



Image/Video Search: Addressing Semantic Gap and User Gap

Shih-Fu Chang

Digital Video and Multimedia Lab

<http://www.columbia.edu/dvmm>

Acknowledgments

- Columbia University
 - Yu-Gang Jiang, Wei Jiang, Eric Liu, Jun Wang, Akira Yanagawa, Eric Zavesky
- IBM Research
 - Lexing Xie, Rong Yan, Paul Natsev, John Smith
- National University of Singapore
 - Tat-Seng Chua
- University of Amsterdam (MediaMill)
 - Ork de Rooij, Cees Snoek, Marcel Worring
- City University, Hong Kong
 - Chong-Wah Ngo
- IRIT, France
 - Elie El Khoury



The Explosive Growth of Image/Video Data

broadcast



70,000 TB/year, 100 million hours



UGC portals



80% of Internet Traffic on Video Streaming/sharing



video blogs



11.5 billion online video views in March 2008



- Explosive growth of videos in broadcast, entertainment, User Generated Content (UGC), blog

- Impact on citizen journalism and social network

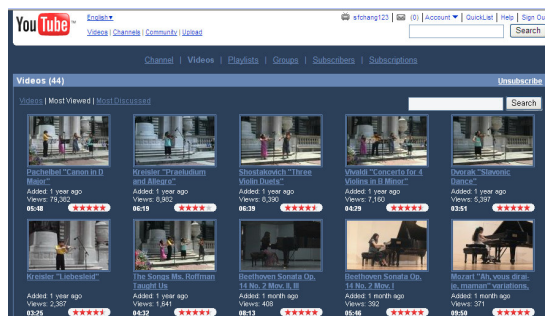
3

(Slide from Lexing Xie of IBM)

Example:

My family's video channel

- 70 videos uploads to Youtube in two years
- Single top video viewed more than 100,000 times
 - More than the total views of my published papers!!!



video

Name	URL	Alexa	Google PR	Language
Yahoo Video	http://video.yahoo.com	1	8	English
Google Video	http://video.google.com	2	8	English
Youtube	http://www.youtube.com	3	9	English
Video.qq	http://video.qq.com	16	6	Chinese
You.video.sina.cn	http://v.sina.com.cn	18	7	Chinese
Flickr	http://www.flickr.com	31	4	English
Photobucket	http://photobucket.com/recent/videos	36	6	English
Youku	http://www.youku.com	56	7	Chinese
Daily Motion	http://www.dailymotion.com	64	7	English
Rediff	http://ls.rediff.com	66	2	English
Rambler Vision	http://vision.rambler.ru	72	7	English
4Shared	http://www.4shared.com	100	6	English
Your File Host	http://www.yourfilehost.com	103	5	English
Tudou	http://www.tudou.com	115	7	Chinese
Metacafe	http://www.metacafe.com	119	7	English
KU6	http://www.ku6.com	122	7	Chinese
TinyPic	http://www.tnypic.com	157	6	English
Cnet TV	http://onetv.cnet.com	158	7	English
Libero	http://video.libero.it	161	6	Italian
Multiply	http://multiply.com	162	7	English
Imeem	http://www.imeem.com	171	7	English
Video.web.de	http://video.web.de	182	6	German
ting	http://www.ning.com	199	8	English
6.cn	http://6.cn	217	6	Chinese
56	http://www.56.com	261	7	Chinese
Vision.Ameba.jp	http://vision.ameba.jp	271	6	Japanese
Justin TV	http://www.justin.tv	299	6	English
Break	http://www.break.com	323	7	English
Hulu	http://www.hulu.com	342	7	English
Video.eksenim.mynet	http://video.eksenim.mynet.com	383	5	Turkish

Top Video Sharing Sites (338)

- 12 in top 100 web traffic ranking (Alexa)
- truly international, no language barrier
- video as a universal language for sharing experience

(survey by Mark Robertson Jan 09)



What's available today

- Example
 - Search over 35 million videos on **blinkx**
- *"find videos about Iran election"*
([video](#))
- *"find videos showing Cruise in Manhattan"*
([video](#))



Stress Tests of Existing Search Engines

"Manhattan Cruise"
"Cruise ship in Manhattan"

The screenshot shows a Google Images search for "cruise ship in manhattan". The search bar at the top contains the text "cruise ship in manhattan" and "Search Images". Below the search bar, there are filters for "All image sizes", "Any content", and "All colors". The search results are displayed in a grid of 24 thumbnails, each with a title and a source URL. The thumbnails show various cruise ships and scenes in Manhattan. The source URLs include nydailynews.com, flickr.com, zazzle.com, and others. The search results are sorted by relevance, and the page shows "Results 1 - 18 of about 236,000 (0.17 seconds)".

A Quick Assessment

- Infrastructure is ready for acquisition, sharing, and streaming.
- User adoption is happening.
- But search solutions are still preliminary
 - relying on textual metadata
- Remember many videos do not have metadata ready for search
 - Surveillance
 - Consumer videos in the shoebox

Vision: Filling the Semantic Gap

- Audio-visual features
- Geo, social features
- User context
- Context fusion

- Rich semantic description based on content analysis

S.-F. Chang, Columbia U.
9

Concept-based Video Search

(Slide from Yu-Gang Jiang)



Important issues

- What are useful and feasible categories?
 - A picture is worth a thousand words ...
 - Which 1000?
- How to build automatic concept classifiers
 - The *semantic gap* problem
- How to search over such analytics models?
 - The *user gap* problem

S.-F. Chang, Columbia U.

11



LSCOM Ontology

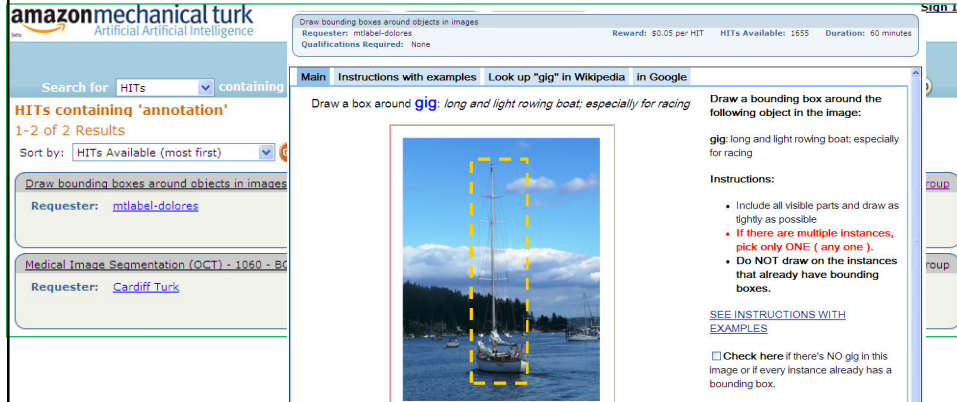
(Naphade, Hauptmann, Smith, Chang, *et al*'06)

- Large Scale Concept Ontology for Multimedia
 - Broadcast news video
 - Defined by experts and actual intelligence users
 - *Useful, observable, detectable* concepts
 - 30M+ labels for 449 concepts annotated over 60,000+ video shots
 - Download by 450+ groups
 - <http://www.ee.columbia.edu/dvmm/lscom/>

S.-F. Chang, Columbia U.

12

Crowd sourcing for image annotation



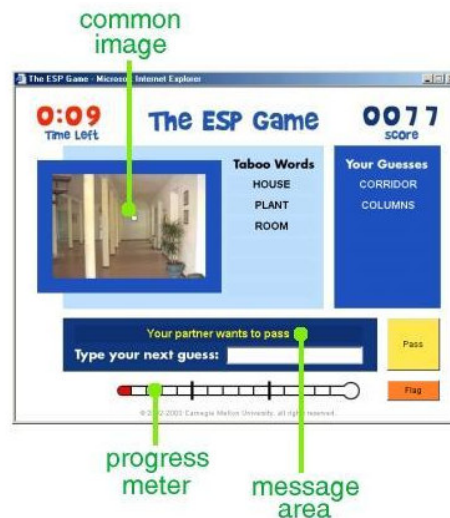
Such task is nicely called Human Intelligence Task (HIT)!

15

Image Labeling via game playing

- Use competitive games to motivate users
- Select labels based on agreement
- Much more fun
- Concerns with the label quality

[Video](#)



(Von Ahn & Dabbish, CHI 04)

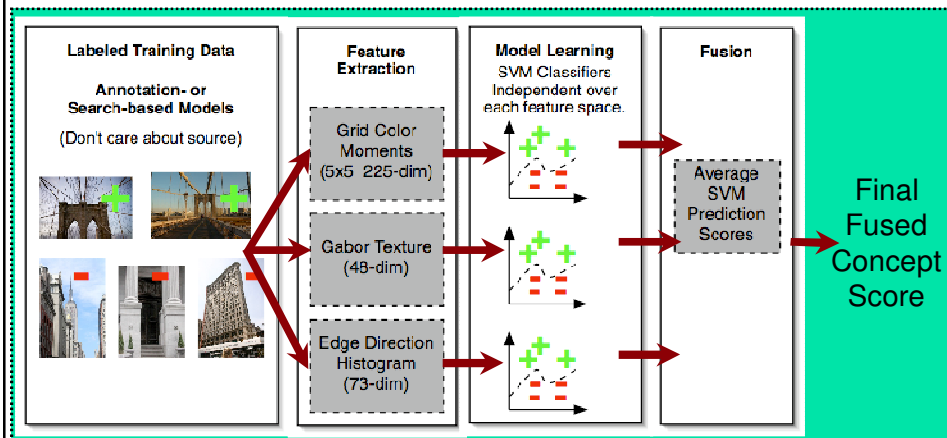
Important issues

- What are useful and feasible categories?
 - A picture is worth one thousand words ...
 - Which 1000?
- How to build automatic concept classifiers
 - The semantic gap problem
- How to search over such analytics models?
 - The user gap problem

S.-F. Chang, Columbia U.

17

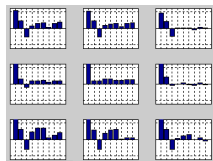
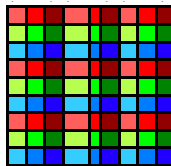
Simple Classifiers using Global Features



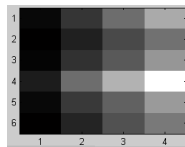
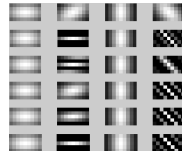
Very Simple Global Visual Features



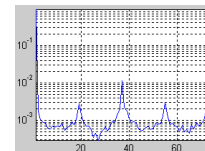
grid color moment



texture



edge direction



- Capture global distributions plus layout
- Popular choice for scene classification

Keypoint Localization



- Goals:
 - Repeatable detection
 - Precise localization
 - Interesting content
- ⇒ *Look for two-dimensional signal changes*

Counting Local Key Points: Bags of Words

The diagram illustrates the process of counting local key points using a Bag of Words (BoW) model. It is divided into two main parts: training and testing.

Training (top): A training set of images is processed through **Keypoint extraction**. The resulting keypoints are then subjected to **Clustering** in a **Keypoint feature space**. This process identifies distinct clusters of keypoints, which are then mapped to a **Visual word vocabulary** represented by different colored shapes (e.g., red circle, green diamond, yellow square, blue triangle).

Testing (bottom): A new image is processed through **Keypoint extraction**. The extracted keypoints are mapped back to the **Visual word vocabulary**. This results in a **BoW histogram**, which is a bar chart showing the frequency of each visual word (labeled $V_1, V_2, V_3, V_4, \dots$) in the image.

(Sivic and Zisserman, ICCV 2003)

S.-F. Chang, Columbia U. 21

Local Features Multi-BoW Spatial Pyramid Kernel

The diagram illustrates the Multi-BoW Spatial Pyramid Kernel (SPK) process, which combines local feature extraction with multi-scale BoW representations.

Vocabulary Generation (left): A **SIFT feature space** is processed through three different keypoint detectors: **DoG**, **Hessian Affine**, and **MSER**. Each detector identifies keypoints that are then mapped to a specific **Visual word vocabulary** (labeled 1, 2, and 3 respectively).

BoW Representation (right): A new image is processed through the same three detectors. The resulting keypoints are mapped to their respective visual word vocabularies. These are then combined into **BoW histograms Using Soft-Weighting** for different spatial pyramid pooling scales. These histograms are then fed into **SVM classifiers** for classification.

[Jiang, Ngo, Yang, CIVR07]
[Lazebnik, Schmid, Ponce, CVPR06]

S.-F. Chang, Columbia U. 22

How to Fuse Kernels from Different Scales/Blocks/Codebooks?

$$K = \sum_{j=1}^J u_j K_j, \quad u_j \geq 0$$

A fast quadratic programming solution:

$$\min_u \|K - Y\|_F^2 + \lambda \|u\|^2$$

$$s.t., K = \sum_{j=1}^J u_j \overline{K_j},$$

$$u_j \geq 0, j = 1, \dots, J, \quad \sum_{j=1}^J u_j = 1,$$

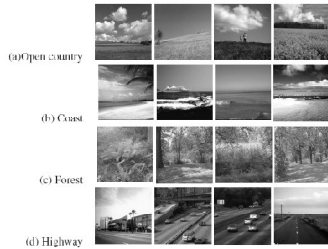


Figure 1. Examples of pictures in the natural scene dataset

“open country” vs. “coast”



“open country” vs. “forest”

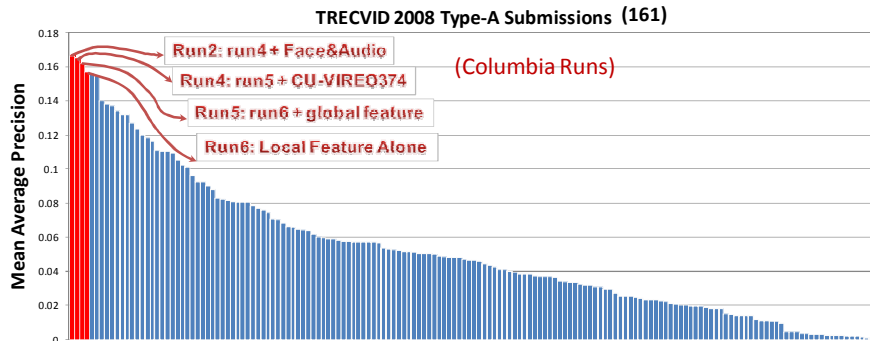


“open country” vs. “highway”



(He, Chang, Xie, CVPR08)

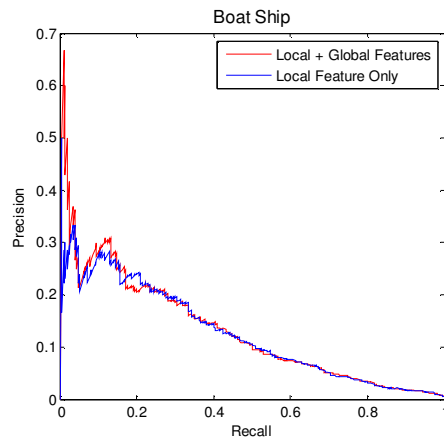
Local Features Prevail in Recent TRECVID



- Spatial local features achieve near top performance (Chang, et al, TRECVID 2008)
- Other features (global, context, face, audio) help
- 374 detection models available for download
 - Global Features (Columbia374) <http://www.ee.columbia.edu/ln/dvmm/columbia374/>
 - Local Features (CU-VIREO374) <http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/>

PR-Curve: Boat-Ship

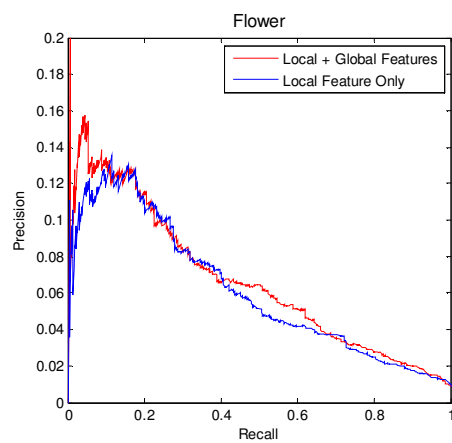
	Avg. precision (AP)
Local + global features	0.238
Local features only	0.228



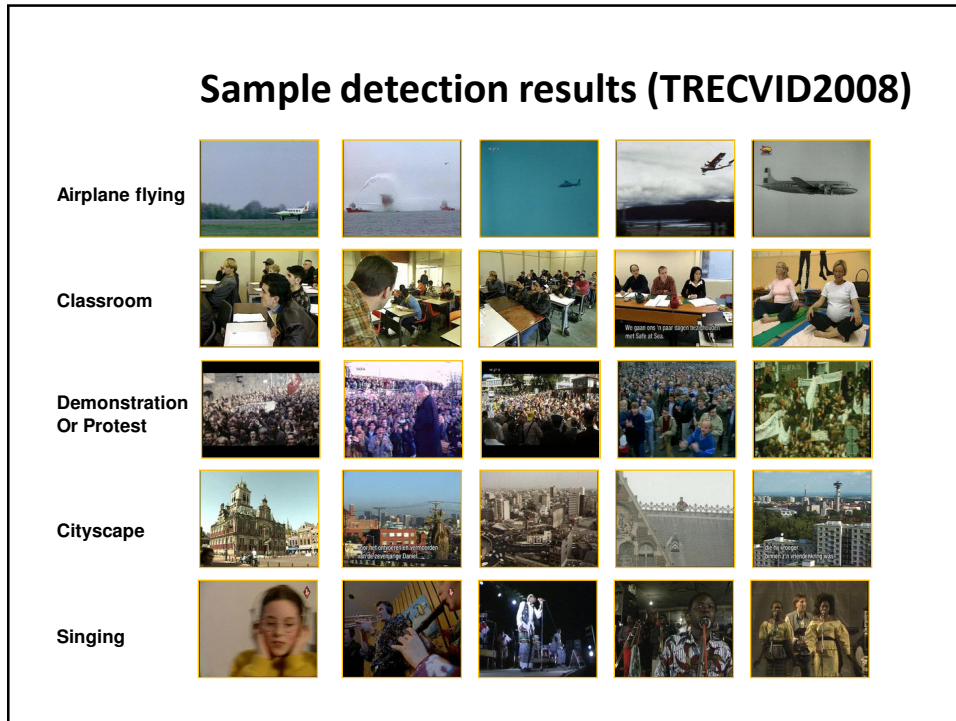
No. of training samples: 437

PR-Curve: Flower


	Avg. precision (AP)
Local + global features	0.091
Local features only	0.082



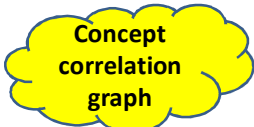
No. of training samples: 582



Explore Concept Correlation: Semantic Diffusion via Graph



Individual Classifier s:
Desert: 0.68; Sky: 0.60;
Weapon: 0.38; Car: 0.43;
Vehicle: 0.35 ...



correlation matrix

Classifier c_j score

$$\mathcal{E}(g) = \frac{1}{2} \sum_{i=1}^C \sum_{j=1}^C W_{ij} \left\| \frac{g(c_i)}{\sqrt{d(c_i)}} - \frac{g(c_j)}{\sqrt{d(c_j)}} \right\|^2$$

$$(g^*, \tilde{W}^*) = \arg \max_{g, \tilde{W}} \mathcal{E}$$

Adapting Graph Weights to New Domain

Broadcast News

Documentary

Graph optimization

$$(g^*, \tilde{W}^*) = \arg \max_{g, \tilde{W}} \mathcal{E}$$

(Jiang, Ngo, and Chang, ICCV09)

Iteration: 00

The correlation model does not fit the new domain

Effects of concept correlation

TRECVID 2006 Test Data

Concept	Baseline	Semantic Diffusion
Sports	0.38	0.45
Weather	0.42	0.45
Office	0.08	0.10
Meeting	0.28	0.32
Desert	0.05	0.08
Mountain	0.18	0.22
Water	0.10	0.13
Corporate-Leader	0.02	0.03
Police_Security	0.02	0.03
Military	0.10	0.13
Animal	0.01	0.01
Computer_TV-screen	0.25	0.28
Flag-US	0.28	0.31
Airplane	0.03	0.04
Car	0.18	0.20
Truck	0.08	0.10
People-Marching	0.06	0.08
Explosion_Fire	0.18	0.20
Maps	0.28	0.32
Charts	0.18	0.20

- Consistent improvement over all concepts: ~10%
- Efficient process, linear complexity $O(mn)$, 2 ms per image

30

Scalability Issue

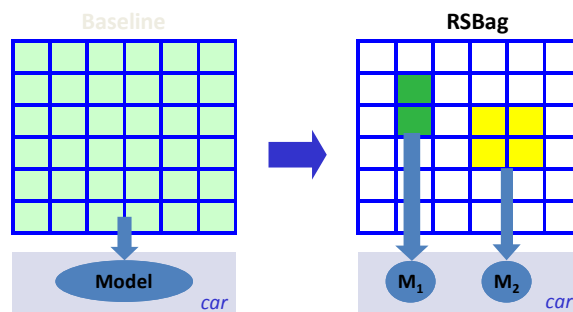
- Training time of a new concept (#pos: ~1000; #neg: ~10,000)
 - 10-20 hours (single P4 3.4GHz CPU)
- Time of processing one image/keyframe

Global feature extraction (color + texture)	0.1 seconds
Local feature extraction and quantization (3 types of keypoints; 500 visual words for each type of keypoint; on average 700+ keypoints per image)	5 seconds
Testing (classification time)	0.4 sec/concept

- Speedup
 - ~1.5 hours (IBM Random Subspace Bagging approach)
 - ~0.1 hours (cluster distributed computing over 16 nodes)

Random Subspace Bagging (RSBag)

Scaling in both data and feature space [TRECVID'07, IBM]



- RSBag: reduce redundancy of concept models in both data and feature space
 - Select multiple small bags of training examples from sampled **data and feature space**
 - Learn a **base model** on each bag of data w. any learning approaches (such as SVMs)
 - Fuse them into a **composite classifier** for each concept
 - Advantage: achieve baseline performance w. **less learning time & smaller model size**

(Slide from Rong Yan of IBM)

Blessing of the Web: Internet Vision

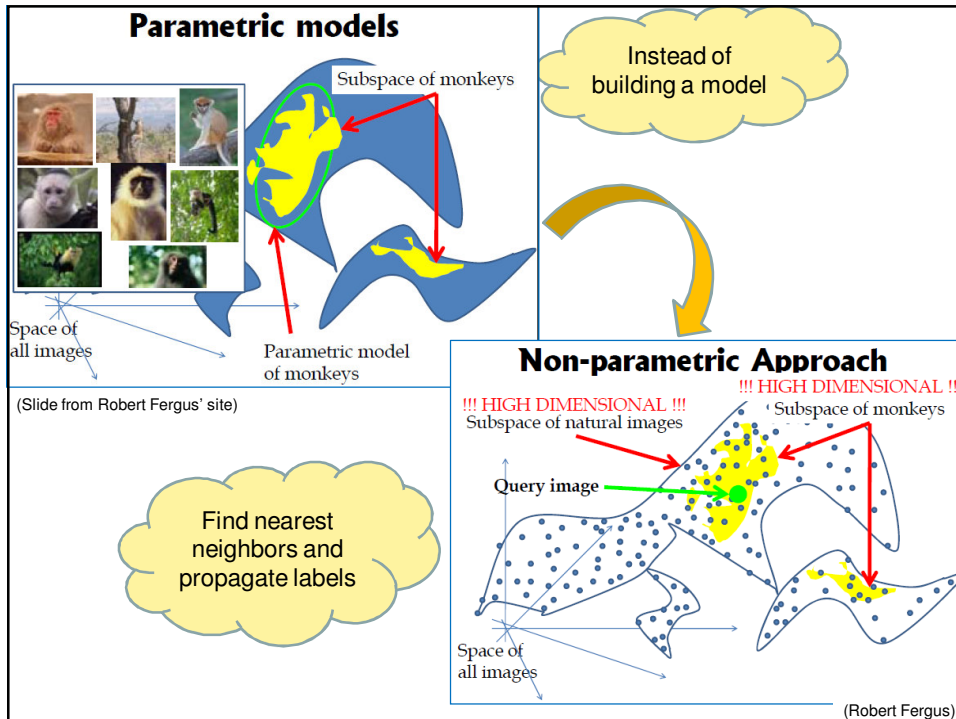
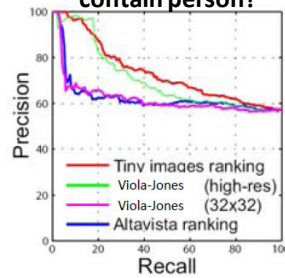
- Billions of images online as dense sampling of visual world
- Tiny images may contain enough information for recognition

1256x256



- Tiny Image (Torralba, Fergus, Freeman '07)
 - 75,378 nouns, 80M images from Internet
 - Simple pixel correlation distance
 - Outperform special detection models for some

Does the picture contain person?



Back to Fundamental Problems

- Nearest neighbor search over billions of images
 - Locality sensitive hashing preserving proximity
 - Hashing considering geometric attributes

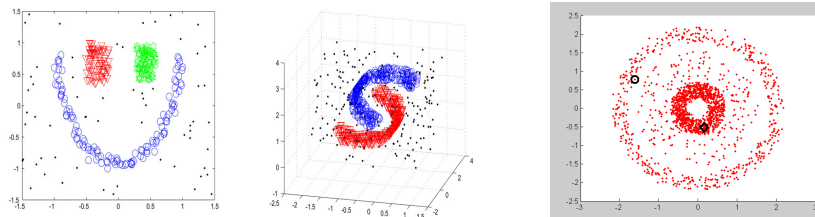
- Similarity metrics
 - Robust over content diversity due to imaging and editing variations
 - Easy to compute

digital video | multimedia lab



Beyond Nearest Neighbors

- Images are not random in the feature space
 - Explore Manifold Structures



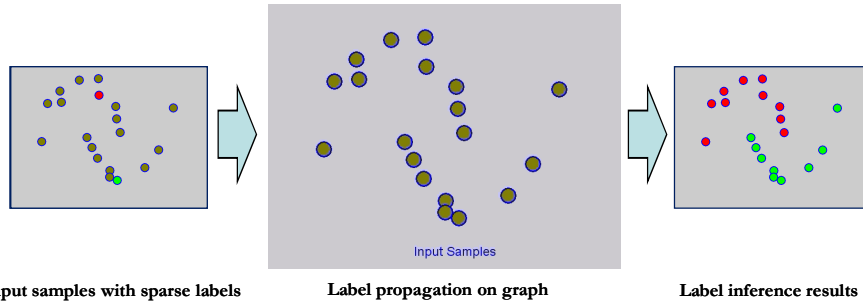
- Manifold structures may be used to aid
 - Visualization, label propagation, retrieval
 - More than nearest neighbors

Graph-based Representation and Semi-Supervised Learning

- Given a small # of labeled data and a large # of unlabeled data

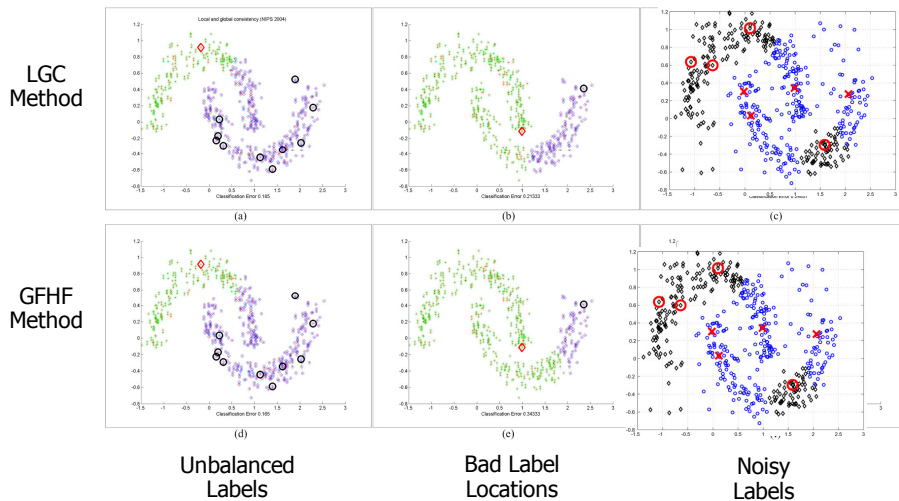
$$f^* = \min_f Q(f, y, \mathcal{G}(V, W))$$

\mathcal{G} -- graph Q -- risk function
 V -- graph node f -- classification
 W -- weight matrix y -- label matrix



Positive Negative
 Unlabeled
 Positive
 Negative

Many Challenging Issues



A hot topic in Machine Learning

- Given initial labels, Y , find classification function F over graph nodes

(Zhou, et al NIPS04)

Label smoothness Fit known labels

$$Q(F) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left\| \frac{F_{i \cdot}}{\sqrt{D_{ii}}} - \frac{F_{j \cdot}}{\sqrt{D_{jj}}} \right\|^2 + \mu \sum_{i=1}^n \|F_{i \cdot} - Y_{i \cdot}\|^2$$

- Gaussian fields & Harmonic functions (Zhu et al ICML03)

$$Q(F) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \|F_{i \cdot} - F_{j \cdot}\|^2$$

- 1) $\Delta F = 0$ on unlabeled data, where $\Delta = D - W$ is the graph Laplacian;
- 2) $F_{i \cdot} = Y_{i \cdot}$ on labeled data.

[demo](#)

39

Graph Transduction via Alternating Minimization (GTAM)

(Wang, Jebara, Chang, ICML08) (Wang and Chang, CVPR09)

-- Bivariate Optimization over Labels (Y) and Prediction (F)

$$Q(F, Y) = \frac{1}{2} \text{tr} \{ \mathbf{F}^T \mathbf{L} \mathbf{F} + \mu (\mathbf{F} - \mathbf{V} \mathbf{Y})^T (\mathbf{F} - \mathbf{V} \mathbf{Y}) \}$$

- **Prediction/Propagation Step**

- Given label (Y), propagate over graph

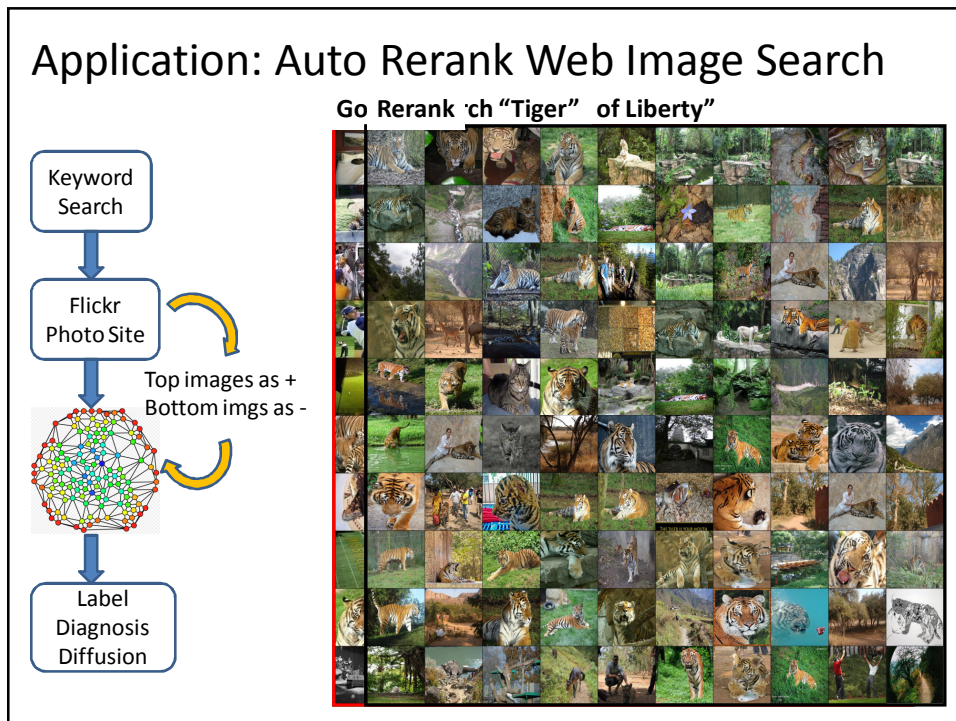
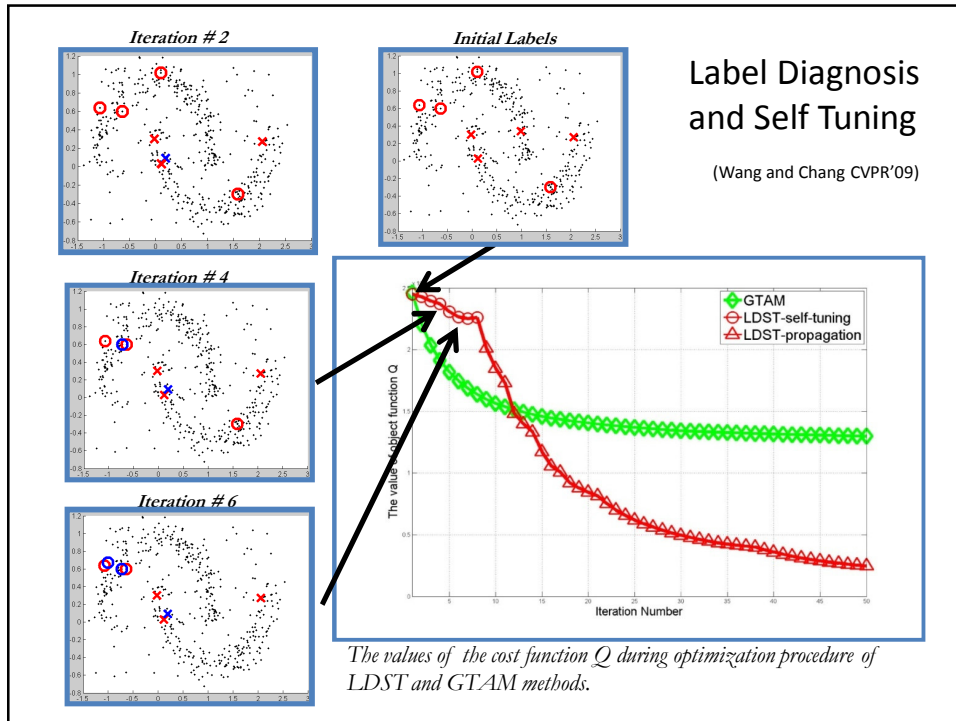
$$\frac{\partial Q}{\partial \mathbf{F}^*} = 0 \Rightarrow \mathbf{F}^* = (\mathbf{L}/\mu + \mathbf{I})^{-1} \mathbf{V} \mathbf{Y} = \mathbf{P} \mathbf{V} \mathbf{Y}$$

- **Label Selection Step**

- Iteratively add good labels or remove bad labels

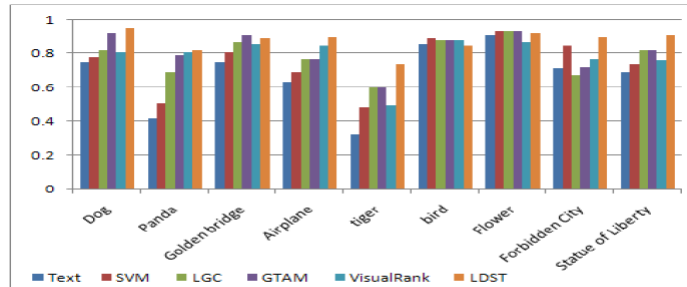
$$Q(\mathbf{Y}) = \frac{1}{2} \text{tr} \left(\mathbf{Y}^T \mathbf{V}^T \left[\mathbf{P}^T \mathbf{L} \mathbf{P} + \mu (\mathbf{P}^T - \mathbf{I})(\mathbf{P} - \mathbf{I}) \right] \mathbf{V} \mathbf{Y} \right)$$

40



Effects of Graph-based reranking

(Wang and Chang, CVPR09)



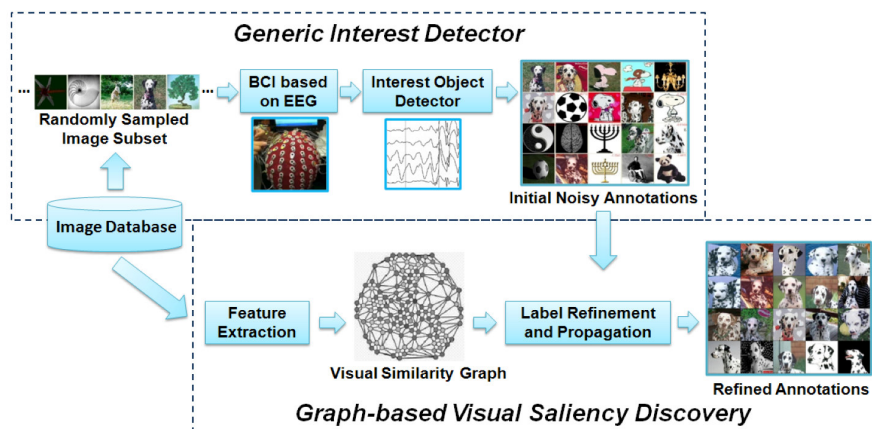
The comparison of the precision of the top 100 ranked images over different categories of images.

Method	Text	SVM	LGC	GTAM	VisualRank	LDST
Accuracy (%)	67.11	74.22	79.89	81.44	79.00	87.56

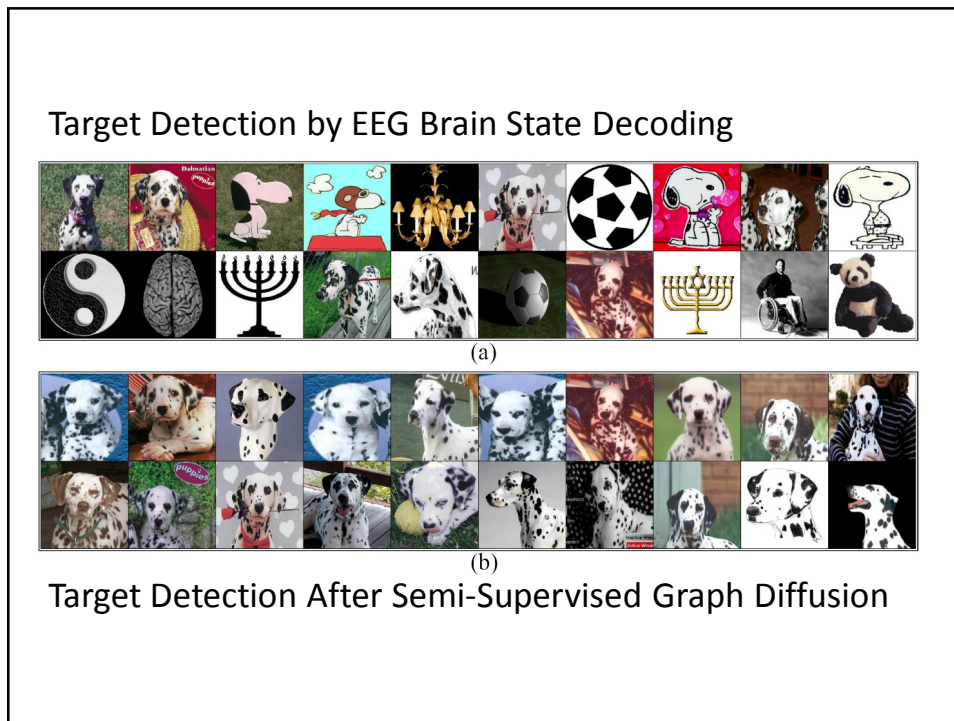
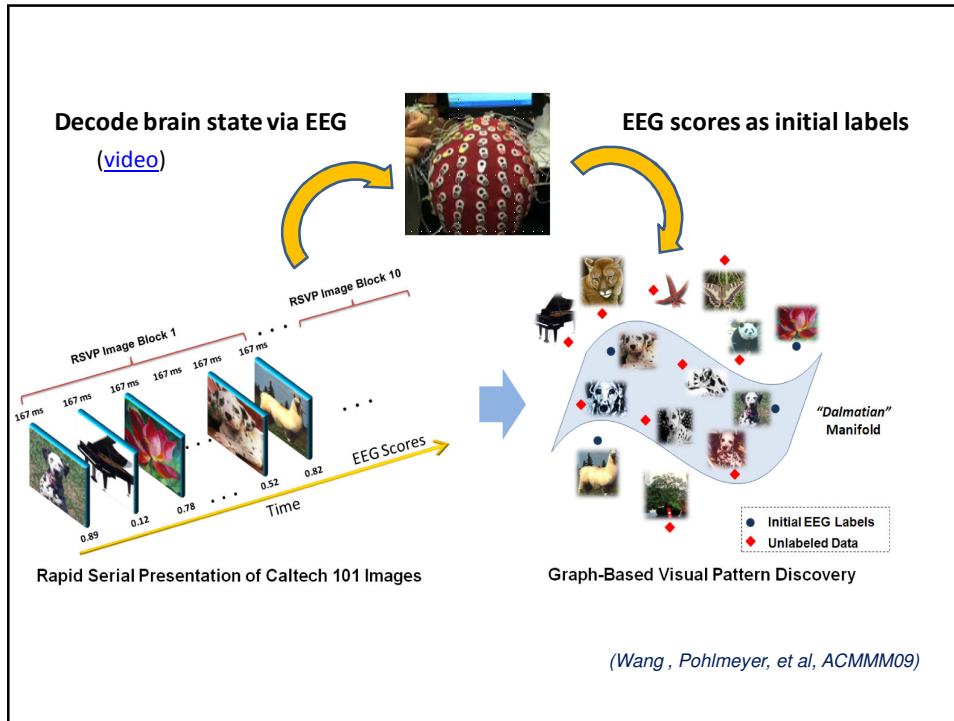
The accuracy of the top ranked Flickr images by different approaches.

Brain Machine Interface for Image Retrieval

- Decode user's interest during image browsing



(joint work with P. Sajda, E. Pohlmeier, B. Hanna, and Y.-G. Jiang)
 (Wang, Pohlmeier, et al, ACM MM09)



Target Detection by EEG Brain State Decoding



(a)



(b)

Target Detection After Graph Diffusion

Images from DigiGlobe)

Important issues

- What are useful and feasible categories?
 - A picture is worth one thousand words ...
 - Which 1000?
- How to build automatic concept classifiers
 - The semantic gap problem
- How to search over such analytics models?
 - The user gap problem

How to Search over a Large Classifier Pool?

- Audio-visual features
- Geo, social features
- User context
- Context fusion

- Rich semantic description based on content analysis

S.-F. Chang, Columbia U.

49

User Gap

- Given an assigned search topic, how to select the most appropriate concepts?

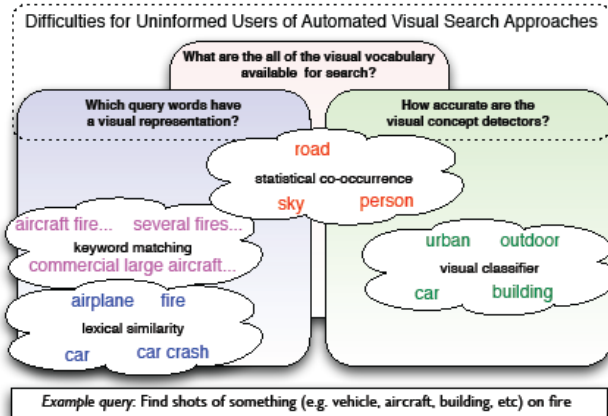
Topic: Find shots of something burning with flames visible

explosion?
smoke?

- What keywords to type in the Goggle box?

“User Gap” Challenge: User’s Inability in Forming Queries

- Difficult to choose words/concepts without in-depth knowledge of data and vocabulary

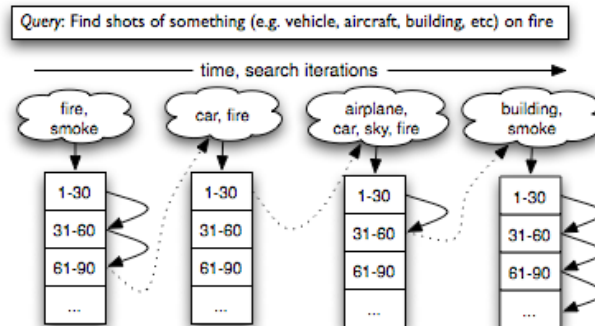


S.-F. Chang, Columbia U.

51

Pains of Frustrated Users

- Forced to take “one shot” searches, iterating queries with a trial and error approach...



S.-F. Chang, Columbia U.

52

State of the Art in Addressing the User Gap

- Scale up to a large number of classifiers
- Explore interactive tools
 - Fast relevance feedback
 - Game-style UI
 - Multi-dimensional browsing
 - Instant keyword mapping and navigation
- VideoOlympics
 - an exciting forum showcasing the video search systems



(<http://www.videolympics.org/>)

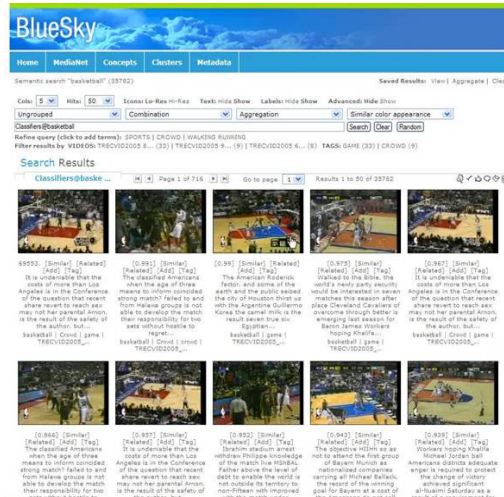
Sample Video Search Systems

- IBM IMARS
 - 1200+ concept models, multimodal retrieval
- VisionGo, National U. Singapore
 - High-level concepts, text
 - Game style UI, relevance feedback
- MediaMill, U. of Amsterdam
 - Multi-dimensional fast browsers
 - Context, similarity, high-level features
- CuZero, Columbia U.
 - Query to concept mapping, 450 concepts
 - Real-time query navigation
- Informedia, CMU
 - Multi-modal search, concept filtering

»digital video | multimedia lab»



Visual semantic classification adds important new dimensions for search (IBM Multimedia Analysis and Retrieval Search System Prototype)



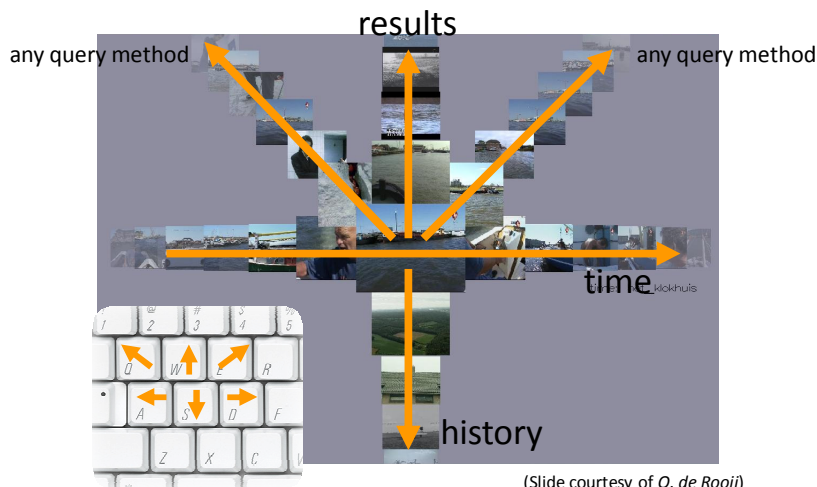
- What IMARS does?**
 - Automatic semantic classification of video contents during ingestion
 - Can be integrated with existing search systems and interfaces
 - Can be invoked as services (SOA, Web Services)
- Novel Approach:**
 - System models visual semantics of video using visual features
 - Automatically classifies visual semantics covering scenes, objects, events, people, sites of interest
 - Also support browsing and retrieval using visual similarity, near duplicates, text and temporal adjacency
- Benefits:**
 - Automates video content indexing reducing costs with manual approaches, making previously inaccessible content searchable
 - Enhances "text-only" and "speech-based" approaches
 - Provides more accurate retrieval (higher precision) by combining visual semantic classification with other search techniques
- Demos & Downloads:**
 - <http://mp7.watson.ibm.com/demos>
 - <http://www.alphaworks.ibm.com/tech/imars>

Video: "Find people eating"

(Slide courtesy of L. Xie)

The MediaMill Multi-Dimensional ForkBrowser

Video: "Find people in front of plants/trees"

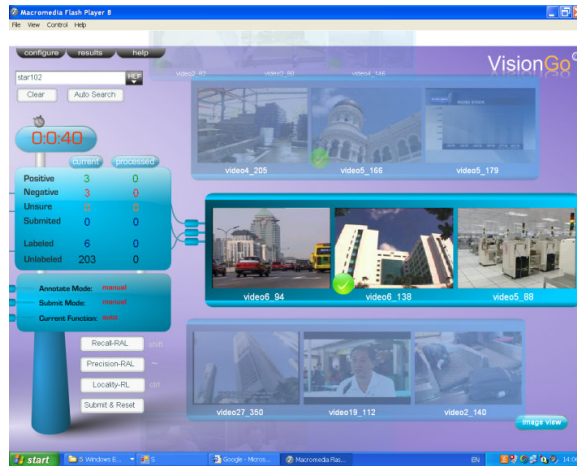


(Slide courtesy of O. de Rooij)

(O. de Rooij, C. Snoek, and M. Worring, CIVR'08)

NUS VisionGo System

- Game style UI, fast relevance feedback

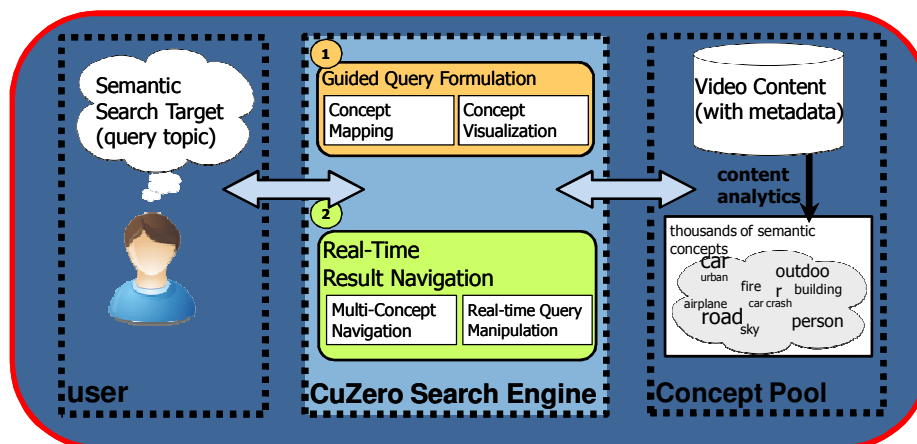


[video](#)

(Slide courtesy of T.-S. Chua) (Luan et al, CIVR'08)



Video: "Find Dakota Building near Central Park"



(Zavesky and Chang , MIR'08)

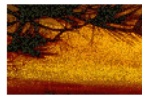
In Retrospect: 15 years after QBIC

IBM QBIC (Flickner *et al* '95)
Columbia VisualSEEk
(Smith & Chang, '96)

Query



results



Video Goggle
(Sivic & Zisserman '03)



Query object

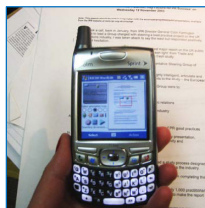


Results

59

A Killer App of CBIR

- (June 16, 2009) SnapTell acquired by A9.com, a subsidiary of Amazon.com [video](#)
- Ricoh Hot Paper, Erol *et al*, 2008



Use camera to point to a document patch and retrieve linked information

Conclusions

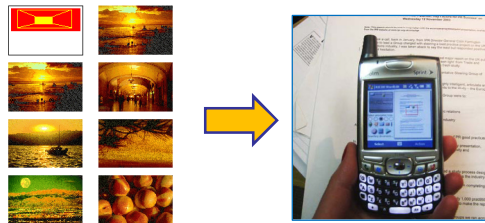
- Golden opportunity for video search research
 - Many practical problems with high impact
- Key exciting research topics
 - Billions of data and associated metadata online
 - Robust features with efficient indexing
 - Intuitive interfaces and search tools
 - Machine learning techniques (e.g., SSL)
 - Scalability in speed, source, model

»digital video | multimedia lab»



Conclusions

- 15 years from CBIR inception to mass products
 - Enabled by phone cameras as search UI



- What's the *enabler* for the next killer app?
 - We are near the 15-yr mark since
 - *Indoor-outdoor image classification, Szummer & Picard '98*
 - *On image classification: city images vs. landscapes, Vailaya, Jain, & Zhang, '98*

»digital video | multimedia lab»



References

(many papers can be found at <http://www.ee.columbia.edu/dvmm>)

- Y Rui, TS Huang, S.-F. Chang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", *Journal of Visual Communication and Image*, 1999.
- A.W. Smeulders, M. Worring, M. Santini, S. Gupta, A. Jain, "Content-Based Image Retrieval at the End of the Early Years," *T-PAMI*, 2000.
- M. Naphade, J.R. Smith, J.Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, J. Curtis, "Large-Scale Concept Ontology for Multimedia," *IEEE MultiMedia* magazine, 2006.
- L. von Ahn, L. and L. Dabbish,, "Labeling images with a computer game," *SIGCHI*, 2004.
- J Sivic, A Zisserman, "Video google: A text retrieval approach to object matching in videos," *ICCV* 2003.
- Y. G. Jiang, C. W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *ACM CIVR*, 2007.
- Y.-G. Jiang, A.Yanagawa, S.-F.Chang, C.-W.Ngo, "CU-VIREO374: Fusing Columbia374 and VIREO374 for Large Scale Semantic Concept Detection," *ADVENT Technical Report #223-2008-1 Columbia University*, August 2008.
(<http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/>)
- S. Lazebnik, C. Schmid and J. Ponce, "Beyond Bags of Features, Spatial Pyramid Matching for Recognizing Natural Scene Categories," *CVPR*, 2006.
- E. Zavesky, S.-F. Chang, "CuZero: Embracing the Frontier of Interactive Visual Search for Informed Users," *ACM MIR*, 2008.
- J. Wang, E. Pohlmeier, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, "Brain State Decoding for Rapid Image Retrieval," *ACM Multimedia Conference*, 2009.

References (2)

- M. Campbell, et al, "IBM Research TRECVID-2007 Video Retrieval System," *NIST TRECVID Workshop*, 2007.
- Y.-G. Jiang, J. Wang, S.-F. Chang, C.-W. Ngo, "Domain Adaptive Semantic Diffusion for Large Scale Context-Based Video Annotation," *ICCV*, 2009.
- A Torralba, R Fergus, Y Weiss, "Small codes and large image databases for recognition," *CVPR* 2008.
- J. Wang, Y.-G. Jiang, S.-F. Chang, "Label Diagnosis through Self Tuning for Web Image Search," *CVPR* 2009.
- J. Wang, T. Jebara, S.-F. Chang, "Graph Transduction via Alternating Minimization.," *ICML* 2008.
- J. Wang, E. Pohlmeier, B. Hanna, Y.-G. Jiang, P.Sajda, and S.-F. Chang , "Brain State Decoding for Rapid Image Retrieval ," *ACM Multimedia* 2009.
- IBM IMARS Video Search System, <http://www.alphaworks.ibm.com/tech/imars/>
- O. de Rooij, C. Snoek, and M. Worring, "Balancing Thread Based Navigation for Targeted Video Search," *ACM CIVR*, 2008.
- H," AC. Luan, et al, "Adaptive multiple feedback strategies for interactive video search," *ACM CIVR* 2008.
- B. Erol, E. Antunez, and J. Hull, "HOTPAPER: multimedia interaction with paper using mobile phones," *ACM Multimedia*, 2008.
- Szummer & Picard , "Indoor-outdoor image classification," 1998.
- Vailaya, Jain, & Zhang, "On image classification: city images vs. landscapes," 1998.