EE 6882
Statistical Methods for Video Indexing and Analysis

Instructors:
Prof. Shih-Fu Chang, Columbia University
Dr. Lexing Xie, IBM T.J. Watson Research

TA:
Eric Zavesky

Fall 2007, Lecture 1
Course web site: http://www.ee.columbia.edu/~sfchang/course/svia

dvmm
DIGITAL VIDEO • MULTIMEDIA LAB

EE E6882 SVIA Lecture #1

- Introduction, Course Syllabus
- Readings (available on course site)
  - Rui et al, Content-Based Image Retrieval Review paper
  - A. Jain et al, "Statistical Pattern Recognition: A Review," IEEE Tran. on Pattern Analysis and Machine Intelligence, vol 22, No 1, Jan. 2000.
  - Gonzalez and Woods, Digital Image Processing, 2nd edition, Prentice Hall, 2001 (Chapter 12, Object recognition)
- Next Week:
  - Sept. 17th 2007 (Prof. Xie)
  - Topic: Content Based Image Retrieval

dvmm
DIGITAL VIDEO • MULTIMEDIA LAB

Topics: Image/Video Search

- Explosive growth of online image/video data, personal media, broadcast news videos, etc.
- 5 billion images on the Web, 31 million hours of TV programs each year
- Successful services like Youtube and Flickr
  - Others: blinkx.com, like.com, etc
- Image/video search exciting opportunity

Different Visual Search Models

- Browsing and Grouping
  - Subject listing (e.g., WebSeek, http://www.ee.columbia.edu/webseek)
  - Animation summary (e.g., http://www.blinkx.com)
- Keyword Search
- Content-Based Search
  - E.g., VisualSeek, like.com

User Expectation in Practice

> "…type in a few words at most, then expect the engine to bring back the perfect results. More than 95 percent of us never use the advanced search features most engines include, …" – *The Search*, J. Battelle, 2003

- Keyword search is the primary search method.

Google Zeitgeist publishes top keywords monthly

Google Image Queries: April 2007

Newsmakers	Getaways
1. virginia tech	1. hawaii
2. knut	2. dubai
3. yuri gagarin	3. mexico
4. shaha riza	4. chelsea
5. kurt vonnegut	5. london

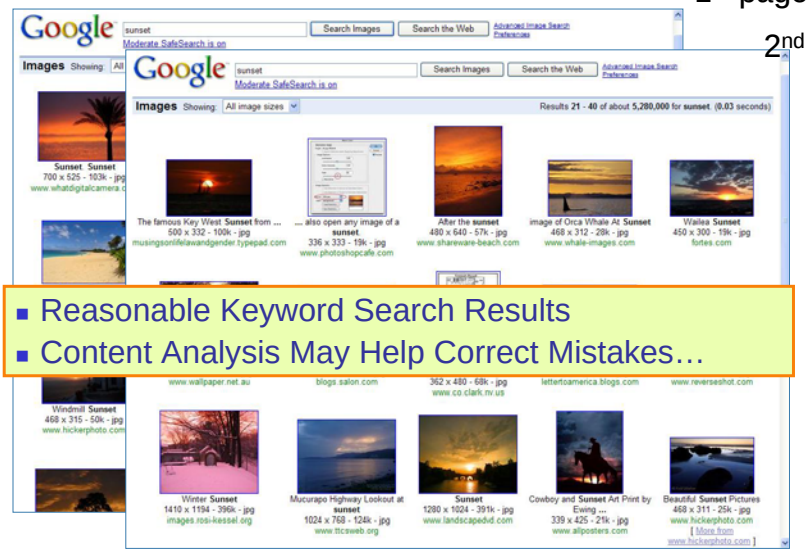
At the cineplex	Global consciousness
1. ghost rider	1. global warming
2. shrek	2. world map
3. borat	3. earth day
4. 300 movie	4. southern cross
5. spiderman 3	5. iceberg

Examples of Keyword Image Search

query: “sunset”

1st page

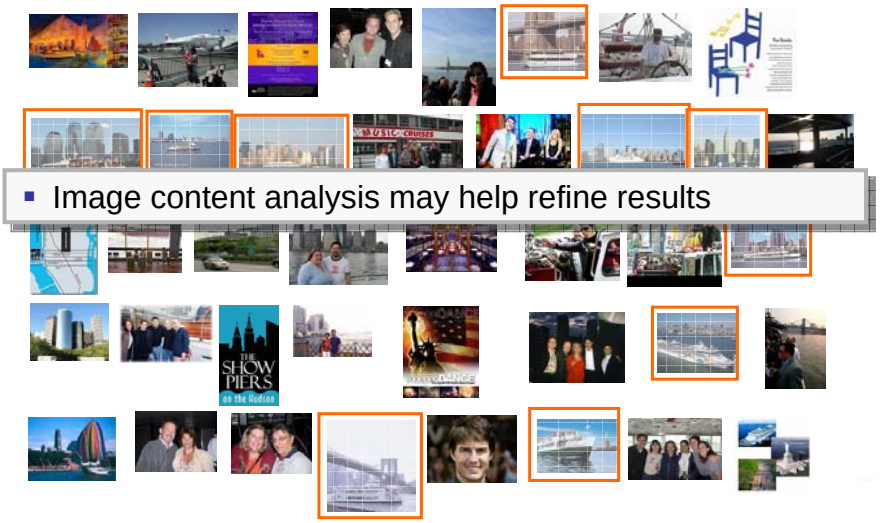
2nd page



- Reasonable Keyword Search Results
- Content Analysis May Help Correct Mistakes…

Example Search

- Text Query on Google: *“Manhattan Cruise”*



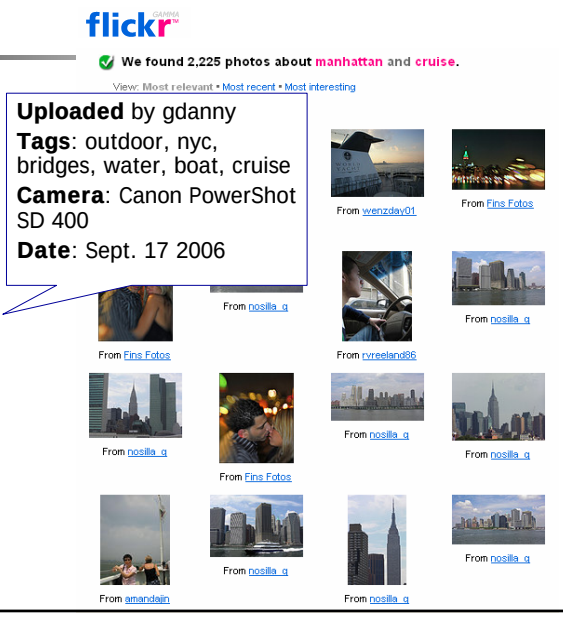
- Image content analysis may help refine results

-8-

How about Social-Net Tagging?

- Yahoo-flickr millions of users,

Social tags may be subjective and incomplete.



Insufficient Precision of Social Tags

- Test New York City landmark labels

	precision
Bronx-Whitestone Br.	1.00
Brooklyn Br.	0.38
Chrysler Building	0.65
Columbia University	0.30
Empire State Building	0.18
Flatiron Building	0.70
George Washington Br.	0.48
Grand Central	0.37
Times Square	0.56
Verrazano Narrows Br.	0.66
World Trade Center	0.13

Many tags from social networks are of low precision
(due to batch uploading?)

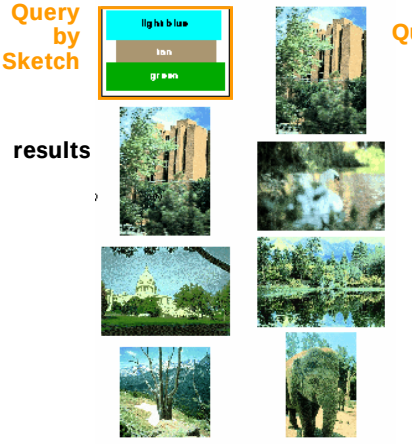
An Interesting Paradigm:
Image Tagging via Game Playing

(Von Ahn & Dabbish, CHI 04)

- Used in *Goggle Image Labeler* (http://images.google.com/imagelabeler/)
- Use competitive games to motivate users
- Has attracted many participants for free!
  - Some users spent hours in a day
- Claim the potential of annotating the whole Web in just few months!
  - 5 Billion images

Seeking the image search tools
-- Content-Based Image Retrieval (CBIR)

IBM QBIC '95, Columbia VisualSEEk '96



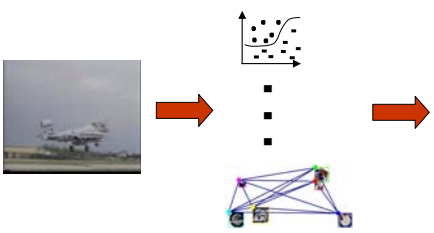
Issues

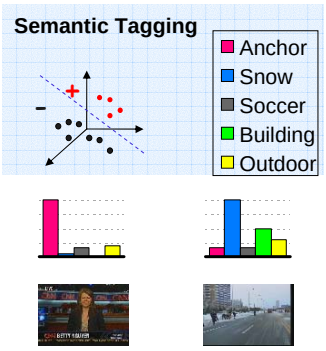
- What image features to extract?
- How to match images and videos?
- How to make it fast?

Opportunity for Content Analysis: Large-Scale Auto. Image Tagging Framework

- Audio-visual features
- Surrounding text
- SVM or graph models
- Context fusion

- Rich semantic description based on content analysis





Large-Scale Concept Detectors from Research Community

- Columbia374
  - 374 baseline detectors for LSCOM multimedia ontology
- MediaMill
  - 491 concept detectors for LSCOM and MediaMill 101 Lexicons
- IBM MARVEL Search System
  - Trials with BBC, CNN
  - Real-time standalone detectors from IBM AlphaWorks
- Others ...

What Concept to Detect?

- One effort: Large Scale Concept Ontology for Multimedia (LSCOM)
  - Joint effort by news/intelligence analysts, librarians, researchers
  - Broadcast News Domain
  - Selection Criteria
    - useful, detectable, observable
  - 834 concepts defined, 449 concepts annotated
  - Labeled over 61,000 shots of TRECVID 2005 data set
    - 33 Million judgments collected, 100 person-month labor
  - Download by 170+ groups so far
    - http://www.ee.columbia.edu/dvmm/lscom/

LSCOM Concepts (449)

- Event/Activity (56 - 13%)
  - Airplane taking off, car crash, explosion, etc
- People (113 - 25%)
  - Person, male/female, firefighter, etc
- Location (89 - 20%)
  - Cityscape, hospital, airfield, etc
- Object (135 - 30%)
  - Vehicle, map, tank, power plant, etc
- Scene (49 - 10%)
  - Vegetation, urban, interview, etc
- Program (7 - 2%)
  - Entertainment, weather, finance, etc

Consumer Video Ontology
(Kodak-Columbia, 2007)

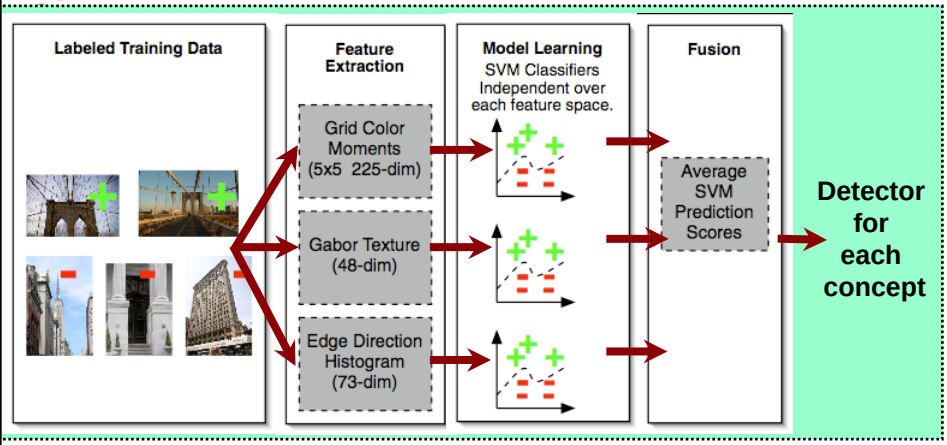
- Activity (6)
- Occasion (16)
- Scene (15)
- Object (25)
- People (11)
- Sound (14)
- Camera Motion (5)
- Object Motion (3)
- Social (4)

Activity:
Occasion :
Scene:
Object:
People:
Sound:
Camera Motion:
Object Motion:
Social: friend, family, classmate, colleague

Research Issues

- How to develop automatic tagging tools?
  - Train automatic recognition models
    - What image features?
    - What statistical models?
  - Explore surrounding information
    - Time, location (e.g., Yahoo! Zonetag, http://zonetag.research.yahoo.com/)
    - Text and metadata

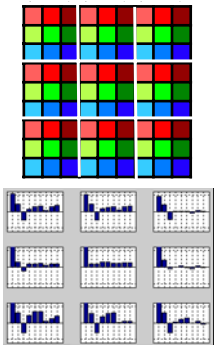
Building Image Classifiers – Basic



- General for all concepts, easy to implement
- 374 baseline detectors (*Columbia 374*) released

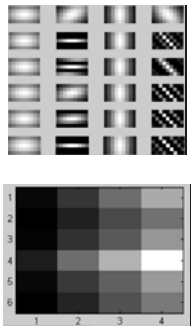
Examples of Basic Image Features

grid layout + color moment



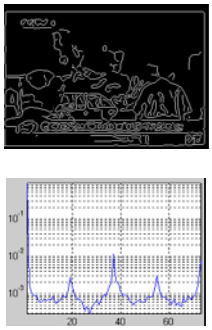
225 dimensions

Gabor texture



48 dimensions

edge direction histogram

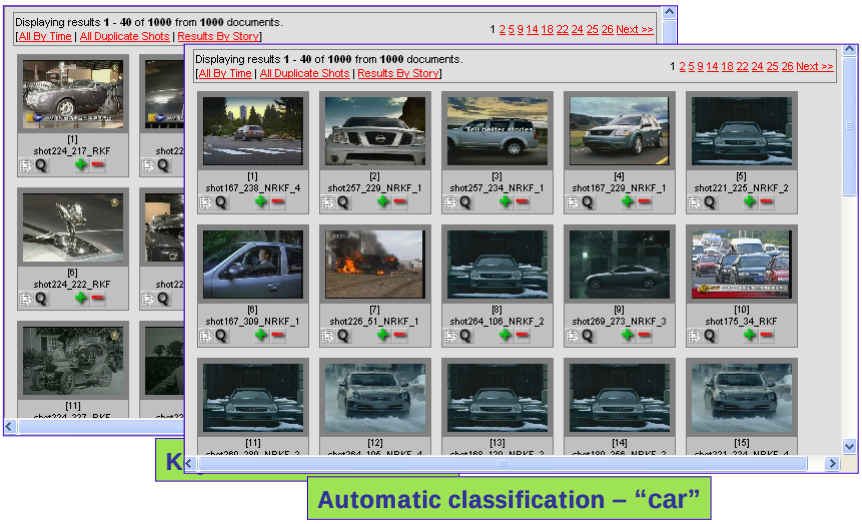


73 dimensions

Text search vs. visual classification

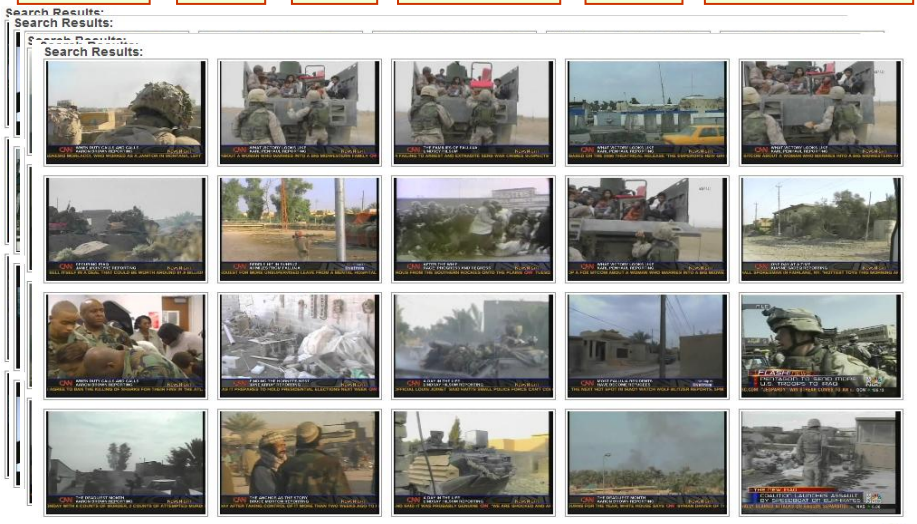
(images from TRECVID)

Text search vs. visual classification

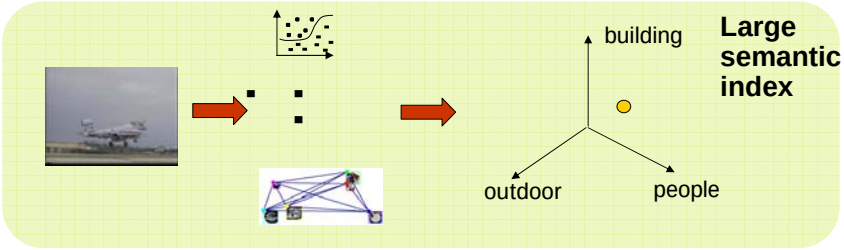


Example: good detectors for LSCOM concept

waterfront | bridge | crowd | explosion fire | US flag | Military personnel



Power of Concept-based Representation



New applications:
Search, Filtering, Pattern Mining

Mapping search topics to concepts

TRECVID search topics

Finds shots with one or more emergency vehicles in motion (e.g., ambulance, police car, fire truck, etc.)

Matched Concepts:
Emergency_Room, Vehicle

Find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible.

Matched Concepts: Building

Find shots with … or entering a ve… re soldiers, police, …oner.

Research issue:
what concept to use?
How to fuse multiple concepts?

Matched Concepts: Person, Vehicle

Matched Concepts:
Guard, Police_Security, Prisoner, Soldier

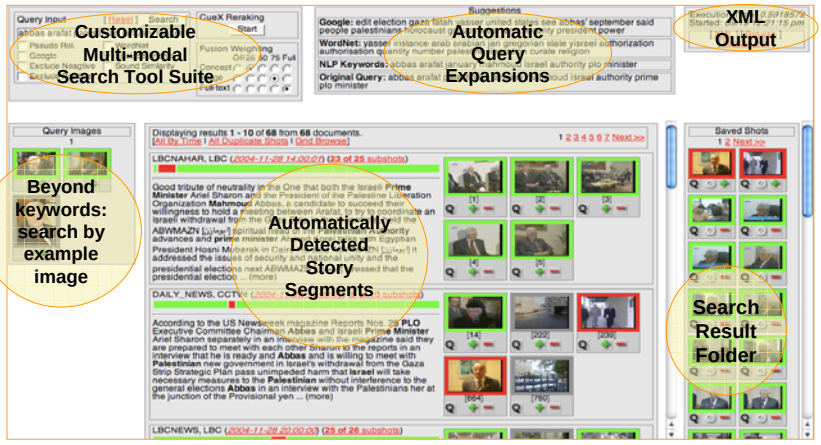
Concept Search Demo

- Interactive demos available at
  http://apollo.ee.columbia.edu/vace/newSearch/
- Concept search case 1 (link)
- Concept search case 2 (link)
- Multimodal search (link)

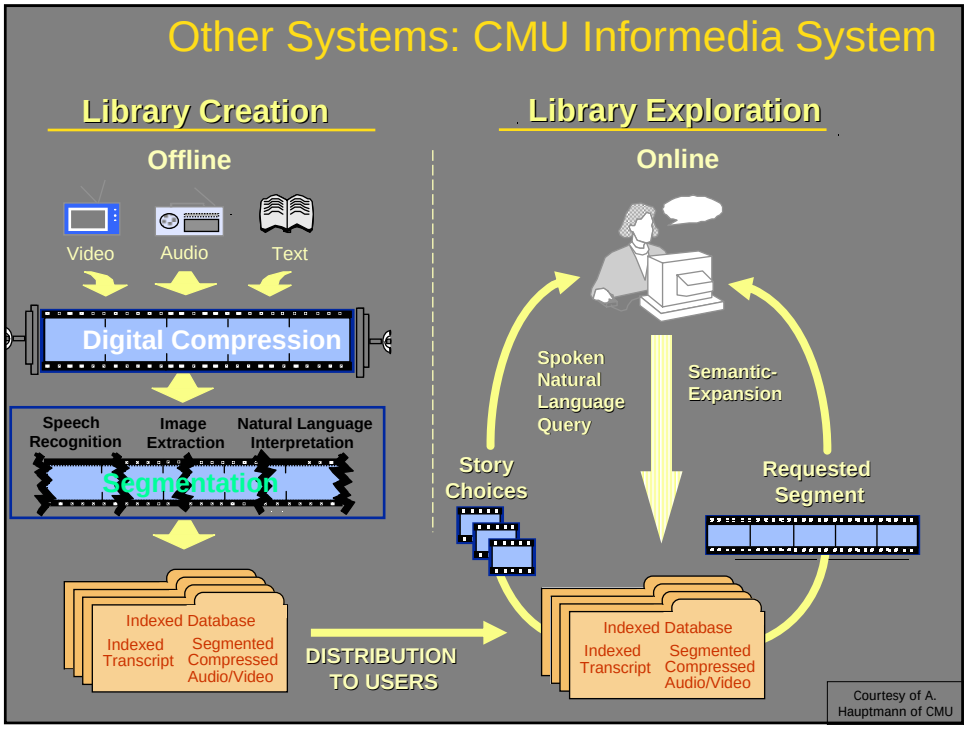
Demos prepared by Eric Zavesky

CuVid: Columbia Video Search System

http://www.ee.columbia.edu/cuvidsearch



Prototype includes 160 hours, 3 languages (English, Arabic, Chinese), 6 channels



Problems Studied in this Course

- Content Based Image Retrieval
  - Feature extraction
  - Image/Video matching methods
  - Efficient indexing: search millions or billions of images
- Image/Video Copy Detection Methods
- Image Annotation Strategies
  - Make image annotation more attractive
- Automatic Classification and Tagging
  - Statistical models
  - Contextual information
- Multimodal Search Using Text, Image, and Others
- Strategies for Searching Media on Social Networks

About the course

- Objectives:
  - Learn how to formulate and solve problems in this field
  - Get insights and experience of recent pattern recognition/machine learning techniques
  - Hands on experiments with image/video classification/indexing problems
- Intended Audience
  - Beginning graduate students or professionals
  - familiar with signal/image processing
  - comfortable with probability, statistics, linear algebra, and some machine learning

Course Format

- Overview Lectures + student presentations + final projects
- We will give several overview lectures at the beginning.
- 1 hands-on homework on image search (assigned in week 2)
- Student paper presentation (starting week 5)
  - One paper assigned to each student
  - assignments determined 2-3 weeks in advance
- Everyone writes comments before class on the web site
- One final term project (1-2 people per team)
- Grading
  - Paper presentation/demo 30%
    Class participation/homework 30%
    Final Project 40%

Paper review and presentation

- Each student discusses paper and experiments with us 3 weeks before class
  - Week 1: review and research
  - Week 2: simulate a toy problem using available data set and tools
  - Week 3: prepare presentation
- Other students post comments and questions before class
- Presentation
  - 30 mins each paper (including demo if available)

Paper Review and Demo (2)

- Review
  - Background review and examples
  - Problem addressed and main ideas
  - Insights about why it works
  - Limitation, generality, and repeatability
  - Alternatives and comparisons
- Experiments
  - Check software and data available and repeatable
  - Reconstruct the method and try on toy data sets
  - Analyze results (not just accuracy numbers, offer explanations and verifiable theories about observations)
  - Demo code archived on class site and shared with others

Resources and Matlab

- Links on the class web site
  - Tutorials on paper writing, Matlab, etc
- Software links on web site to
  Matlab, Neural Network, HMM, Netlab, SVM
- SVIA EE6882 Class Dataset
  - Benchmark data set, a few thousands of images from broadcast news and stock photos
  - Extracted features and labels
  - Available through TA
- Matlab is often used for programming, C/Java welcome
  - Accessible on university computers
  - Very brief introduction next week

Paper Review last year
(www.ee.columbia.edu/~sfchang/course/svia-F04)

- Feature Selection for SVM
- Fast multiresolution image querying
- Relevance Feedback in Image Retrieval
- MPEG-7 Color and Texture Features
- SVM Image Classification
- SVM Active Learning
- Maximum Entropy for Story Segmentation
- HMM for Video Parsing
- Relevance Model for Image Retrieval
- Video Fingerprinting

Final Projects last time (2004)

- Many students extend topics chosen for paper review/experiments
  - SVM feature selection for news story segmentation
  - Wavelet multiresolution image retrieval
  - Comparison of relevant feedback methods for image retrieval
  - Object Search over 3D VR object database
  - Michael and Graham
    Relevance Feedback for music retrieval
  - SVM image classification
  - HMM for news story segmentation
  - Motion based object segmentation and classification
  - MPEG-7 CSS Shape feature evalution

Other information

- Student presentations and codes from last year will be available
- Office Hours
  - Instructors: Mondays 3-4, Mudd 1300
  - TA: Eric Zavesky, emz2101@columbia.edu, Wed. 3:30-5pm, CEPSR 708