From Pixels to Semantics

Research on automatic indexing and retrieval of large collections of images

James Z. Wang
PNC Technologies Career Development Professorship
School of Information Sciences and Technology
Penn State University

http://wang.ist.psu.edu

Outline

- Introduction
- Our SIMPLIcity work
- Experiments and results
- Our recent ongoing research
  - Automatic modeling and learning of concepts
- Conclusions and future work

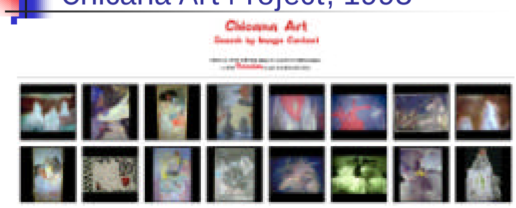
Research: Main Areas

- Multimedia information retrieval
  - Image retrieval
  - Image classification
- Biomedical informatics
  - DNA/Protein sequence analysis
  - Biodiversity informatics
- Information security
  - Content-based security filtering

Brief History of Our Work

- 1995: Stanford Art Library funded our project on searching of a Chicana Art image database
- 1996-2000: our work was funded by
  - IBM QBIC group (used by SF Museum of F.A.)
  - NEC AMORA group (used by the Getty Museum)
  - SRI (DARPA-funded, stereo matching for battlefield images)
- 1999-2000: NSF DLI2 project, Stanford
  - Use wavelets, statistical classification, and integrated region-based approach for image retrieval; image security filtering
- 2000-now: Penn State University
  - Automatic modeling and learning of concepts for image indexing

Chicana Art Project, 1995



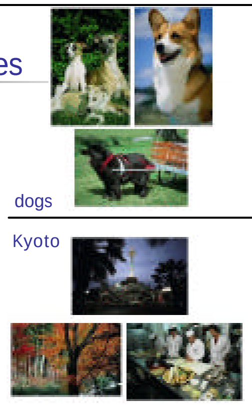
- 1000+ high quality paintings
- Goal: help students and researchers to find similar paintings
- Used wavelet-based features [Wang+,1997]

Introduction: The Problem

- Content-based Image Retrieval
  - The retrieval of relevant images from an image database on the basis of automatically -derived image features
- Applications
  - Biomedicine (X-ray, pathology, CT, MRI, … …)
  - Government (radar, aerial, trademark, … …)
  - Commercial (fashion catalogue, journalism, … …)
  - Cultural (museums, art galleries, … …)
  - Education and training
  - Entertainment, WWW, … …

Major Challenges

- Size
  - 1 million images
    → 1000 GB of space
    → 30 GB compressed
- Understandability & Vision
  - "meaning" depend on the *point-of-view*
  - Hard to translate contents and structure into linguistic terms



(cont.)

- Query formulation
  - SIMILARITY: look similar
  - OBJECT: contains a bike
  - OBJECT RELATIONSHIP: contains a dog near a person
  - MOOD: a happy picture
  - TIME/PLACE: Yosemite sunset

Related Work

- Many image search engines
  - IBM, VIRAGE, NEC, Interpix, scour.net,......
  - MIT, Stanford, Berkeley, Columbia, CMU, UCSB,......
- Speed: None is capable of handling the images on the Web
- Accuracy: None is near the human level of accuracy
- An active research area

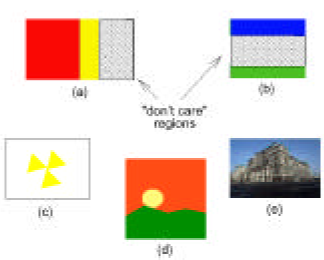
Text-based Approach

- Index images using keywords or descriptions (e.g., google.com)

\+ Easier to design and implement, fast execution
\+ Surrounding text in the Web page
\+ Accepted approach for high value pictures

-- Often too expensive
-- A picture is worth, and can require 1000 words
-- Query word may NOT appear as a keyword
-- Surrounding text may NOT describe the image

Feature-based Approach



\+ Handles low-level semantic queries
\+ Many features can be extracted

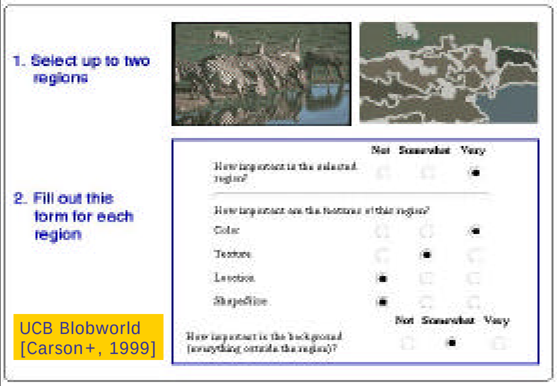
-- Cannot handle higher-level queries (e.g.,objects)

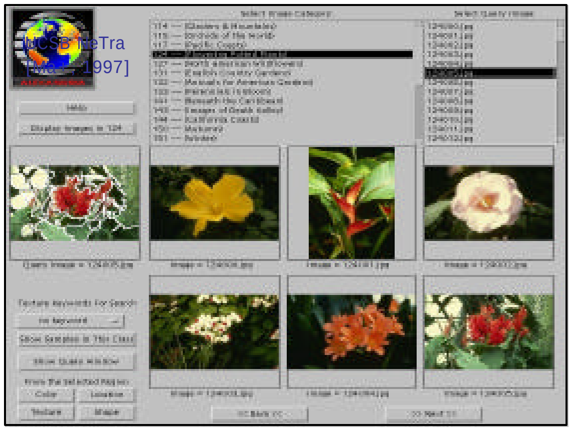
Region-based Approach

- Extract objects from images first

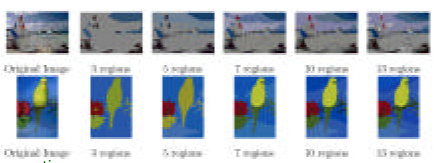
\+ Handles object-based queries
e.g., find images with objects that are similar to some given objects
\+ Reduce feature storage adaptively

-- Object segmentation is very difficult
-- User interface: region marking, feature combination





Motivations



- 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

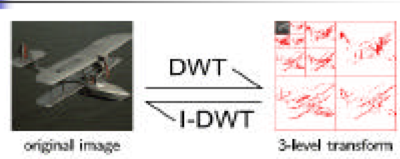
Outline

- Introduction
- Our SIMPLIcity work
- Experiments and results
- Our recent ongoing research
- Conclusions and future work

Our SIMPLIcity Work
[Wang+, D-LIB, 1999][Wang+, PAMI, 2001]

- **SIMPLIcity** system
  - **S**emantics-sensitive **I**ntegrated **M**atching for **P**icture **LI**braries
  - Combine low-level statistical semantic classification with image retrieval
  - Wavelet-based feature extraction for fast segmentation
  - Integrated Region Matching (IRM)

Wavelets



- 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

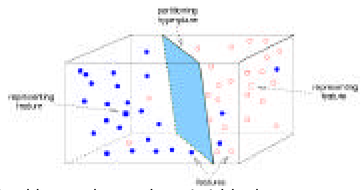
$\phi_{j,k}(x) = 2^{j/2}\phi_j(2^j x - k),\ j, k \in \mathbb{Z}$

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

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

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

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

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

$$\sum_{j=1}^{n} s_{i,j} = p_i, \; i = 1, \ldots, m$$
$$\sum_{i=1}^{m} s_{i,j} = p'_j, \; j = 1, \ldots, n$$

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

Recent Extensions

- Scalable IRM:
  - Indexing region-based feature space using statistical clustering
  - [Wang+Du, JCDL, 2001]
- Fuzzy matching:
  - Fuzzy region matching to further reduce sensitivity to the average number of regions segmented
  - [Chen+Wang, PAMI, 2002]

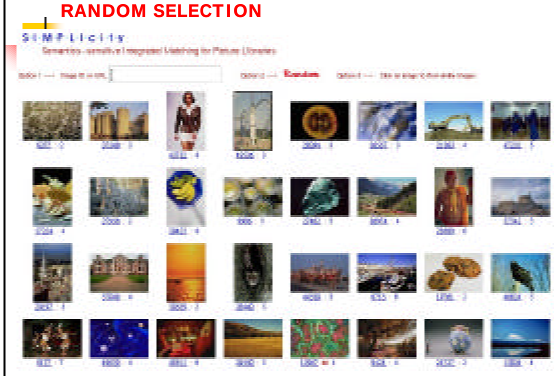
Outline

- Introduction
- Our SIMPLIcity work
- Experiments and results
- Our recent ongoing research
- Conclusions and future work

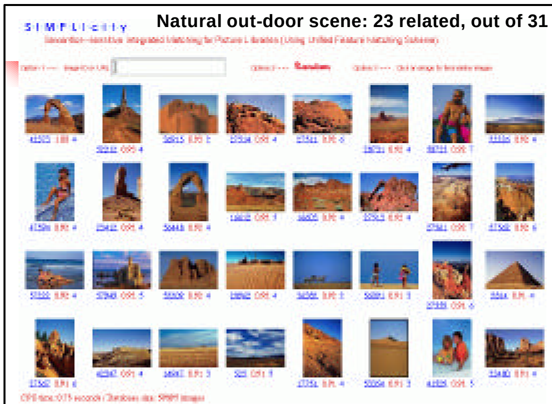
Experiments and Results

- Speed
  - 800 MHz Pentium PC with LINUX OS
  - Database: 200,000 COREL image DB (60,000 photographs + 140,000 hand-drawn arts)
  - 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

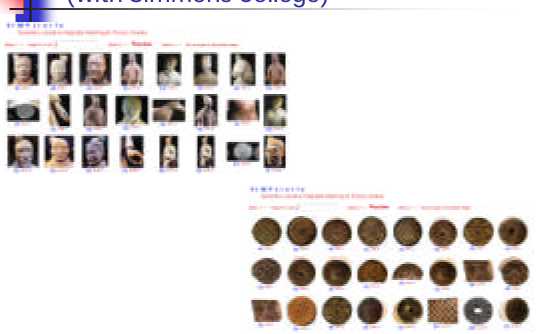


Natural out-door scene: 23 related, out of 31



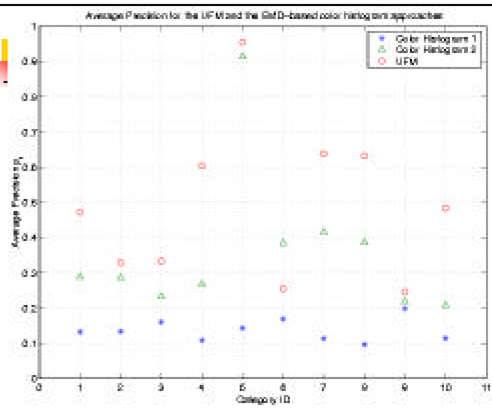
External Query

Searching Terracotta Warriors (with Simmons College)



Objective Accuracy Test on Image Categorization

- Compare with EMD [Rubner+, 1999]
  - Two setups: avg of 13.1 filled bins, 42.6 filled bins
- Subset of the COREL database
  - 10 categories, each containing 100 pictures
  - Africa, breach, buildings, buses, dinosaurs, elephants, flowers, horses, mountains, food
  - 1000 queries were tested
  - Average precision p within the best 100 matches is computed



4/12/2002 J. Z. Wang, Penn State University 31

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

4/12/2002 J. Z. Wang, Penn State University 32

Outline

- Introduction
- Our SIMPLIcity work
- Experiments and results
- Our recent ongoing research
- Conclusions and future work

4/12/2002 J. Z. Wang, Penn State University 33

Automatic Modeling and Learning of Concepts for Image Indexing

- Key observations:
  - Human beings are able to build models about objects or concepts from images
  - The learned models are stored in the brain and used in the recognition process
- Hypothesis: Computers can learn a large collection of concepts by 2D or 3D image-based training
- Trained Concepts: Basic building blocks in determining the semantic meanings of images
  - Basic Object: flower, beach
  - Object composition: building+grass+sky+tree
  - Location: Asia, Venice
  - Time: night sky, winter frost
  - Abstract: sports, sadness

4/12/2002 J. Z. Wang, Penn State University 34

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

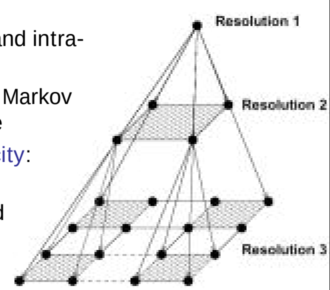
4/12/2002 J. Z. Wang, Penn State University 35

Initial Model: 2-D Wavelet MHMM

- Model: Inter-scale and intra-scale dependence
- States: hierarchical Markov mesh, unobservable
- Features in SIMPLIcity: multivariate

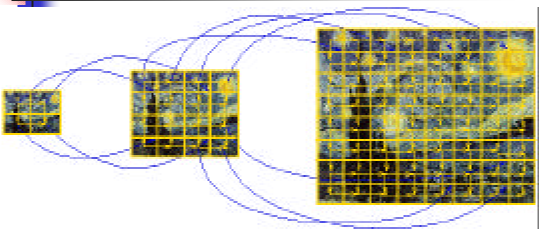
  Gaussian distributed

  given states



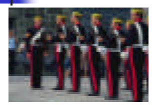
4/12/2002 J. Z. Wang, Penn State University 36

2-D MHMM

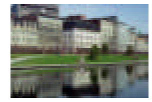


- Capture cross-resolution and intra-resolution context information
- Statistical dependence across resolution is assumed to be Markovian

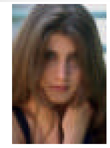
Preliminary Results



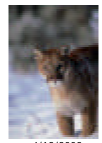
Computer Prediction: people, Europe, man-made, water



Building, sky, lake, landscape, Europe, tree



People, Europe, female



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

Food, indoor, cuisine, dessert

Outline

- Introduction
- Our SIMPLIcity work
- Experiments and results
- Our recent ongoing research
- Conclusions and future work

Conclusions

- A robust integrated region-based image retrieval algorithm
- Implemented in our SIMPLIcity system
  - Tested on 200,000 images
  - Improved accuracy and robustness, compared with some systems
  - Fast execution
- On-going: 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)
  - very large databases

Acknowledgments

- NSF DLI2
- The PNC Foundation
- SUN Microsystems
- Lockheed Martin Corp
- US Army NMTB (pending)
- Earlier funding: IBM QBIC, NEC AMORA, SRI AI, Stanford Lib/Math/Biomedical Informatics/CS

More Information

Papers in PDF,
60,000-image DB download,
demo, etc.

http://wang.ist.psu.edu