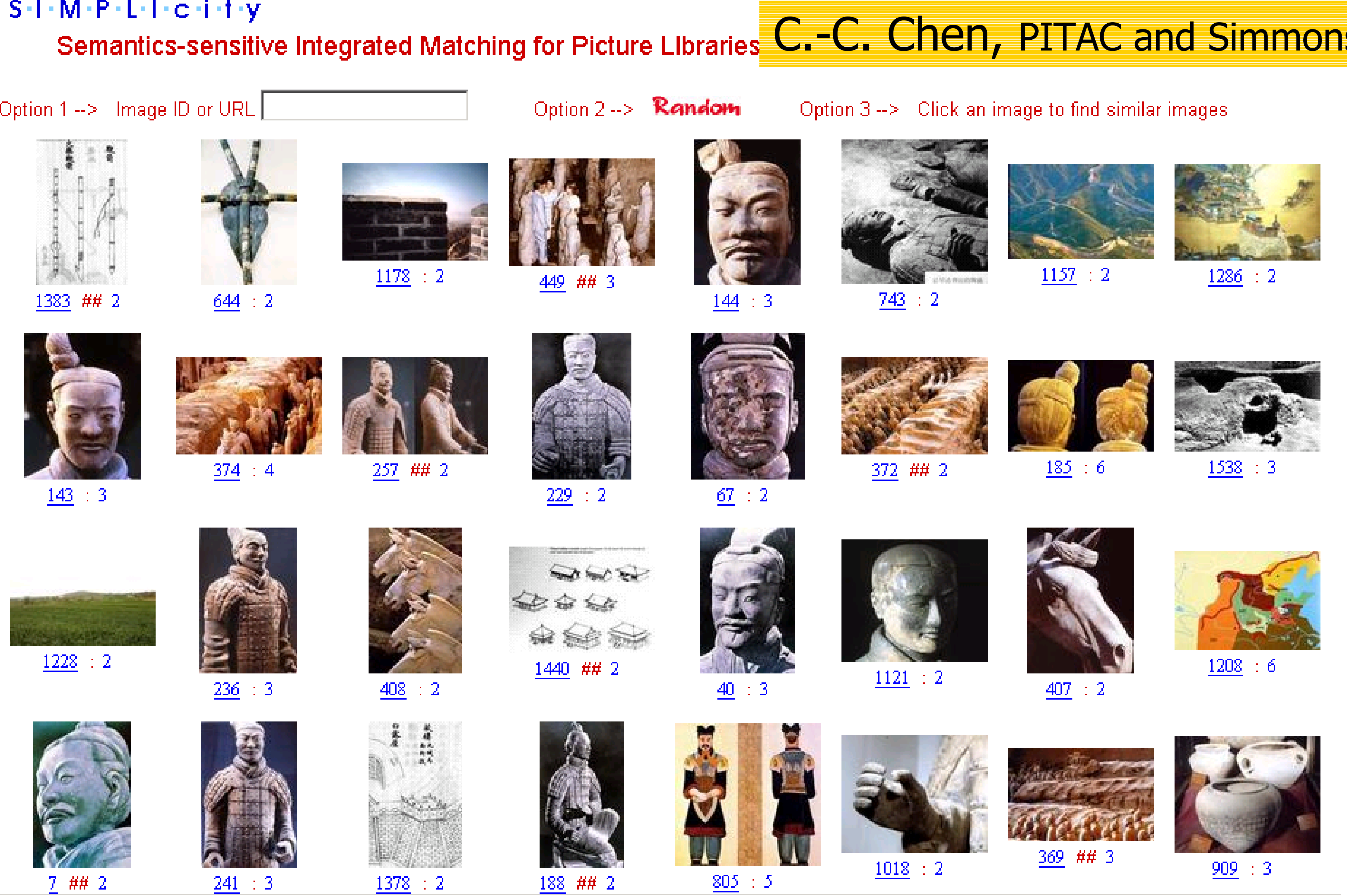


Self portrait	Chinese painting, by **JIN Nong** (1687-1764), Qing Dynasty JIN Nong: Most famous of the Eight Eccentrics of Yang Zhou. (more info, signature)	Cicada and red leafs	Chinese painting, by **QI Baishi** (1863-1957), Qing Dynasty - China QI Baishi: One of the greatest contemporary artists of the traditional Chinese art. (more info, signature)
Flower	Chinese painting, by **JIN Nong** (1687-1764), Qing Dynasty JIN Nong: Most famous of the Eight Eccentrics of Yang Zhou. (more info, signature)		Chinese painting, by **REN Bonian** (1840-1896), Qing Dynasty REN Bonian: His painting style combined fine brushwork with freehand brushwork, traditional Chinese skills with Western skills, and the work of the Chinese literati with folk paintings. (more info, signature)
	Chinese painting, by **REN Bonian** (1840-1896), Qing Dynasty REN Bonian: His painting style combined fine brushwork with freehand brushwork, traditional Chinese skills with Western skills, and the work of the Chinese literati with folk paintings. (more info, signature)	Mountains and waters in Wu	Chinese painting, by **SHEN Zhou** (1427-1509), Ming Dynasty SHEN Zhou: Landscape painter. His flower and bird paintings created the new style of freehand drawing in the Ming Dynasty. (more info, signature)

Database: most significant Asian paintings

Question: can we build a “dictionary” of different painting styles?

C.-C. Chen, PITAC and Simmons



SIMPLIcity

Semantics-sensitive Integrated Matching for Picture LIbraries

Option 1 --> Image ID or URL

Option 2 --> Random

Option 3 --> Click an image to find similar images

1383 ## 2 | 644 : 2 | 1178 : 2 | 449 ## 3 | 144 : 3 | 743 : 2 | 1157 : 2 | 1286 : 2

143 : 3 | 374 : 4 | 257 ## 2 | 229 : 2 | 67 : 2 | 372 ## 2 | 185 : 6 | 1538 : 3

1228 : 2 | 236 : 3 | 408 : 2 | 1440 ## 2 | 40 : 3 | 1121 : 2 | 407 : 2 | 1208 : 6

7 ## 2 | 241 : 3 | 1378 : 2 | 188 ## 2 | 805 : 5 | 1018 : 2 | 369 ## 3 | 909 : 3

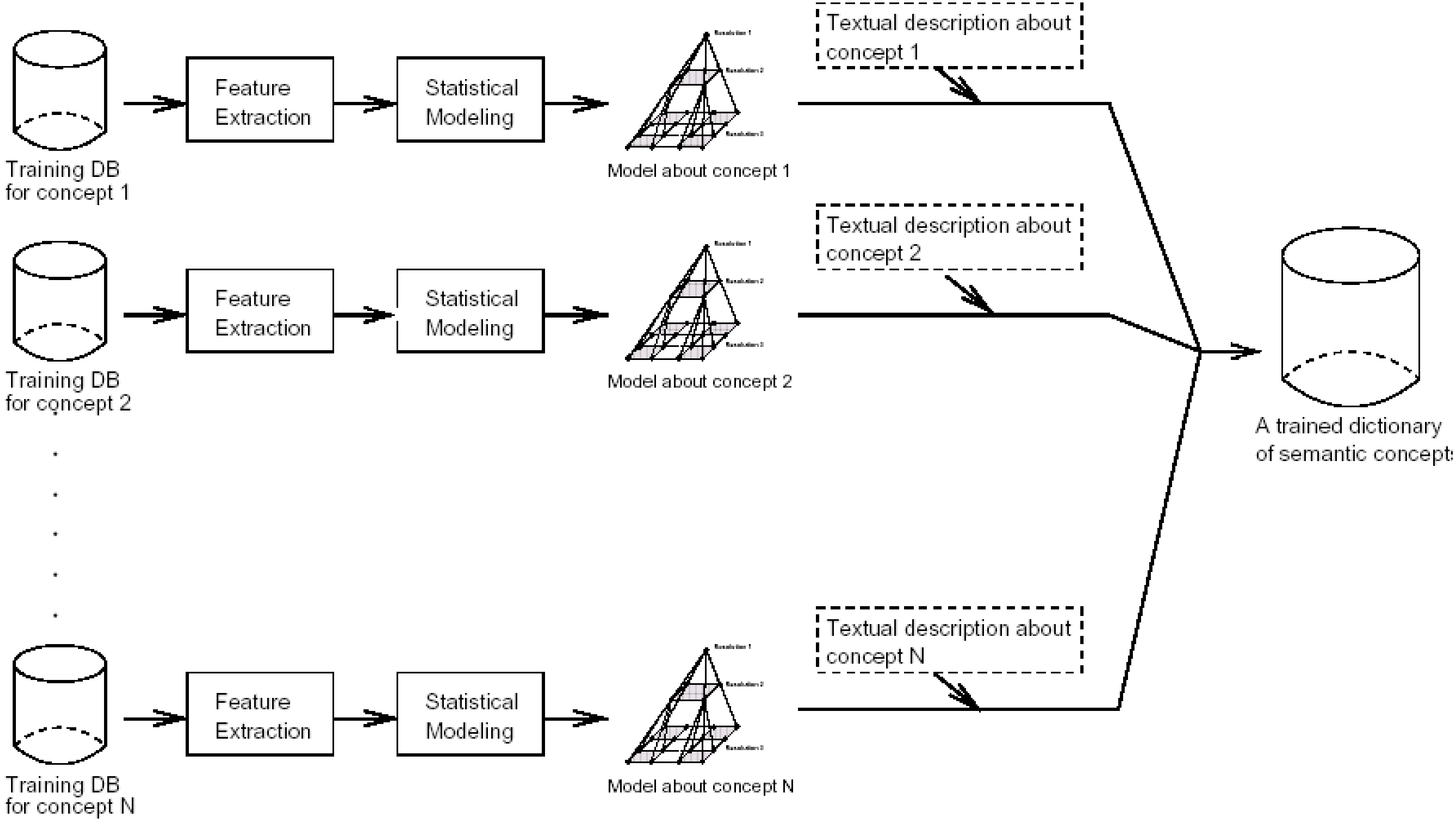
Database: terracotta soldiers of the First Emperor of China

Question: can we train the computer to be an art historian?

System Design

- Train statistical models of a dictionary of concepts using sets of training images
  - 2D images are currently used
  - 3D-image training can be much better
- Compare images based on model comparison
- Select the most statistical significant concept(s) to index images linguistically
- Initial experiment:
  - 600 concepts, each trained with 40 images
  - 15 minutes Pentium CPU time per concept, train only once
  - highly parallelizable algorithm

Training Process



Training DB for concept 1

Feature Extraction

Statistical Modeling

Model about concept 1

Textual description about concept 1

Training DB for concept 2

Feature Extraction

Statistical Modeling

Model about concept 2

Textual description about concept 2

Training DB for concept N

Feature Extraction

Statistical Modeling

Model about concept N

Textual description about concept N

A trained dictionary of semantic concepts

Automatic Annotation Process

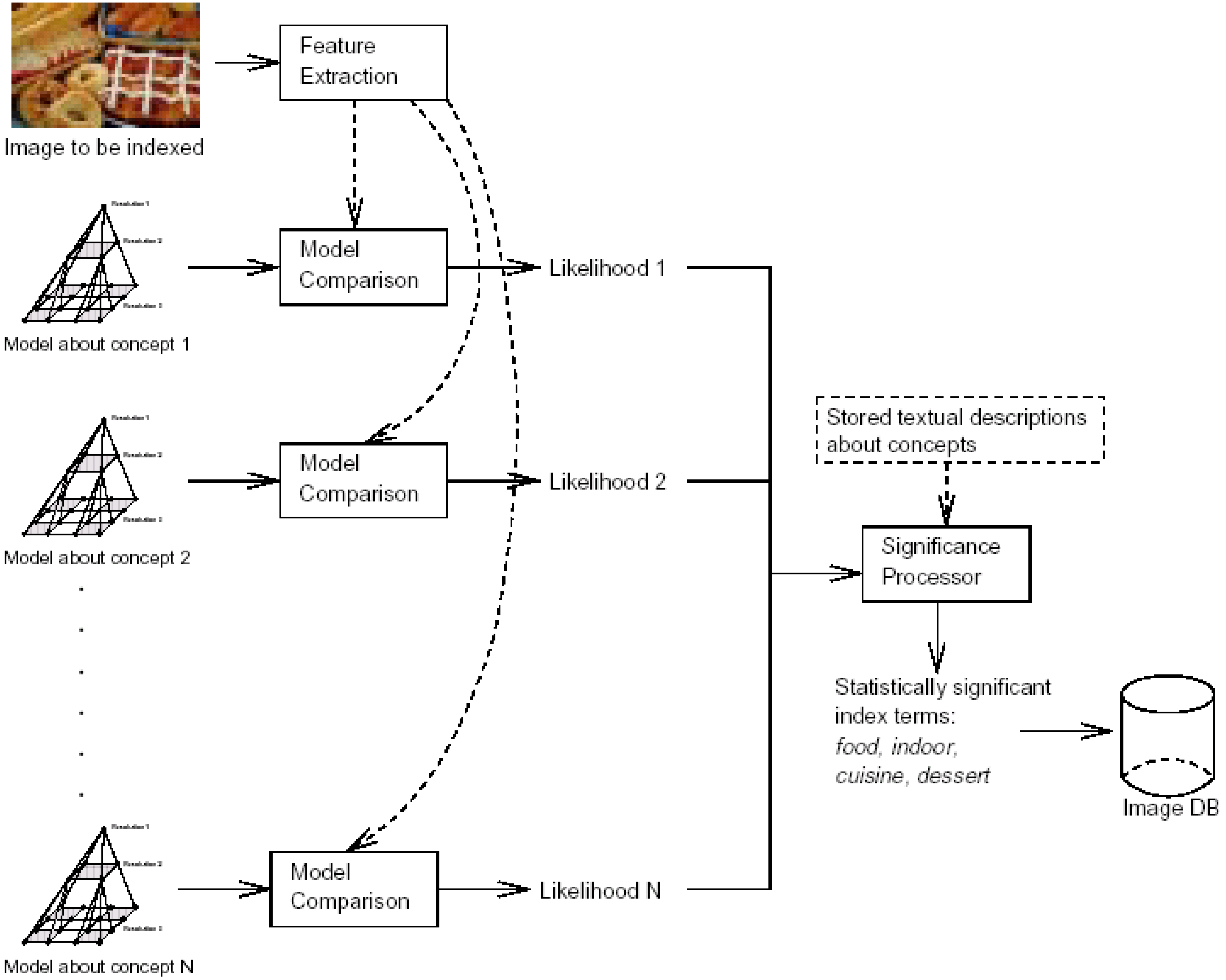


Image to be indexed

Feature Extraction

Model about concept 1

Model Comparison

Likelihood 1

Model about concept 2

Model Comparison

Likelihood 2

Model about concept N

Model Comparison

Likelihood N

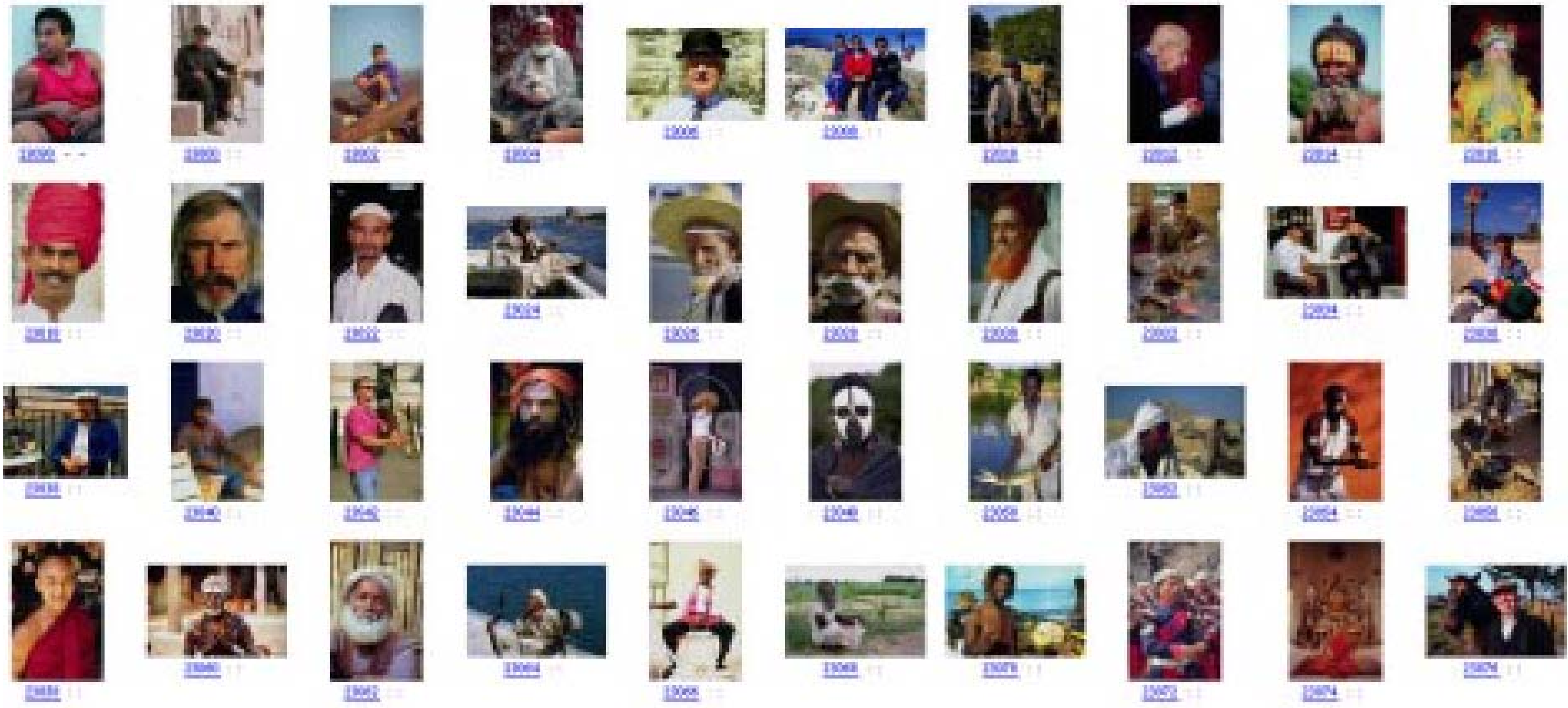
Stored textual descriptions about concepts

Significance Processor

Statistically significant index terms: *food, indoor, cuisine, dessert*

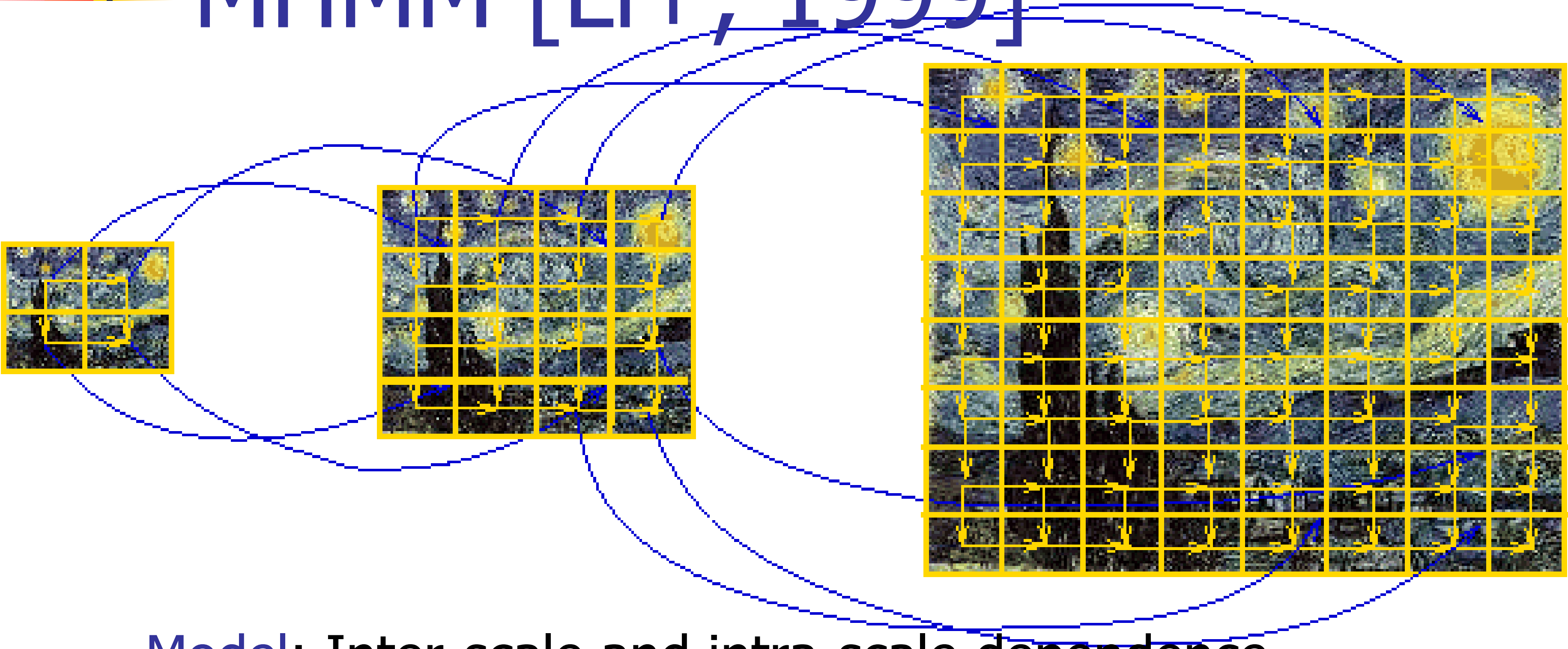
Image DB

Training



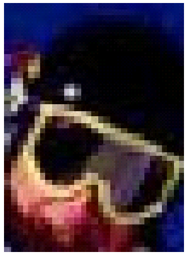
Training images used to train the concept "male" with description "man, male, people, cloth, face"

Initial Model: 2-D Wavelet MHMM [Li+, 1999]



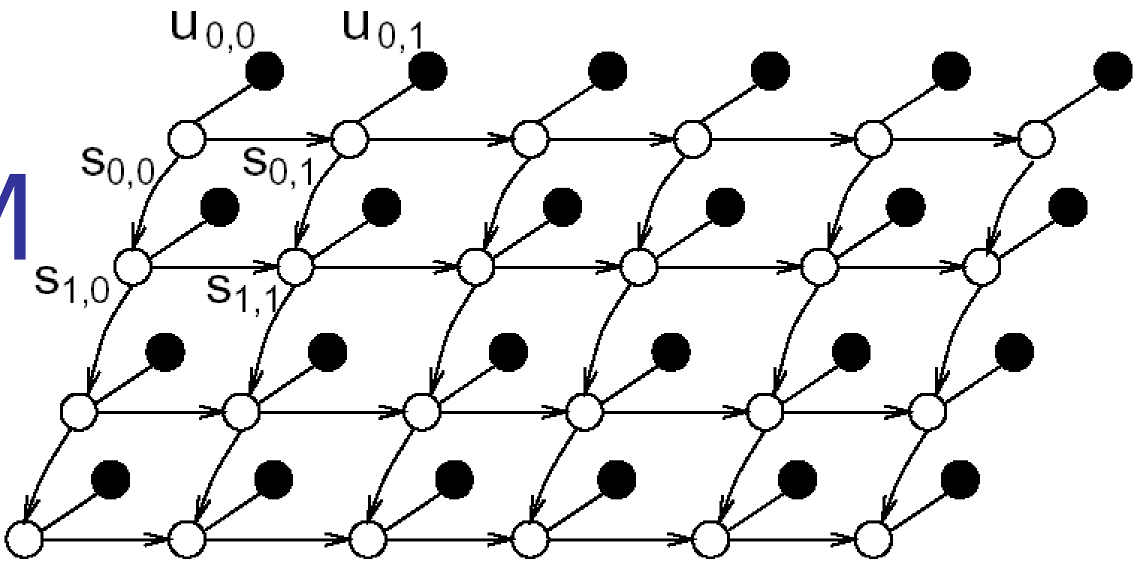
- Model: Inter-scale and intra-scale dependence
- States: hierarchical Markov mesh, unobservable
- Features in SIMPLIcity: multivariate Gaussian distributed given states
- A model is a knowledge base for a concept

2D MHMM



a region of an image

the whole image

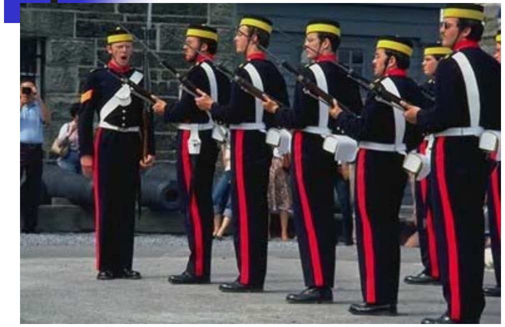


u | s: $N(\mu, \Sigma)$

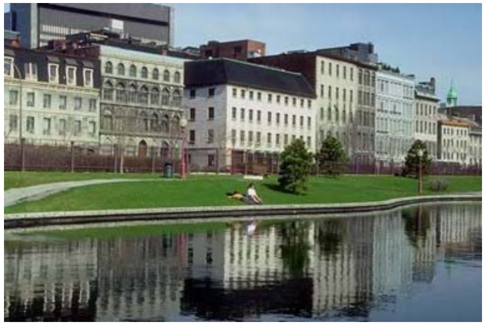
$$
\begin{aligned}
& P\{s_{i,j}^{(r)}, u_{i,j}^{(r)}; r \in \mathcal{R}, (i,j) \in \mathbb{N}^{(r)}\} \\
= \ & P\{s_{i,j}^{(1)}, u_{i,j}^{(1)}; (i,j) \in \mathbb{N}^{(1)}\} \cdot P\{s_{i,j}^{(2)}, u_{i,j}^{(2)}; (i,j) \in \mathbb{N}^{(2)} \,|\, s_{k,l}^{(1)}; (k,l) \in \mathbb{N}^{(1)}\} \cdots \\
& P\{s_{i,j}^{(R)}, u_{i,j}^{(R)}; (i,j) \in \mathbb{N}^{(R)} \,|\, s_{k,l}^{(R-1)}; (k,l) \in \mathbb{N}^{(R-1)}\}
\end{aligned}
$$

- Start from the conventional 1-D HMM
- Extend to 2D transitions
- Conditional Gaussian distributed feature vectors
- Then add Markovian statistical dependence across resolutions
- Use EM algorithm to estimate parameters

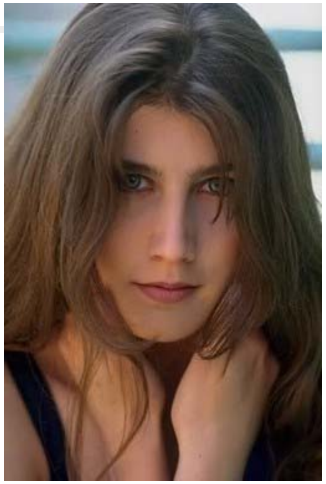
Preliminary Results



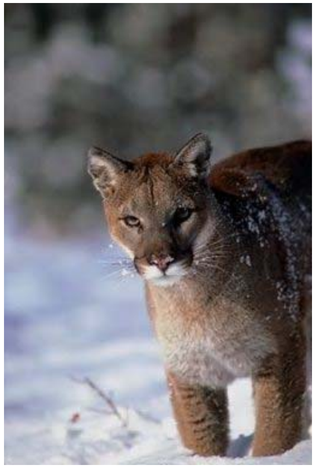
Computer Prediction:
people, Europe, man-made, water



Building, sky, lake, landscape, Europe, tree

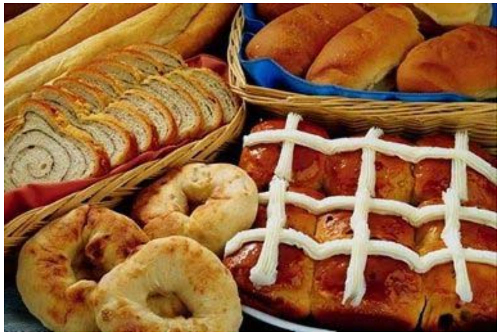


People, Europe, female



Snow, animal, wildlife, sky, cloth, ice, people

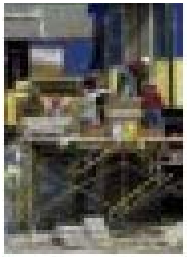
Food, indoor, cuisine, dessert



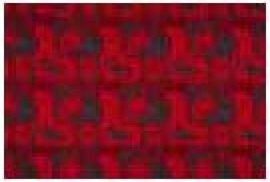
More Results

skyline, sky, New York, landmark

plant,flower, garden



modern,parade, people

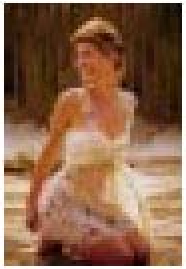


pattern,flower, red,dining

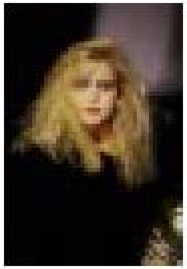
ocean,paradise, San Diego, Thailand, beach,fish

flower,flora, plant,fruit, natural,texture

ancestor, drawing, fitness, history, indoor

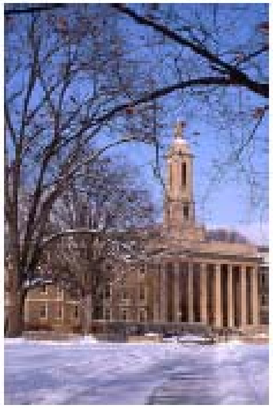


hair style, occupation,face, female,cloth

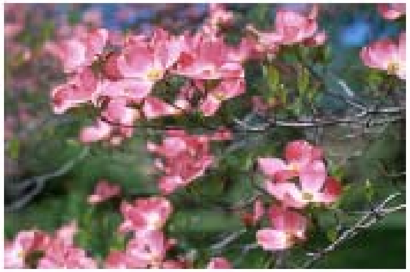


night,cyber, fashion,female

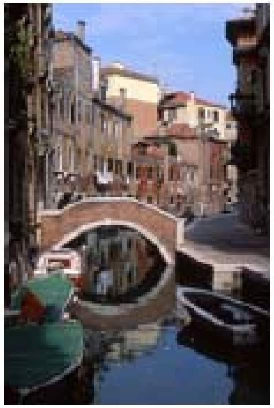
Results: using our own photographs



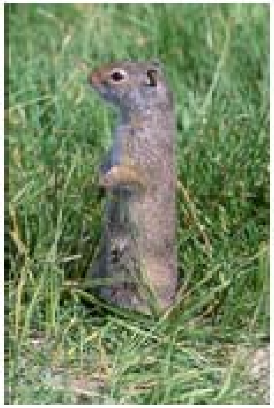
P: building, snow, sky, tree, landscape



P: flower



P: historical building, (bridge), river, Italy, sky, Europe



P: animal, grass (squirrel)

- **P**: Photographer annotation
- Underlined words: words predicted by computer
- (Parenthesis): words not in the learned "dictionary" of the computer

Advantages of Our Approach

- Incremental mining and learning
- Highly scalable (unlike CART, SVM, ANN)
- Flexible: Amount of training depends on the complexity of the concept
- Context-dependent: Spatial relations among pixels taken into consideration
- Universal image similarity: statistical likelihood rather than relying on segmentation

Outline

- Introduction
- Our related SIMPLIcity work
- ALIP: Automatic modeling and learning of concepts
- Conclusions and future work

Conclusions

- We propose a new research direction:
    - **A**utomatic **L**inguistic **I**ndexing of **P**ictures
    - Highly challenging but crucially important
    - Interdisciplinary collaboration is critical
- Our SIMPLIcity image indexing system
- Our ALIP System: Automatic modeling and learning of semantic concepts
    - 600 concepts can be learned automatically

Future Work

- Explore new methods for better accuracy
  - refine statistical modeling of images
  - learning from 3D
  - refine matching schemes
- Apply these methods to
  - special image databases
    (e.g., art, biomedicine, intelligence)
  - very large databases
- Integration of ALIP with large-scale information mining systems
- ……

Acknowledgments

- NSF ITR (since 08/2002)
- Endowed professorship from the PNC Foundation
- Equipment grant from SUN Microsystems
- Penn State Univ.

- Earlier funding (1995-2000): IBM QBIC, NEC AMORA, SRI AI, Stanford Lib/Math/Biomedical Informatics/CS, Lockheed Martin, NSF DL2

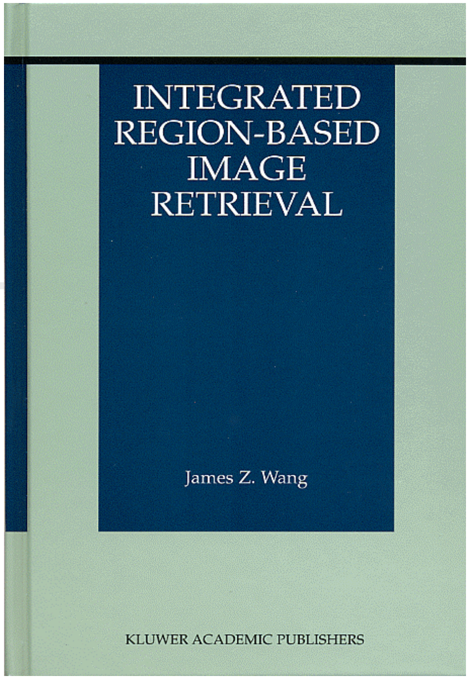
More Information

Papers in PDF,
image databases, downloads,
demo, etc

CALL FOR PAPER:
WWW2003-Multimedia Track, 11/15

http://wang.ist.psu.edu

PENNSTATE



INTEGRATED
REGION-BASED
IMAGE
RETRIEVAL

James Z. Wang

KLUWER ACADEMIC PUBLISHERS

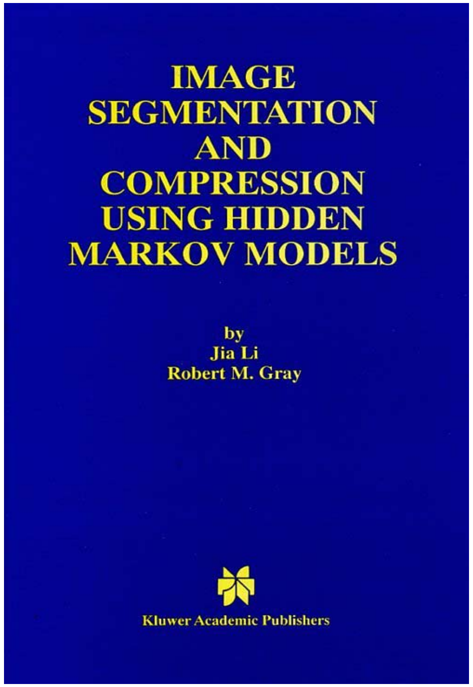


IMAGE
SEGMENTATION
AND
COMPRESSION
USING HIDDEN
MARKOV MODELS

by
Jia Li
Robert M. Gray

Kluwer Academic Publishers