Understanding Multimedia Content *Using* Web Scale Social Media Data

Jiebo Luo

Kodak Research Labs

Kodak

Outline

- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search unannotated personal image by textual queries
- Propagating labels from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

2 Kodak

Outline

- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search unannotated personal image by textual queries
- Propagating labels from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

3

Web 2.0

- Origin of the term
  - Tim O’Reilly in 2004
- Top 100 Web 2.0 applications

What is Web 2.0?

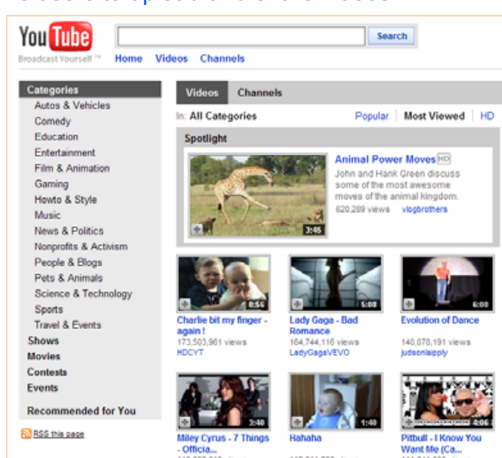
- A comparison between Web 1.0 and Web 2.0

Web 1.0 (1993-2003) Pretty much HTML pages viewed through a browser		**Web 2.0** (2004- beyond) Web pages, plus a lot of other “content” shared over the web, with more interactivity; more like an application than a “page”
“Read”	**Mode**	“Write” & Contribute
“Page”	**Primary Unit of content**	“Post / record”
“Static”	**State**	“Dynamic”
Web browser	**Viewed through…**	Browsers, RSS Readers, anything
“Client Server”	**Architecture**	“Web Services”
Web Coders	**Content Created by…**	Everyone
“Geeks”	**Domain of…**	“Mass amatuerization”

Kodak

YouTube™

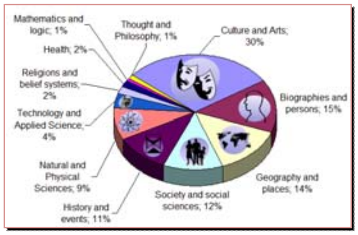
- Video sharing website which allows users to upload and share videos
- Launched in 2005
- Reached 1 billion views per day in 2009!



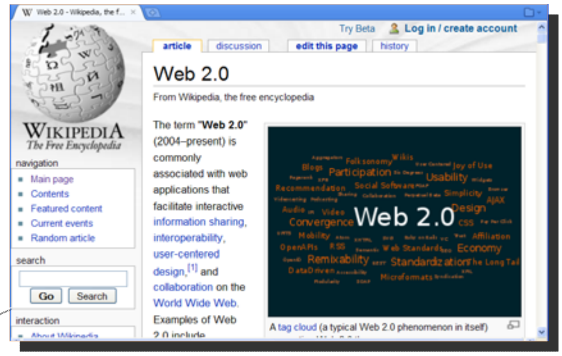
Kodak

Wikipedia

- Collaborative dictionary being edited in real time by millions of users around the world.
- Wikipedia covers a wider variety of topics than any print resource in existence!



Wikipedia content by subject



Sample Wiki article

Kodak

---

facebook

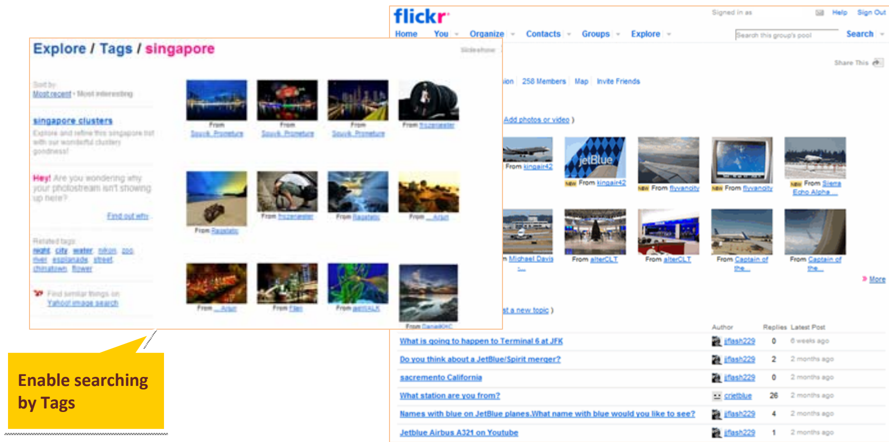
- Social networking website with over 350,000,000 users
- Basic features include networking with others and posting on a "wall" or "commenting" on pictures.

Wikipedia fan's page and Bill Gate's page

Kodak

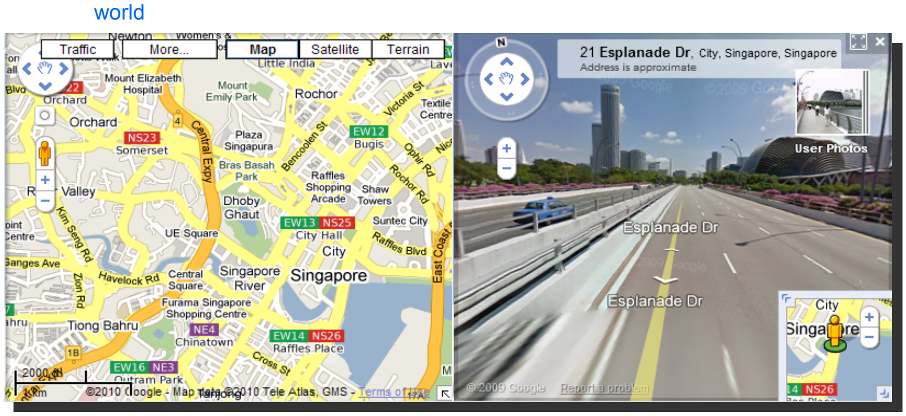
flickr

- Online photo management and sharing website
- 4+ billion users uploaded pictures, 2+ million uploads per day



Google Maps

- Easy to embed Google maps on any Web sites
- Easy to assign or visualize geotags
- **Google Street View** - provides panoramic views along many streets around the world



Characteristics of Web 2.0

- Rich user experience
- User participation
- Dynamic content
- Scalability
- Openness

What is Social Media?

A group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, which allows the creation and exchange of **user-generated content**

Andreas M. Kaplan, Michael Haenlein (2010). "Users of the world, unite! The challenges and opportunities of Social Media". Business Horizons 53 (1): 59–68.

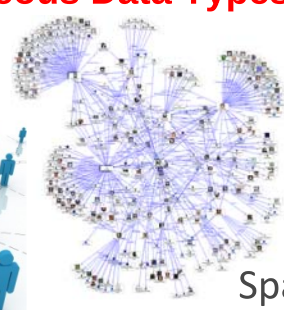
Social Media is Huge

- Goldman Sachs invested in *Facebook* at **$50 Billion** valuation in January 2011
- Private trading raised *Twitter* valuation to **$7.7 Billion** early in May 2011
- *LinkedIn* shares more than doubled in the first hours of trading as a public company, giving it a market value of more than **$10 billion**

Kodak

Heterogeneous Data Types

Link



Image/Video

Spatial-temporal Data

Text

Kodak

Links in Social Media

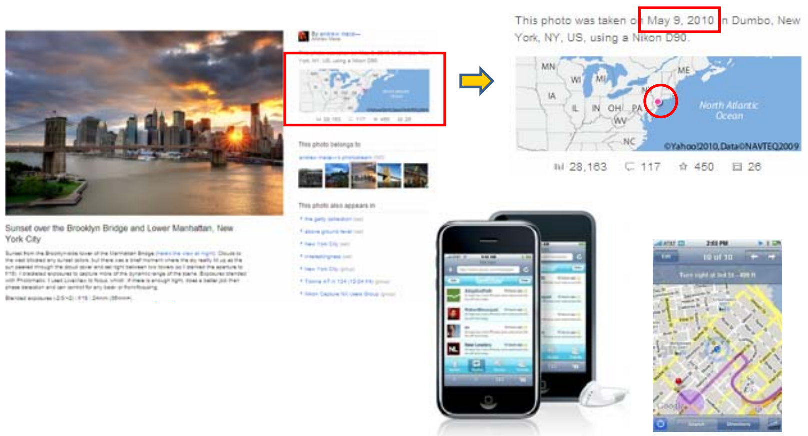
- Facebook
  - Friendship
- LinkedIn
  - Colleague/Classmate
- Twitter
  - Follower/followee

Text in Social Media

- News article
- Blog
- Micro-blogging
- User profile
- Social tagging

Spatial-Temporal Data in Social Media

- Geo-tagged and time-stamped photos



Characteristics of Social Media Data

- Multimodality
  - Visual data (and audio data)
  - Metadata and surrounding text
  - User ratings and comments
  - Links
- Noisy
  - User supplied information may not be reliable
  - Misinformation/spam too
  - Domain mismatch
- Large scale
  - Millions of data points can be easily acquired
- Time-sensitive
  - *Public* data can be collected instantly, or tracked over time

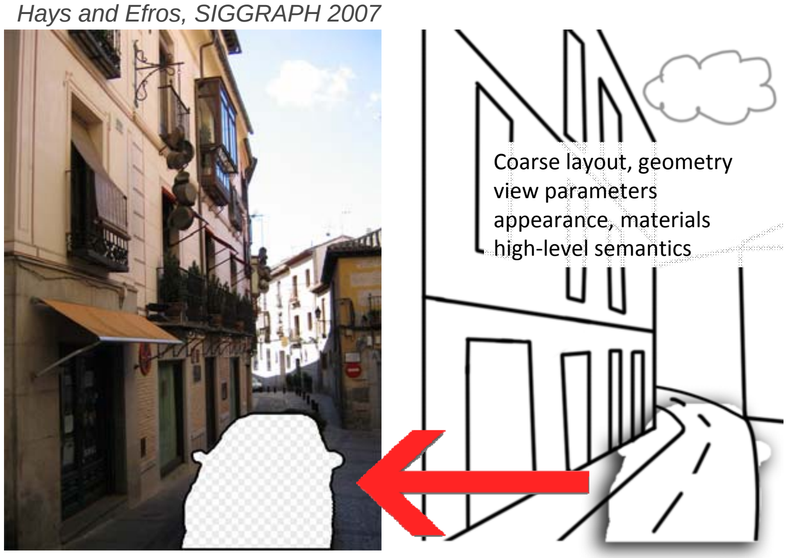
Outline

- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search unannotated personal image by textual queries
- Propagating labels from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

Kodak

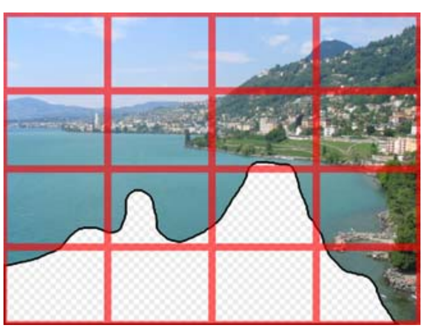
Scene Matching for Image Completion

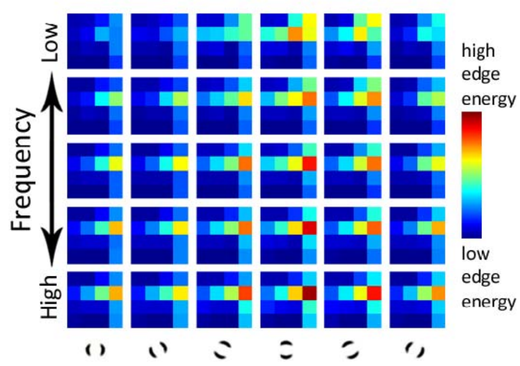
Hays and Efros, SIGGRAPH 2007



Kodak

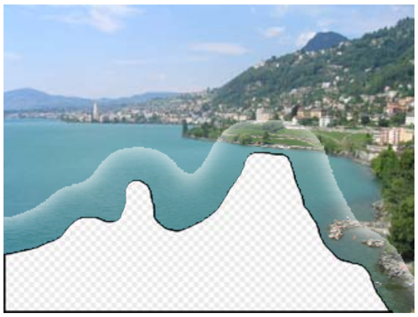
Scene Descriptor and Scene Matching





Scene Gist Descriptor
(Oliva and Torralba 2001)

Context Matching

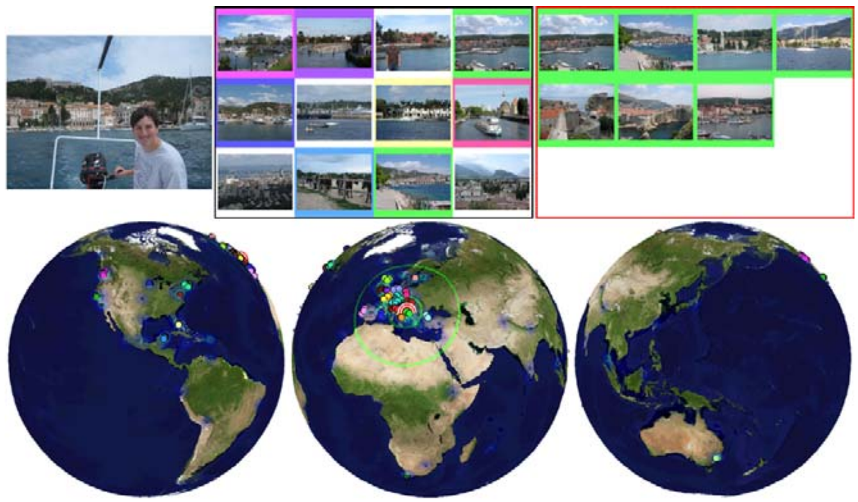


26

Kodak

Scene Matching + Lazy Learning = Geolocation

Hays and Efros, CVPR 2008



‘Beauty’ of Brute Force Data-Driven Approach

- “Data is King!”
- “Any feature will do when you have enough data”
- “KNN is all you need”
- “My dataset is bigger than yours, stupid.”

- Is machine learning becoming *irrelevant*?
  - Transfer Learning for Consumer Photo and Video Understanding

Outline

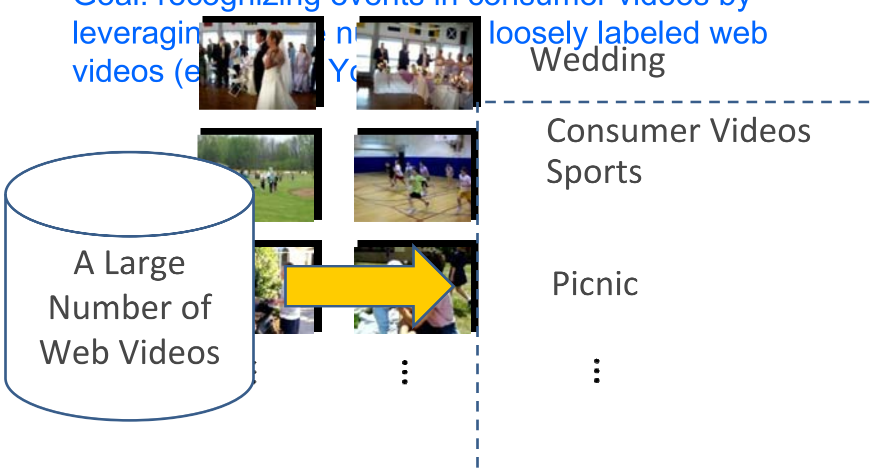
- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search annotated personal image by textual queries
- Propagating from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

Kodak

Video Event Recognition by Learning from Web Data

*Duan, Xu,. Tsang, Luo, CVPR 2010, **Best Student Paper***

- Goal: recognizing events in consumer videos by leveragin n loosely labeled web videos (e Y

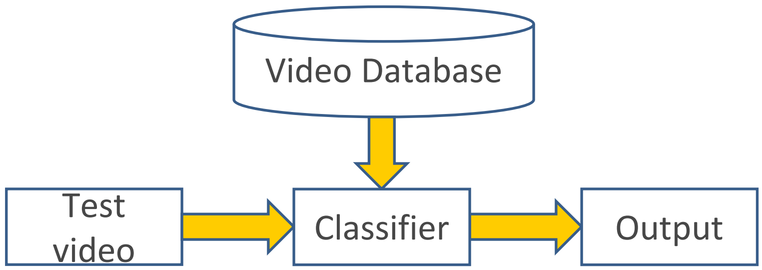


Large intra-class variability; limited labeled videos

Kodak

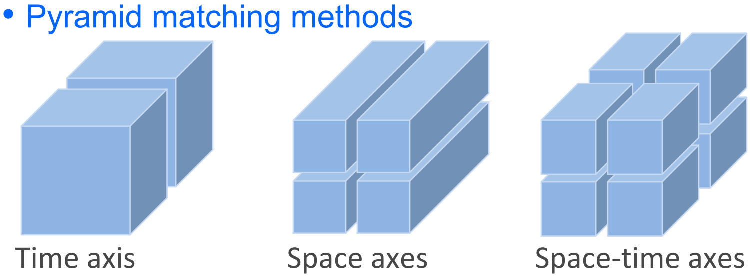
Overview

- Flowchart of the system



Similarity between Videos

- Pyramid matching methods



  - Temporally aligned pyramid matching, D. Xu and S.-F. Chang [1]
  - Unaligned space-time pyramid matching, I. Laptev [2]

Similarity between Videos

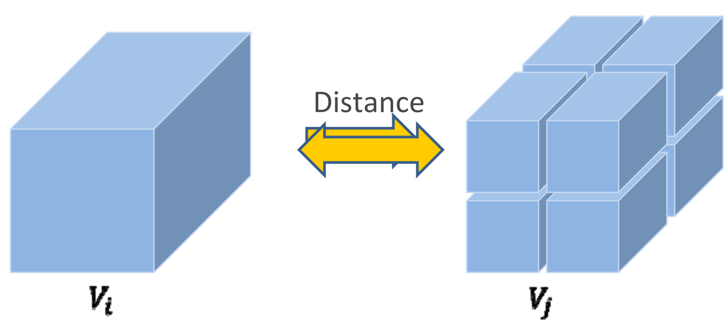
- Aligned Space-Time Pyramid Matching
  - Each video is divided into $2^l \times 2^l \times 2^l$ non-overlapped space-time volumes, where $l = 0,1$.
  - Greater variability
- Two-step approach
  - Distances between space-time volumes: solved by existing methods such as bag-of-words model, I. Laptev [2]

34

Kodak

Similarity between Videos

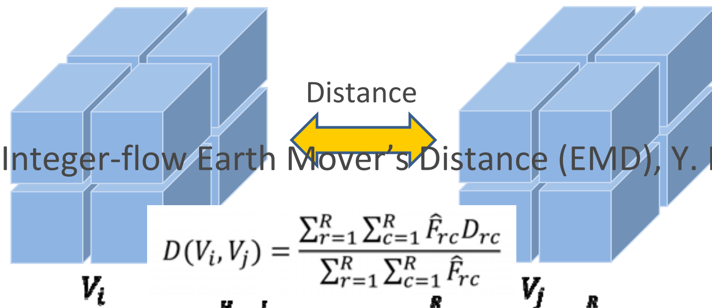
- Aligned Space-Time Pyramid Matching
  - Level 1



35

Kodak

Similarity between Videos



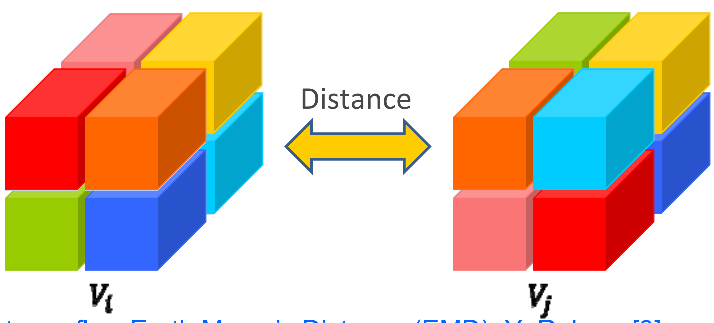
- Integer-flow Earth Mover's Distance (EMD), Y. Rubner [3]

$$D(V_i, V_j) = \frac{\sum_{r=1}^{R}\sum_{c=1}^{R}\hat{F}_{rc}D_{rc}}{\sum_{r=1}^{R}\sum_{c=1}^{R}\hat{F}_{rc}}$$

$$\hat{F}_{rc} = \arg\min_{F_{rc}\in\{0,1\}} \sum_{u=1}^{H}\sum_{v=1}^{I} F_{rc}D_{rc} \quad \text{s.t.} \quad \sum_{c=1}^{R} F_{rc} = 1, \forall r; \sum_{r=1}^{R} F_{rc} = 1, \forall c.$$

36

Similarity between Videos



- Integer-flow Earth Mover's Distance (EMD), Y. Rubner [3]

$$D(V_i, V_j) = \frac{\sum_{r=1}^{R}\sum_{c=1}^{R}\hat{F}_{rc}D_{rc}}{\sum_{r=1}^{R}\sum_{c=1}^{R}\hat{F}_{rc}}$$

$$\hat{F}_{rc} = \arg\min_{F_{rc}\in\{0,1\}} \sum_{u=1}^{H}\sum_{v=1}^{I} F_{rc}D_{rc} \quad \text{s.t.} \quad \sum_{c=1}^{R} F_{rc} = 1, \forall r; \sum_{r=1}^{R} F_{rc} = 1, \forall c.$$

37

Cross-Domain Problem

- Data distribution mismatch between consumer videos and web videos
  - Consumer videos (D^T): naturally captured
  - Web videos (D^A): often selected, usually edited
- Maximum Mean Discrepancy (MMD), K. M. Borgwardt [4]

$$\mathrm{DIST}_k(D^A, D^T) = \left\| \frac{1}{n_A}\sum_{i=1}^{n_A} \varphi(\mathbf{x}_i^A) - \frac{1}{n_T}\sum_{i=1}^{n_T} \varphi(\mathbf{x}_i^T) \right\|_{\mathcal{H}}$$

$$\Rightarrow \quad \mathrm{DIST}_k^2(D^A, D^T) = \mathrm{tr}(\mathbf{KS})$$

where $\mathbf{K} = \begin{bmatrix} \mathbf{K}^{A,A} & \mathbf{K}^{A,T} \\ \mathbf{K}^{T,A} & \mathbf{K}^{T,T} \end{bmatrix}$, $\mathbf{S} = \mathbf{s}\mathbf{s}'$ and $\mathbf{s} = [\underbrace{\frac{1}{n_A}, \ldots, \frac{1}{n_A}}_{n_A}, \underbrace{\frac{-1}{n_T}, \ldots, \frac{-1}{n_T}}_{n_T}]'$.

Kodak

Cross-Domain Problem

- Suppose there are P pre-learned classifiers f_p
- f_p is learned by SVM with the labeled training data from both domains
- Proposed target decision function

Prior information

$$f^T(\mathbf{x}) = \underbrace{\sum_{p=1}^{P} \beta_p f_p(\mathbf{x})} + \Delta f(\mathbf{x})$$

where β_p is the linear combination coefficient and $\Delta f(\mathbf{x})$ is the perturbation function.

Kodak

Cross-Domain Problem

- Motivated by Multiple Kernel Learning (MKL) (F. Bach [5]), perturbation function $\Delta f(\mathbf{x})$

$$\Delta f(\mathbf{x}) = \sum_{m=1}^{M} d_m w_m' \varphi_m(\mathbf{x}) + b$$

where $\sum_{m=1}^{M} d_m = 1, d_m \geq 0.$

- MKL: $\mathbf{K} = \sum_{m=1}^{M} d_m \mathbf{K}_m$, where $k_m(\mathbf{x}_i, \mathbf{x}_j) = \varphi_m'(\mathbf{x}_i)\varphi_m(\mathbf{x}_j)$
- MMD

$$\Omega(\mathbf{d}) \coloneqq \mathrm{DIST}_k^2(D^A, D^T) = \mathrm{tr}(\mathbf{KS}) = \mathbf{h}'\mathbf{d}$$

where $\mathbf{h} = [\mathrm{tr}(\mathbf{K}_1\mathbf{S}), \ldots, \mathrm{tr}(\mathbf{K}_M\mathbf{S})]', \mathbf{d} \in \mathcal{D} = \{\mathbf{d} \in \Re^M | \mathbf{d}'\mathbf{1} = 1, \mathbf{d} \geq 0\}.$

40

Kodak

Cross-Domain Problem

- Adaptive Multiple Kernel Learning (A-MKL)

$$\min_{\mathbf{d} \in \mathcal{D}} \ G(\mathbf{d}) = \frac{1}{2}\underbrace{\Omega^2(\mathbf{d})}_{\text{MMD}} + \theta \cdot \underbrace{J(\mathbf{d})}_{\text{Structural risk functional}}$$

where

$$J(\mathbf{d}) = \min_{\mathbf{w}_m, \boldsymbol{\beta}, b, \xi_i} \frac{1}{2}\left(\sum_{m=1}^{M} d_m \|\mathbf{w}_m\|^2 + \lambda\|\boldsymbol{\beta}\|^2\right) + C\sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i\left(\sum_{p=1}^{P} \beta_p f_p(\mathbf{x}) + \sum_{m=1}^{M} d_m w_m' \varphi_m(x) + b\right) \geq 1 - \xi_i, \xi_i \geq 0$$

41

Kodak

Cross-Domain Problem

- Feature Replication (FR), H. Daumé III [6]
  - Augment features
- Domain Transfer SVM (DTSVM), L. Duan [7]
  - No prior information
- Adaptive SVM (A-SVM), J. Yang [8]
  - $f^T(\mathbf{x}) = \sum_{p=1}^{P} \gamma_p f_p(\mathbf{x}) + \Delta f(\mathbf{x})$
  - γ_p is pre-defined
  - $\Delta f(\mathbf{x})$ is modeled by SVM

Experiments

- Data set
  - 195 consumer videos and 906 web videos collected by ourselves and from Kodak Consumer Video Benchmark Data Set [5]
  - 6 events: “wedding”, “birthday”, “picnic”, “parade”, “show” and “sports”
  - Training data: 3 videos per event from consumer videos plus all web videos
  - Test data: The rest of consumer videos

Experiments

- Two types of features
  - Space-time (ST) feature, Laptev *et al.* [1]
  - SIFT feature, Lowe [2]
- Four types of base kernels
  - Gaussian: $\exp\left(-\gamma D^2(V_i, V_j)\right)$
  - Laplacian: $\exp\left(-\sqrt{\gamma} D(V_i, V_j)\right)$
  - Inverse Square Distance: $\frac{1}{\gamma D^2(V_i, V_j)+1}$
  - Inverse Distance: $\frac{1}{\sqrt{\gamma} D(V_i, V_j)+1}$

Kodak

Experiments

- Aligned Space-Time Pyramid Matching (ASTPM) vs. Unaligned Space-Time Pyramid Matching (USTPM)
  - ASTPM is better than USTPM at Level 1

Table 1. Means and standard deviations (%) of MAPs at different levels using SVM with the default kernel parameter for SIFT features.

	Gaussian	Laplacian	ISD	ID
Level-0	41.4 ± 3.7	44.2 ± 3.8	45.0 ± 3.5	46.2 ± 4.0
Level-1 (Unaligned)	43.0 ± 2.7	47.7 ± 1.7	49.0 ± 1.6	48.2 ± 1.5
Level-1 (Aligned)	50.4 ± 3.7	53.8 ± 1.8	52.9 ± 3.6	51.0 ± 2.5

Table 2. Means and standard deviations (%) of MAPs at different levels using SVM with the default kernel parameter for ST features.

	Gaussian	Laplacian	ISD	ID
Level-0	22.2 ± 1.8	36.1 ± 0.8	22.0 ± 3.8	35.6 ± 0.7
Level-1 (Unaligned)	20.1 ± 1.0	33.9 ± 0.6	21.8 ± 0.7	33.4 ± 0.7
Level-1 (Aligned)	20.6 ± 0.7	35.8 ± 1.7	22.3 ± 1.1	35.9 ± 1.8

Kodak

Experiments

$$f^T(\mathbf{x}) = \sum_{p=1}^{P} \beta_p f_p(\mathbf{x}) + \sum_{m=1}^{M} d_m w'_m \varphi_m(x) + b$$

- 80 base kernels in total: 2 pyramid levels, 2 types of features, 5 kernel parameters and 4 types of kernels
- Average classifiers at Level l ($l = 0,1$)
  - $f_l^{SIFT}(\mathbf{x})$: 20 base classifiers learned by SVM
  - $f_l^{ST}(\mathbf{x})$: 20 base classifiers learned by SVM
  - Pre-learned classifiers $f_p(\mathbf{x})$: 4 average classifiers

Kodak

Experiments

- Comparisons of cross-domain learning methods

Table 3. Means and standard deviations (%) of MAPs of all methods over the six events in three cases.

	SVM_T	SVM_AT	FR	A-SVM	MKL	DTSVM	A-MKL
MAP-(a)	42.3 ± 5.2	53.3 ± 4.4	53.8 ± 1.8	38.7 ± 7.6	42.4 ± 2.4	48.5 ± 2.7	56.2 ± 2.7
MAP-(b)	33.4 ± 1.3	25.3 ± 0.5	29.2 ± 1.5	25.1 ± 0.7	35.2 ± 1.5	35.3 ± 1.0	37.2 ± 2.0
MAP-(c)	42.0 ± 4.9	34.6 ± 1.4	46.0 ± 1.6	31.9 ± 4.4	42.5 ± 4.6	52.7 ± 2.4	**57.9 ± 1.7**

- MKL-based methods
  - Better fuse SIFT features and ST features
  - Handle noise in the loose labels
- Relative improvements
  - SVM_T: 36.9%
  - SVM_AT: 8.6%
  - Feature Replication (FR) [6]: 7.6%
  - Adaptive SVM (A-SVM) [7]: 49.6%
  - Domain Transfer SVM (DTSVM) [8]: 9.9%

Kodak

Outline

- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search annotated personal image by textual queries
- Propagating from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

Kodak

Textual Query Based Retrieval for Consumer Photos

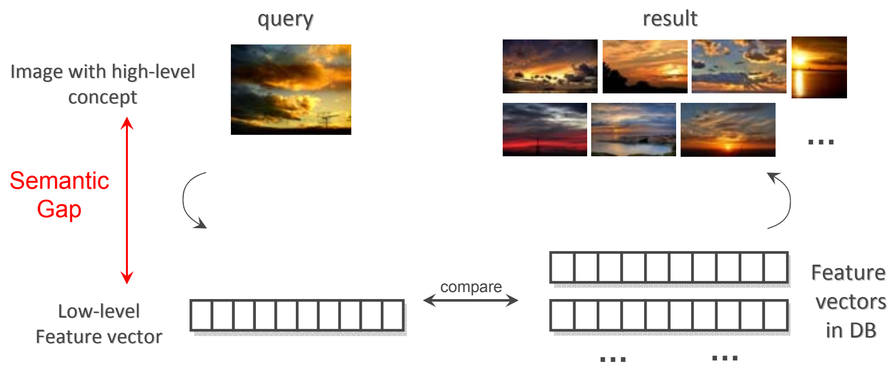
Liu, Xu,. Tsang, Luo, ACM Multimedia 2009

- Digital cameras and mobile phone cameras produce an enormous amount of personal photos
- Retrieving images from increasingly large collections of personal photos becomes a challenge

Kodak

Previous Work

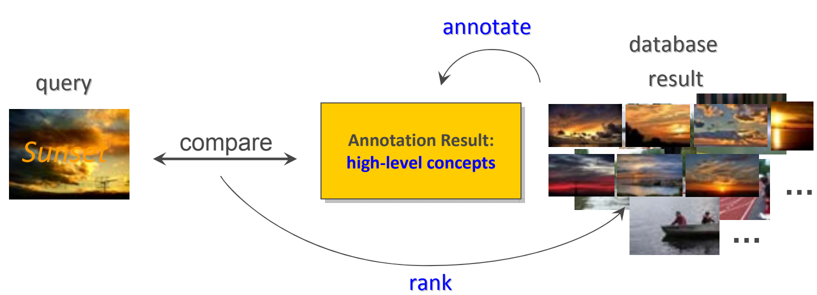
- Content-Based Image Retrieval (CBIR)
  - Users provide images as queries to retrieve personal photos.
- The paramount challenge - *semantic gap*
  - The gap between the low-level visual features and the high-level semantic concepts.



Kodak

A More Desirable Way For Consumer Photos

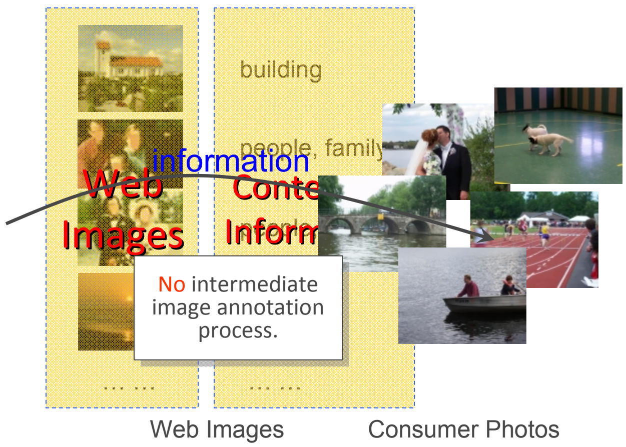
- Let the user to retrieve the desirable personal photos using textual queries.
- Image annotation is used to classify images w.r.t. high-level semantic concepts.
- An intermediate stage for textual query based image retrieval.
  - Semantic concepts are analogous to the textual terms describing document contents.



Kodak

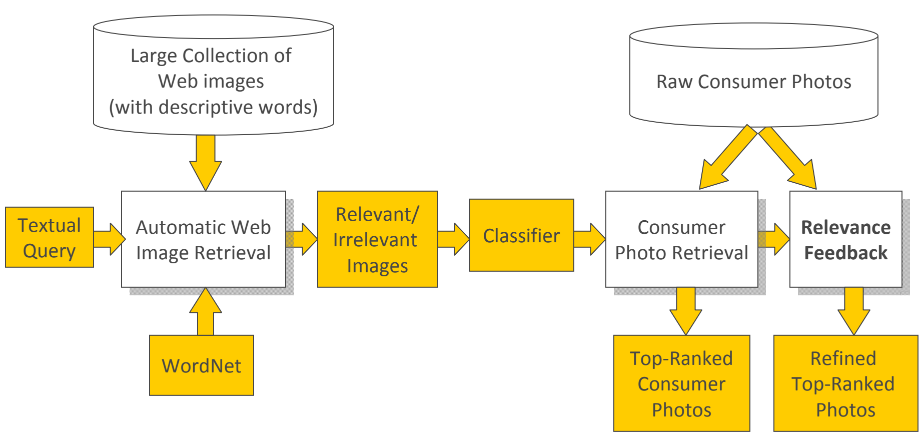
Our Goal

- A textual query based consumer photo retrieval system without any intermediate annotation stage.

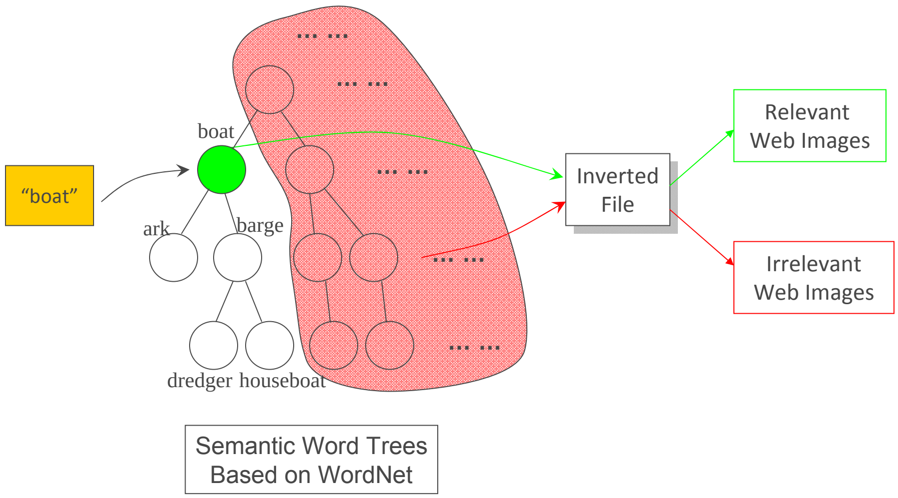


52

System Framework



Automatic Web Image Retrieval



- For any text query from the user, first search it in the semantic word trees.
- The web images containing the query word are considered as "relevant web images".
- … "irrelevant web images".

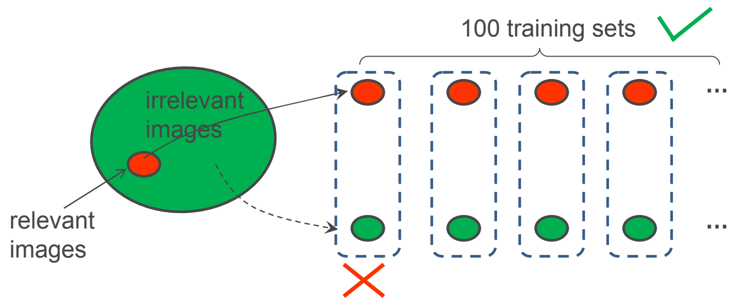
Classifier for Initial Photo Retrieval

- **Our goal**: a (quasi) **real-time** retrieval system
  - Low training cost
  - Low testing cost
  - Easy to parallelize
- We choose decision stump ensemble classifier and linear SVM.

Kodak

Asymmetric Bagging

- Imbalance: # irrelevant images >> #relevant images
  - Side effects, e.g. overfitting.
- Solution: asymmetric bagging
  - Repeat 100 times using different randomly sampled irrelevant web images.



Kodak

Dataset and Experimental Setup

- Web Image Dataset (Training dataset)
  - 1.3 million photos from photoSIG
  - Relatively professional photos
  - Relatively reliable text descriptions
- Text descriptions for web images
  - Title, portfolio, and categories accompanied with web images
  - Remove the common high-frequency words
  - Remove the rarely-used words
  - Finally, 21377 words in our vocabulary

Kodak

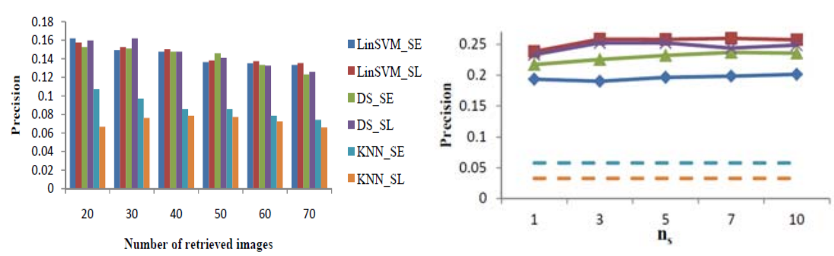
Visual Features

- Grid-Based color moment (225D)
  - Three moments of three color channels from each block of 5x5 grid.
- Edge direction histogram (73D)
  - 72 edge direction bins plus one non-edge bin.
- Wavelet texture (128D)
- We also compare early fusion and late fusion methods to fuse three different types of features.

Kodak

Retrieval without Relevance Feedback

- **kNN_S**: rank consumer photos using the average distance between the “query photo” and the 200 closest web images.
- **DS_S**: decision stump ensemble classifier
- **LinSVM_S**: linear SVM classifier



(a) Kodak (b) NUS-WIDE (top 20)

- Note: “E” is from early fusion and “L” is from late fusion

Kodak

Relevance Feedback

- The user marks n_l relevant or irrelevant consumer photos.
  - Use these labeled photos to further improve the retrieval results;
- Challenge 1: n_l is usually small;
- Challenge 2: Cross-domain Problem
  - Source classifier is trained using the images from the web domain.
  - The user-labeled photos are from the consumer domain.

Kodak

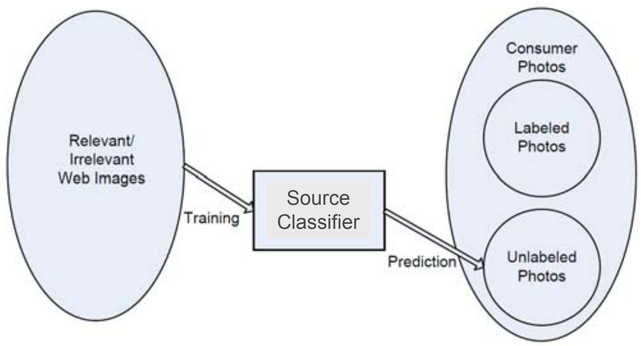
Method 1: Cross-Domain Combination of Classifiers

- Re-train classifiers with data from both domain?
  - Neither effective nor efficient
- A simple but effective method
  - Train an SVM (referred to as target classifier) using user-labeled consumer photos
  - Fuse the source classifier and the target classifier
- Referred as DS_S+SVM_T or LinSVM_S+SVM_T

Kodak

Method 2: Cross-Domain Regularized Regression

- Construct a linear regression function $f^T(x)$:
  - **For labeled photos**: $f^T(\mathbf{x}_i) \approx y_i$;
  - **For unlabeled photos**: $f^T(\mathbf{x}_i) \approx f^s(\mathbf{x}_i)$;



Kodak

User-labeled images $\mathbf{x}_1,\ldots,\mathbf{x}_l$ — $f^T(\mathbf{x})$ should be the user's label $y(\mathbf{x})$

Other images — $f^T(\mathbf{x})$ should be $f^s(\mathbf{x})$

$$\min_{f^T} \Omega(f^T) + C\left(\frac{\lambda}{2n_l}\|f_l^T - y_l^T\|^2 + \frac{1}{2n_u}\|f_u^T - f_u^s\|^2\right)$$

A regularizer controling the complexity of the target classifier $f^T(\mathbf{x})$

- This problem can be solved with least square solver. / Assume a target linear classifier $f^T(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$

Kodak

Retrieval with Relevance Feedback

- In each round, the user marks at most 1 positive and 1 negative images from the top-40 retrieved consumer photos
- Methods for comparison
  - **SVM_T**: train SVM based on the user-labeled images in the target domain
  - **A-SVM**: Adaptive SVM
  - **MR**: Manifold Ranking based relevance feedback method

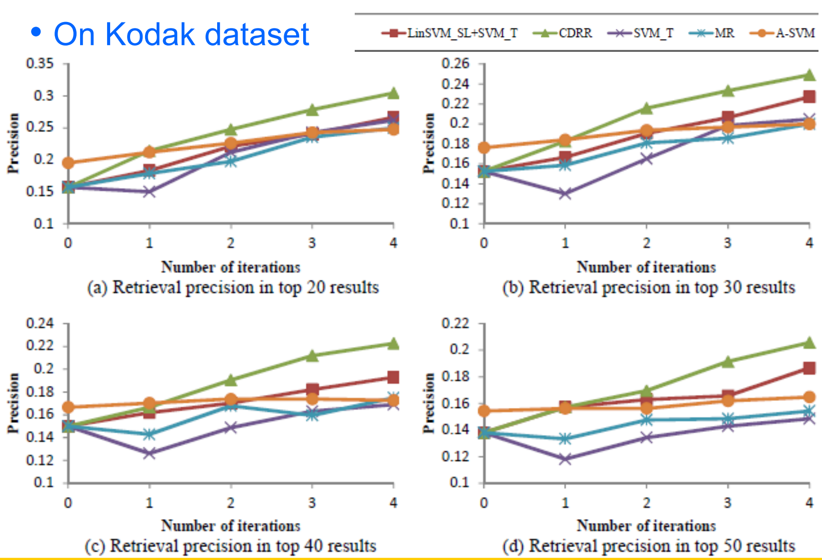
Figure 6: Top-10 initial retrieval results for query 'pool' on Kodak dataset. Incorrect results are highlighted by green boxes.

Fig. 8. Top-10 retrieval results for query "animal" on the NUS-WIDE dataset. (a) Initial results; (b) Results after 1 round of relevance feedback (one positive and one negative images are labeled in each round). Incorrect results are highlighted by green

Kodak

Retrieval with Relevance Feedback

- On Kodak dataset



Kodak

Outline

- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search unannotated personal image by textual queries
- Propagating weak labels from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

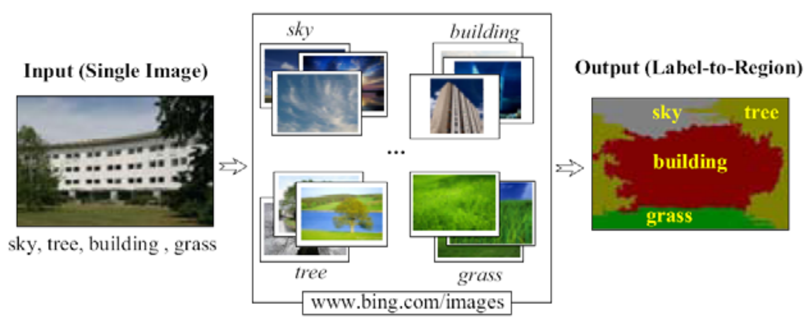
68

Kodak

Nonparametric Label-to-Region by Search

Liu, Yan, Luo, Tang, Huang, Jin, CVPR 2010

- Goal: assigning annotated labels for a given single image from the image-level to their corresponding semantic regions by utilizing the auxiliary knowledge from Internet image search with the annotated image labels as queries.

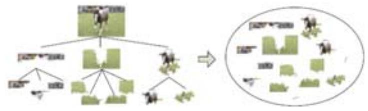


69

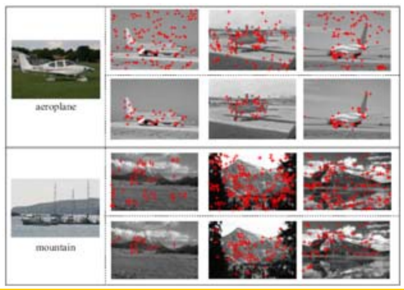
Kodak

Flowchart of the L2R-by-Search Procedure

- Bag of patches



- Label-specific feature mining by search



1st step

2nd step

1st step

2nd step

Kodak

Sparse Region Coding with Continuity-Prior

Figure 6. Example reconstruction results with different priors. The 1^{st} column shows a candidate region and its source image. In other columns, the top row shows the patches selected using sparsity prior only, and the bottom row shows the patches selected using both the sparsity and continuity-biased priors. The input image is from MSRC dataset [9] and the online images are from BING.

Kodak

L2R Assignment via Sparse Representation

Algorithm 1 . Label-to-Region Assigning via Sparse representation.

1: **Input:** selected patch-level label representation $A = [X_1, \ldots, X_{N_C}]$; feature of one candidate region $y \in R^{N_W}$;
2: Normalize the columns of A and y to have unit ℓ_2-norm; Initialize the label confidence vector $z_y \in R^{N_C}$ of y, as $z_y = 0$;
3: Solve the optimal solution $\hat{\alpha}$ according to (5);
4: For each label c annotated with the input image, calculate for y the confidence of belonging to the c-th label based on the reconstruction residual, namely, $z_y(c) \propto exp\{-\|y - A\hat{\alpha}_c\|_2\}$;
5: **Output:** $z_y \in R^{N_C}$;

Algorithm 2 . Post-processing for L2R-by-Search

1: **Input:** label confidence vector z of the input image I; label confidence vectors for all the atomic patches in image I, denoted as $\{z_i\}$, $i = 1, \ldots, N_A$, where N_A is the number of atomic patches in the BOP representation of I;
2: Set K as the number of image labels provided for I;
3: Cluster the atomic patches by grouping all the patch-level label confidence vectors $\{z_1, \ldots, z_{N_A}\}$ into K clusters, denoted as $\{O_1, \ldots, O_K\}$;
4: For each cluster $O_c \in \{O_1, \ldots, O_K\}$
4.3.1: Let z_m denote the summed label vector for each cluster, calculated as $z_m = \sum_{z_j \in O_c} z_j$;
4.3.2: Set z_m as the label vector of each atomic patch belonging to the cluster O_c;
5: Merge those patches within the same cluster to form a semantic region, and set its label as the one with the largest value in the label vector and without overlapping the label with other regions.
6: **Output:** Merged patches with semantic labels;

Kodak

Challenges and Future Directions

Kodak

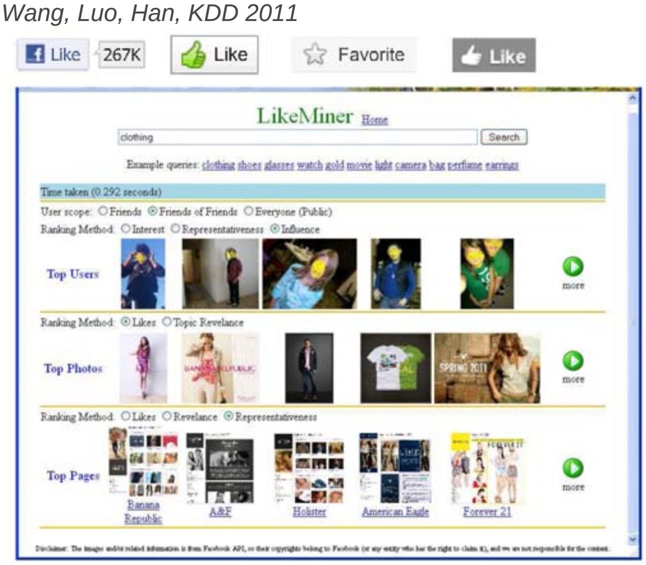
Outline

- Introduction
  - General concepts and themes
- Scene matching for graphics and vision
  - Scene completion and IM2GPS
- Learning from social media data
  - Video event recognition using few labeled examples
- Image search using social media data
  - Search unannotated personal image by textual queries
- Propagating labels from social media data
  - Label to region by search
- Conclusions
  - Challenges and future directions

Kodak

Mining Other Forms of User Interaction

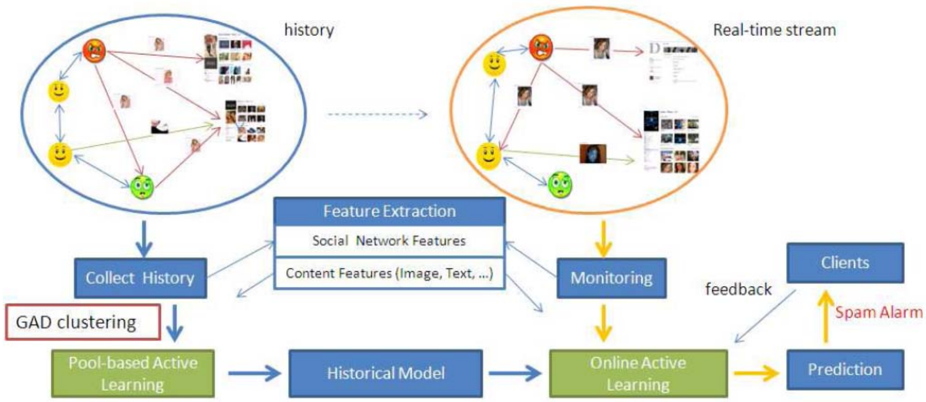
Jin, Wang, Luo, Han, KDD 2011



Kodak

Spam Detection System for Social Media Networks

Jin, Lin, Luo, Han, VLDB 2011



Kodak

Wisdom of Social Media for Prediction and Forecast

Jin, Gallagher, Cao, Luo, Han, Brave New Ideas, ACM Multimedia 2010

- Related work
  - "**The Wisdom of Crowds**" (James Surowiecki's 2005 book): *under the right conditions, a crowd of non-experts can lead to decisions that are even smarter than the experts within the crowd*
  - Actions of individual Internet users can indicate macro trends, most notably: using **Google Trends** to monitor influenza rates 1-2 weeks ahead of the CDC reports (Ginsberg et al., *Nature*, 2009)

Kodak

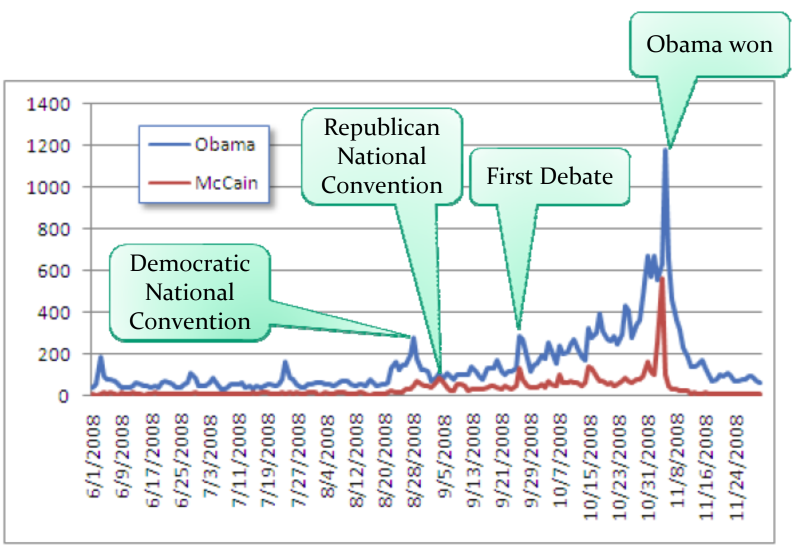
Example Flickr Images

- Flickr images of Nokia, Hillary, PS3, iPod, Obama and Mac

Kodak

2008 General Election

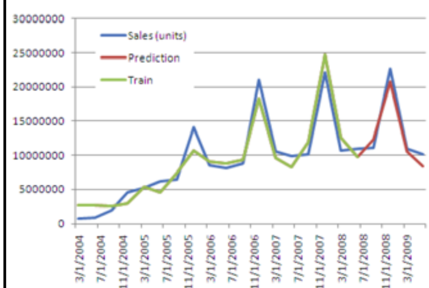
#users/day

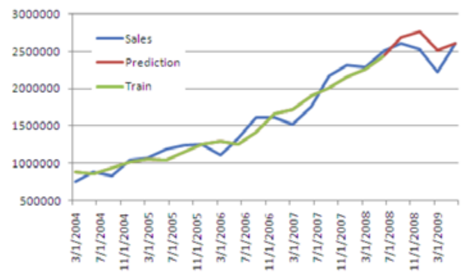


Kodak

(2) Product Sales Prediction

- Blue is the real sales, green is the history data for training, red is the predicted quarterly sales based on history data



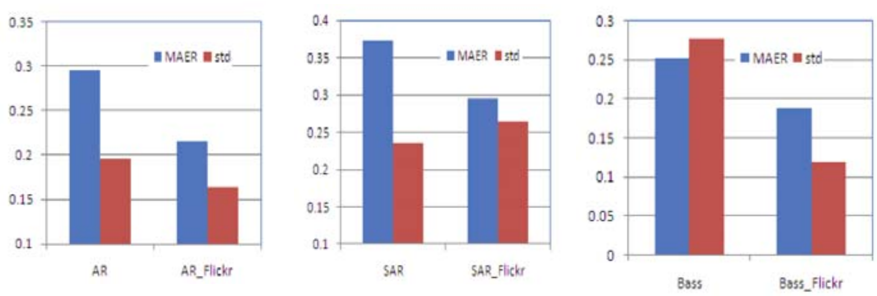


- iPod
- Mac

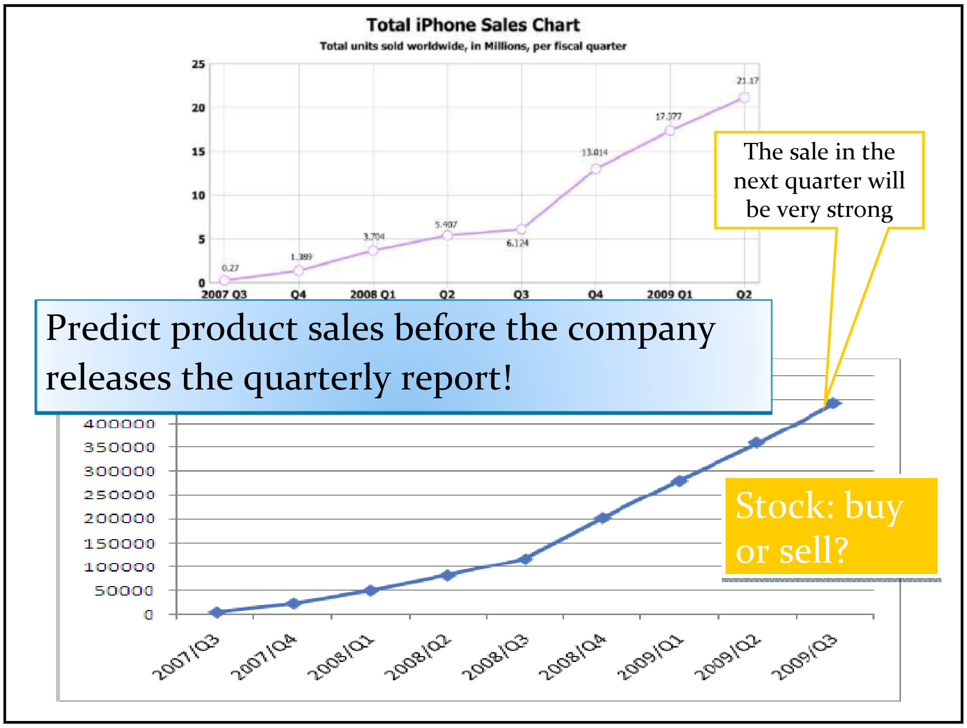
Kodak

Performance Comparison

- Average performance on popular products: iPod, Dell, Mac, Nokia, iPhone, Motorola, PS3.
- Traditional vs. Flickr-boosted models
- Measures: MAER (Mean Absolute Error Ratio) and std
- Flickr based models produce lower prediction errors and are more robust



Kodak



Conclusions

- Social media is a rich source of auxiliary data for machine learning
- The explosive growth of social media data has had a great impact on multimedia content understanding, where great progresses in image annotation, visual recognition, image search are being made
- It is important to handle multi-modality and heterogeneous features properly
- It is crucial to address machine learning issues when using large scale and noisy Web data (esp. cross-domain transfer learning)

Kodak

Challenges and Future Directions

- Challenges
  - Collecting large scale and high quality data
  - Developing efficient indexing and search technologies
  - Handling noise in the user generated content

- Future Directions
  - Exploiting of links and user sentiments
  - Data mining

Kodak

Final Remark

"… considering many votes is like averaging many samples to separate the signal from the noise. If none of the votes contains any signal … the vote is pure noise"

- J. Surowiecki. *The Wisdom of Crowds* (2005)

Kodak