Machine Learning Approaches to Image Retrieval

Yixin Chen
Department of Computer Science
University of New Orleans
http://www.cs.uno.edu/~yixin

Outline

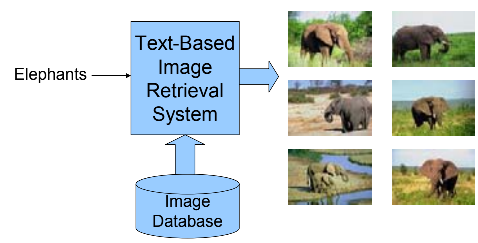
- Introduction
- Region-based image categorization using Multiple-Instance Learning
- Content-based image retrieval by clustering
- Conclusions and future work

Image Retrieval

- The driving forces
  - Internet
  - Storage devices
  - Computing power
- Two approaches
  - Text-based approach
  - Content-based approach

Text-Based Approach

- Input keywords descriptions

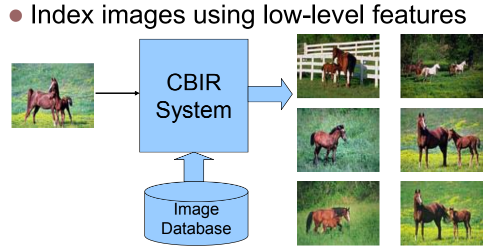


Text-Based Approach

- Index images using keywords (Google, Lycos, etc.)
  - Easy to implement
  - Fast retrieval
  - Web image search (surrounding text)
  - Manual annotation is not always available
  - A picture is worth a thousand words
  - Surrounding text may not describe the image

Content-Based Approach

- Index images using low-level features



Content-based image retrieval (CBIR): search pictures as pictures

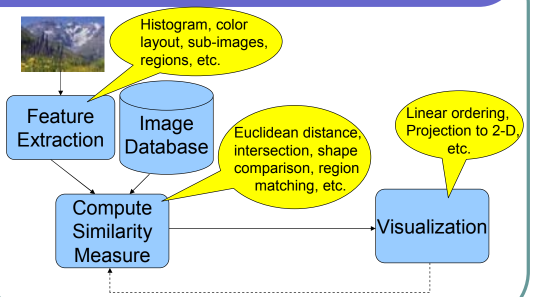
CBIR

- Applications
  - Commerce (fashion catalogue, ……)
  - Biomedicine (X ray, CT, ……)
  - Crime prevention (security filtering, ……)
  - Cultural (art galleries, museums, ……)
  - Military (radar, aerial, ……)
  - Entertainment (personal album, ……)

Previous Work on CBIR

- Starting from early 1990s
- General-purpose image search engines
  - IBM QBIC System and MIT Photobook System (two of the earliest systems)
  - VIRAGE System, Columbia VisualSEEK and WebSEEK Systems, UCSB NeTra System, UIUC MARS System, Stanford SIMPLIcity System, NECI PicHunter System, Berkeley Blobworld System, etc.

A Data-Flow Diagram

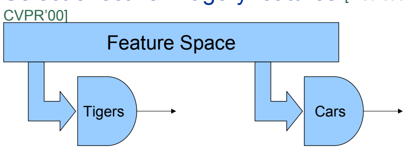


Open Problem

- Nature of digital images: arrays of numbers
- Descriptions of images: high level concepts
  - Sunset, mountains, dogs, ……
- Semantic gap
  - Discrepancy between low-level features and high-level concepts
  - High feature similarity may not always correspond to semantic similarity

Narrowing the Semantic Gap

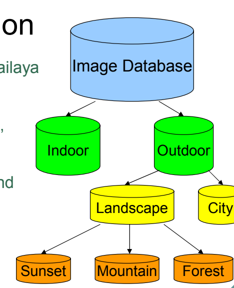
- Imagery features and similarity measure
  - Select effective imagery features [Tieu et al., IEEE CVPR'00]



  - Subjective experiments [Mojsilovic et al., IEEE Trans. IP 9(1)]

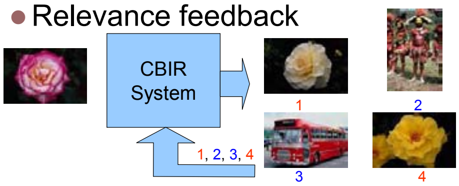
Narrowing the Semantic Gap

- Image Categorization
  - Vacation images [Vailaya et al., IEEE Trans. IP 10(1)]
  - SIMPLIcity [Wang et al., IEEE Trans. PAMI 23(9)]
  - Indoor/outdoor [Yu and Wolf, SPIE 1995]
  - ALIP [Li et al., IEEE Trans. PAMI 2003]



Narrowing the Semantic Gap

- Relevance feedback



  - Adjusting similarity measure [Picard et al., IEEE ICIP'96], [Rui et al., IEEE CSVT 8(5)], [Cox et al., IEEE Trans. IP 9(1)]
  - Support vector machine [Tong et al. ACM MM'01]

Outline

- Introduction
- Region-based image categorization using Multiple-Instance Learning
- Content-based image retrieval by clustering
- Conclusions and future work

Image Categorization

- Image categorization
  - Labeling of images into one of a number of predefined categories
- Difficulties
  - Variable and uncontrolled imaging conditions
  - Complex and hard to describe objects
  - Occlusion
  - Semantic gap

Motivation

- (a) to (d) belong to winter category since we see snow in them
- (b) to (f) belong to people category since there are people in them
- (b) to (d) belong to skiing category since we see people and snow
- (a) to (g) belong to outdoor scene category since they all have a region or regions corresponding to snow, sky, sea, trees, or grass

Problem Formulation

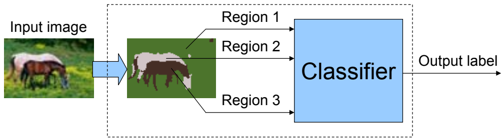
- Goal
  - Design a computer program that can "learn" image concepts from the implicit information of objects contained in images
- What is an "object"?
  - In the physical world: anything that is visible or tangible and is relatively stable in form
  - In an image: a region that is a projection of an object in the physical world

Problem Formulation

- An image is represented as a collection of regions obtained from segmentation

Problem Formulation

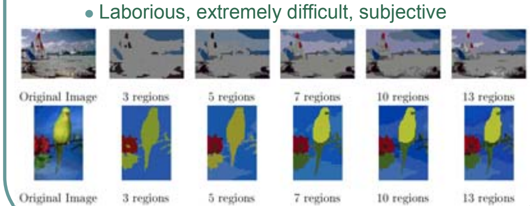
- An overview of the classification system



$$image_i = \{region_1,\ region_2,\ ......,\ region_{m_i}\} \subset R^d$$

Problem Formulation

- Training set: a set of labeled images
  - Region labels are unknown
    - Laborious, extremely difficult, subjective



Problem Formulation

- Learning with incomplete information
  - The classifier uses region features
  - Labels are associated with images instead of individual regions
  - A generalization of supervised learning
  - Simple tricks does not work well

Multiple-Instance Learning

- Bag (image), instance (region)
- Predict bag labels using instances
- The training data is a set of labeled bags

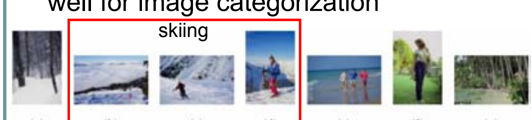
Multiple-Instance Learning

- Previous formulation [Dietterich, et al., AI'97], [Andrews, et al., NIPS'03], [Maron, et al., ICML'98], [Zhang, et al., ICML'02]
  - A bag is positive if at least one of its instances is a positive example; otherwise the bag is negative
  - Build an instance classifier
  - Bag label is equal to the label of its most "positive" instance

Multiple-Instance Learning

- Previous formulation does not perform well for image categorization

skiing



Snow	People	Sky	Trees
(a)	(e) (f)	(e) (f)	(a) (f) (g)

DD-SVM: An Extension of Multiple-Instance Learning

- A bag must contain some number of instances satisfying various properties
  - Find instance prototypes using Diverse Density [Maron et al., NIPS'98]
  - Define a bag feature space using instance prototypes
  - Design a maximal margin classifier in the bag feature space

Experiments

- 20 image categories, each containing 100 images

Africa	Beach	Dogs	Waterfall
Buildings	Buses	Lizard	Antiques
Dinosaurs	Elephants	Fashion	Battle ships
Flowers	Horses	Sunsets	Skiing
Mountains	Food	Cars	Dessert

Experiments

Experiments

Categorization Performance

- Classification accuracy (10-class)

DD-SVM	81.5% ± 2.2%
Hist-SVM	66.7% ± 1.8%
MI-SVM [Andrews et al., NIPS'03]	74.7% ± 0.5%

An Image Classification Example

- Confusion matrix

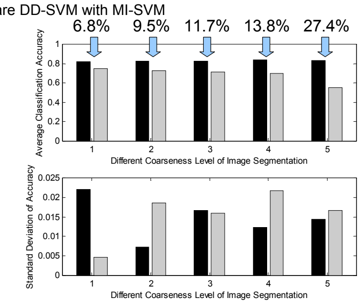
%	Africa	Beach	Buildings	Buses	Dinosaurs	Elephants	Flowers	Horses	Mountains	Food
Africa	72	2	4	2	0	12	2	0	0	6
Beach	0	54	8	6	0	2	2	0	26	2
Buildings	4	0	84	2	0	6	0	0	2	2
Buses	0	4	0	94	0	0	0	0	0	2
Dinosaurs	0	0	0	0	100	0	0	0	0	0
Elephants	14	0	6	0	0	74	0	4	0	2
Flowers	2	0	0	0	0	0	92	0	0	6
Horses	2	2	0	0	0	0	0	96	0	0
Mountains	0	8	12	4	0	4	0	0	72	0
Food	4	2	0	2	0	0	4	0	0	88

Categorization Performance

Some errors between **Beach** and **Mountains** categories

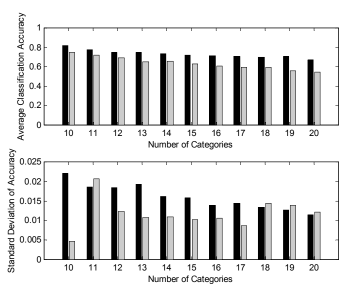
Sensitivity to Image Segmentation

Compare DD-SVM with MI-SVM



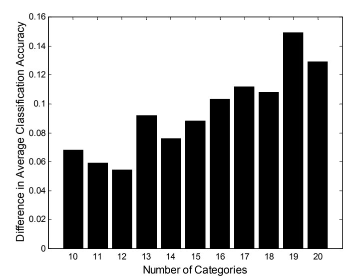
Scalability

Compare DD-SVM with MI-SVM



Scalability

Difference in classification accuracy between DD-SVM and MI-SVM



Outline

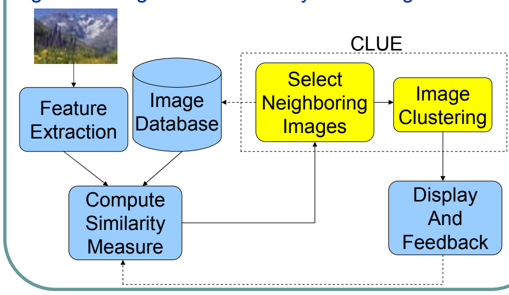
- Introduction
- Region-based image categorization using Multiple-Instance Learning
- Content-based image retrieval by clustering
- Conclusions and future work

CLUE: CLUsters-based rEtrieval of images by unsupervised learning

- Basic idea
  - All CBIR methods assume some correlation between image semantics and distance measure
  - Why not using this information to the furthest extent

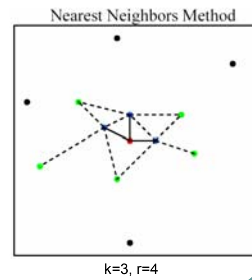
System Overview

A general diagram of a CBIR system using the CLUE



Neighboring Images Selection

- Nearest neighbors method
  - Pick *k* nearest neighbors of the query as seeds
  - Find *r* nearest neighbors for each seed
  - Take all distinct images as neighboring images



Weighted Graph Representation

- Graph representation
  - Vertices denote images
  - Edges are formed between vertices
  - Nonnegative weight of an edge indicates the similarity between two vertices

$$w_{ij} = e^{-\frac{d(i,j)^2}{s^2}}$$

- Recursive Ncut
  - Bipartition the largest sub graph each time

User Interface

An Experimental System

- Similarity measure
  - UFM [Chen et al. IEEE PAMI 24(9)]
- Database
  - COREL
  - 60,000

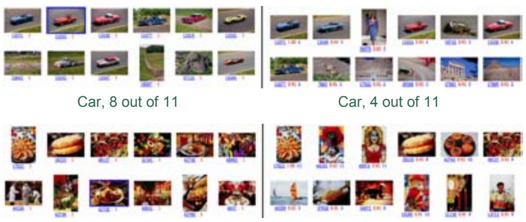
Query Examples

- Query Examples from 60,000-image COREL Database

Bird, car, food, historical buildings, and soccer game

Query Examples

CLUE | UFM

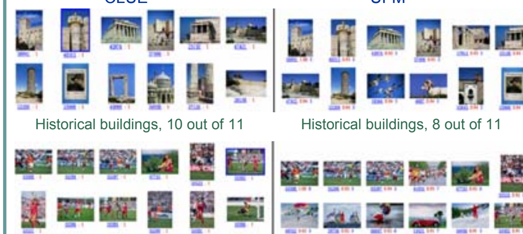


Car, 8 out of 11 | Car, 4 out of 11

Food, 8 out of 11 | Food, 4 out of 11

Query Examples

CLUE | UFM



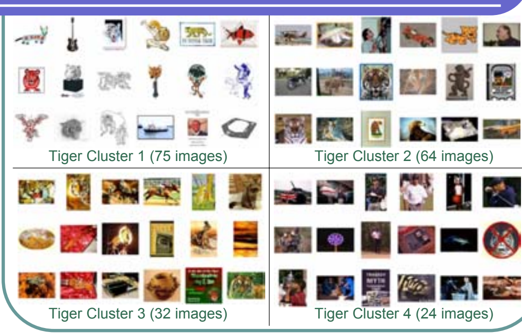
Historical buildings, 10 out of 11 | Historical buildings, 8 out of 11

Soccer game, 10 out of 11 | Soccer game, 4 out of 11

Clustering WWW Images

- Google Image Search
  - Keywords: tiger, Beijing
  - Top 200 returns
  - 4 largest clusters
  - Top 18 images within each cluster

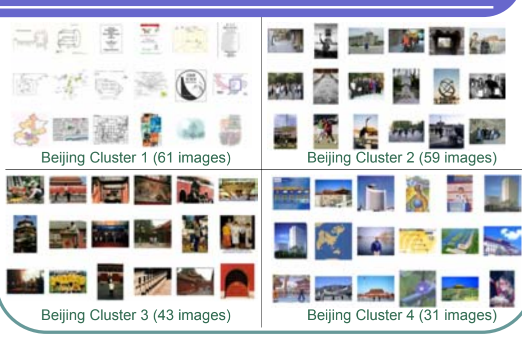
Clustering WWW Images



Tiger Cluster 1 (75 images) | Tiger Cluster 2 (64 images)

Tiger Cluster 3 (32 images) | Tiger Cluster 4 (24 images)

Clustering WWW Images

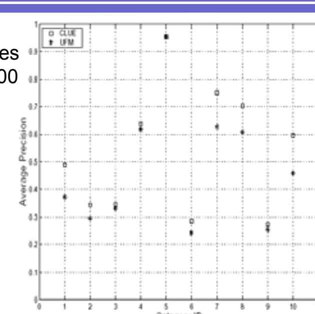


Beijing Cluster 1 (61 images) | Beijing Cluster 2 (59 images)

Beijing Cluster 3 (43 images) | Beijing Cluster 4 (31 images)

Retrieval Accuracy

10 image categories each containing 100 images



Outline

- Introduction
- Region-based image categorization using Multiple-Instance Learning
- Content-based image retrieval by clustering
- Conclusions and future work

Summary

- DD-SVM
- CLUE

Limitations

- Image categorization
  - Diverse Density
- CLUE
  - Recursive Ncut
  - Representative images
  - Sparsity

Future Work

- Bag generator
- Generative model
- Applications

Supported by

- The National Science Foundation
- The Pennsylvania State University
- The PNC Foundation
- SUN Microsystems
- NEC Research Institute
- University of New Orleans
- Research Institute for Children

Acknowledgment

- Dissertation committee at Penn State
  - Professor James Z. Wang, IST and CSE
  - Professor Lee C. Giles, IST and CSE
  - Professor John Yen, IST and CSE
  - Professor Jia Li, STAT
  - Professor Donald Richards, STAT
- NEC Research Institute
  - Dr. Robert Krovetz
- Siemens Medical Solutions
  - Dr. Jinbo Bi
- Kind host
  - Dr. Andrés Castaño

More Information

- Papers in PDF, demonstrations, data sets, etc.

  http://wang.ist.psu.edu/IMAGE
  http://www.cs.uno.edu/~yixin

  yixin@cs.uno.edu