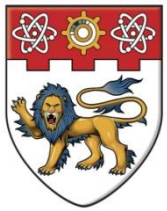


NANYANG TECHNOLOGICAL UNIVERSITY

User Profile Enhanced Geolocation Suggestion for Social Images

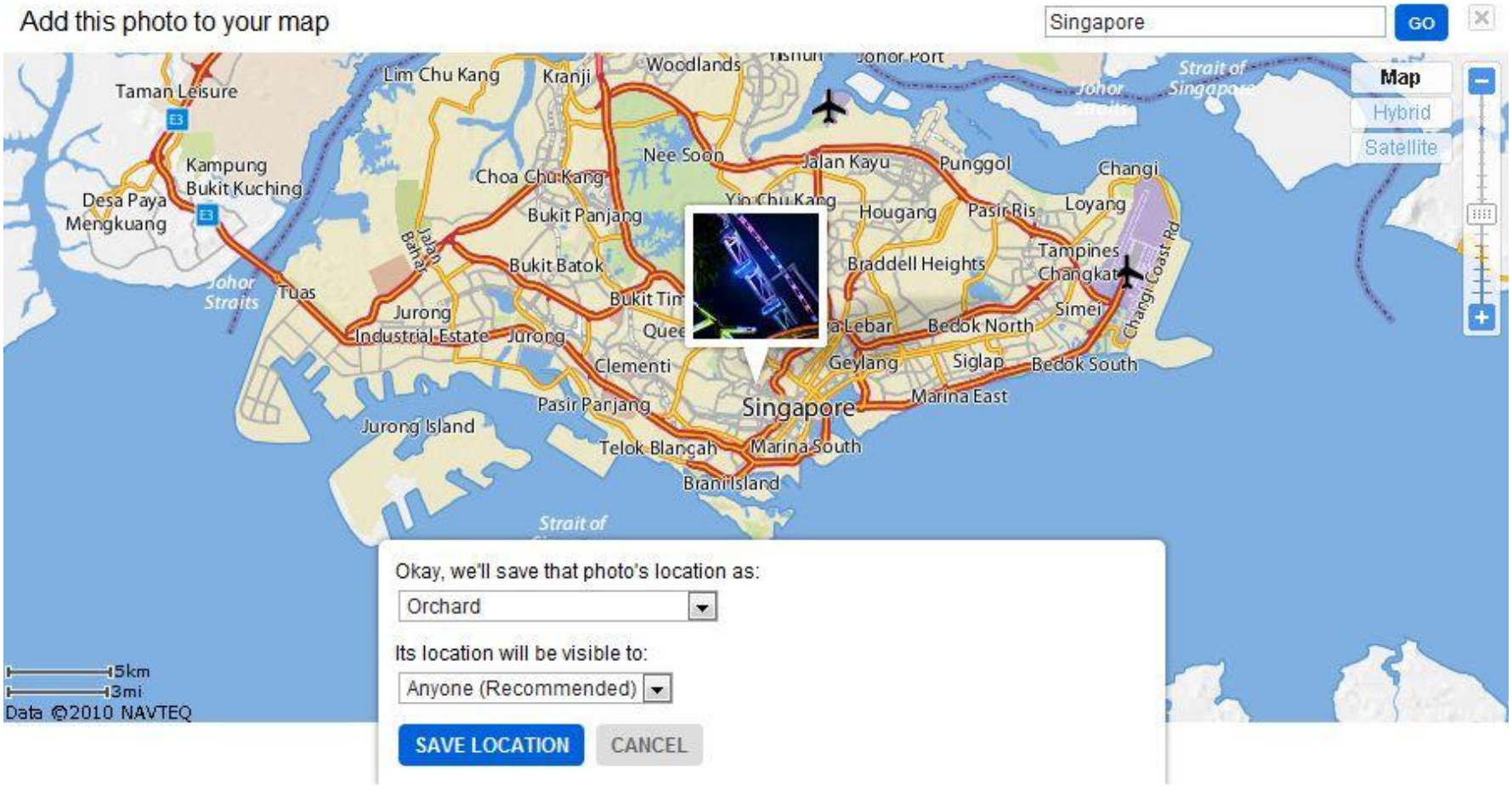
Yuan Quan
SCE & IMI
Supervisor:
Asst. Prof. Gao Cong, School of Computer Engineering
Co-supervisor:
Prof Nadia M. Thalmann, Institute for Media Innovation

24/04/2012

Outlines

- Motivation & Objective
- Related Work
- Geotagging Behavior For Social Images
- Proposed Method
- Experiments & Results
- Conclusion

Motivations & Objectives

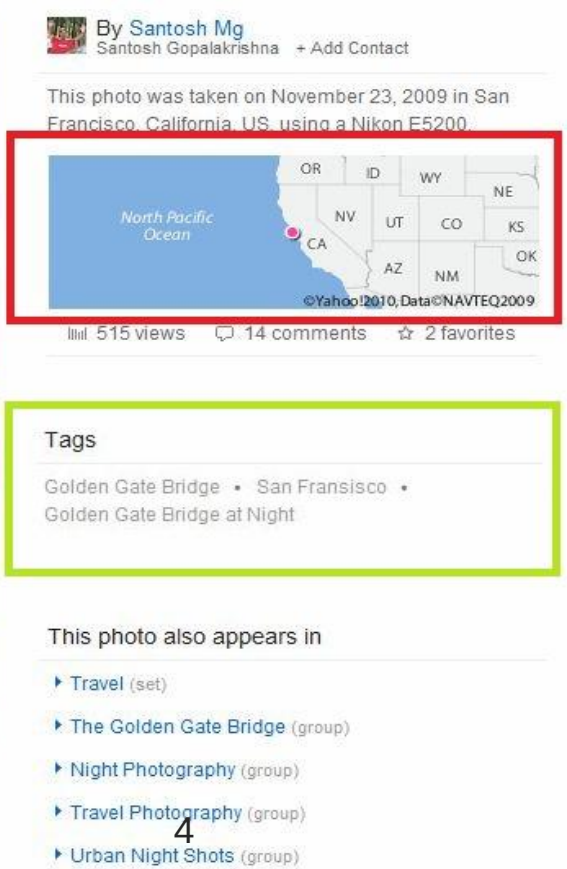


Motivations & Objectives

- Develop an effective and efficient method to suggest geolocations for images using textual tags.

Golden Gate Bridge

This is the Golden Gate Bridge in San Fransisco CA



media innovation
"innovation through collaboration"

'ANG
OGICAL
UNIVERSITY

Related Work

- Geolocation Suggestion for Images
  - IM2GPS: estimating geographic information from a single image, J. Hyas, et al, CVPR 2008
    - Visual features, time-comsuming.
  - Placing flickr photos on a map, P. Serdyukov et al. SIGIR'09
    - Textual tags, many images have no tag.
  - Mapping the world's photos, D. Crandall et al. WWW'09
    - Textual & visual features, difficult to scale
- Geolocation Suggestion for Other Contents
  - You are where you tweet: a content-based approach to geo-locating twitter users, Z. Cheng et al. CIKM'10
    - For users
  - Spatial variation in search engine queries, L. Backstrom, WWW'08
    - For queries
  - Geolocating blogs from their textual content, C. Fink et al. AAAI'09
    - For blogs

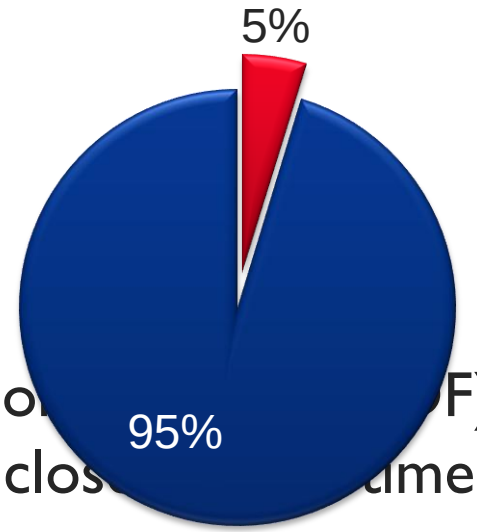
Geotagging Behavior - Dataset

- Collect a dataset of Flickr images uploaded by a set of randomly selected users

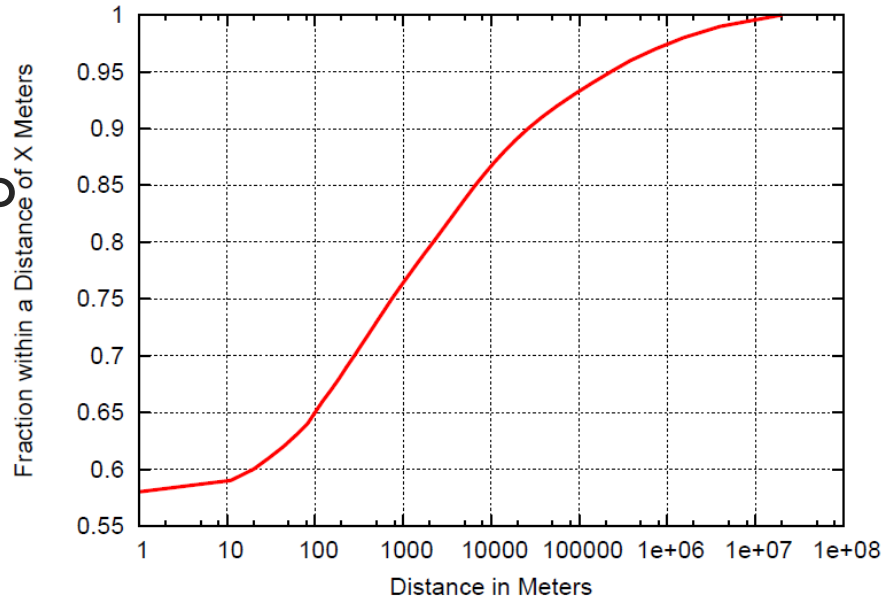
	Total	Textual Tagged	Geotagged	Both Tagged
Images	221,801,183	99,649,530 (44.9%)	17,355,876 (7.8%)	13,268,992 (5.9%)
Users	2,252,758	468,555 (20.7%)	106,289 (4.7%)	97,061 (4.3%)

- More than half of the images have no tag
- A small number of users have geotagged images

Geotagging Behavior



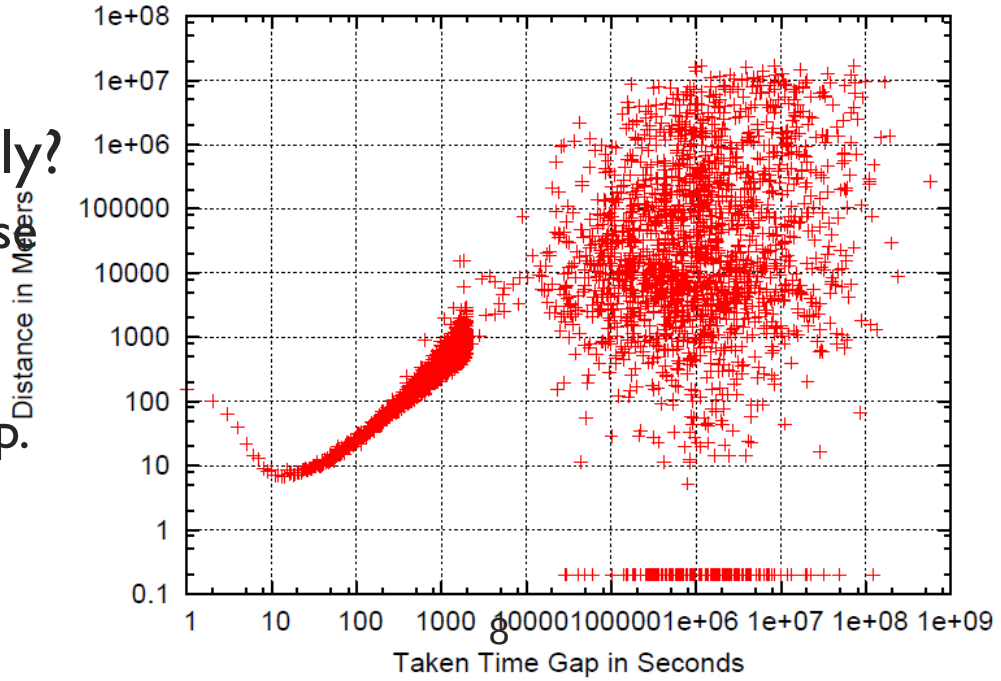
- We analyze the Cumulative Distributio F) of distance between two images with clos ime by the same user.
- Most images are close to the preceding ones.
- The geolocation of the preceding image can be used to geotag the new image.
- However…



Geotagging Behavior

- Turn to textual tags
- Generate user profile using the tags of historical images
  - Close locations share similar tags
  - Most images are spatially close to preceding image
  - Query image is likely to be taken within locations represented by user profile.

- Treat historical tags equally?
  - Close in taken time →Close in space
  - Analyze relation between distance and taken time gap.
  - Tags of recent images are more valuable.



Proposed Method
—Location Representation

- Place a grid over the world map, and define a location as a cell on it.
- Different granularity: 1, 10, 100km over latitude.
- Map the photos to the corresponding cells.

Proposed Method —Using Image Tags

- k-Nearest Neighbors Method (kNN)
  - Each image is represented by a vector on tags. Predict location of an image using k nearest neighbors.
- Language Model Based Method (LM)
  - The textual tags of photos are used to derive a language model that represents a location.

$$P(l \mid T) = \frac{P(T \mid l)P(l)}{P(T)} \propto P(T \mid l) = \prod_{t \in T} P(t \mid l) \qquad P(t \mid l) = \frac{|l|}{|l| + \lambda} P(t \mid l)_{ML} + \frac{\lambda}{|l| + \lambda} P(t \mid G)_{ML}$$

- Naïve Bayes Method (NB, NB+RT, NB+CT)
  - Expand a query image's tag set by adding with its owner's history tags until |T|=K.

$$P(l \mid T) = \frac{P(T \mid l)P(l)}{P(T)} \propto P(T \mid l)P(l) = \prod_{t \in T} P(t \mid l)P(l)$$

Proposed Method
—Example of Language Model

$$P(l \mid T) = \frac{P(T \mid l)P(l)}{P(T)} \propto P(T \mid l) = \prod_{t \in T} P(t \mid l) \quad P(t \mid l) = \frac{|l|}{|l|+\lambda} P(t \mid l)_{ML} + \frac{\lambda}{|l|+\lambda} P(t \mid G)_{ML}$$

Term Frequency

tag	L	G
NTU	3	3
SG	1	2
UK	0	3

Maximum Likelihood

tag	L	G
NTU	0.75	0.375
SG	0.25	0.25
UK	0	0.375

$\lambda = 2$

Smoothed

tag	L
NTU	0.5625
SG	0.25
UK	0.1875

Testing Image

tag	L
NTU	2
SG	1
UK	0

$$P(l \mid T) = 0.5625^2 \times 0.25 \times 0.1875 \approx 0.015$$

Proposed Method —Using User Profile

- Model User Profile (P, WP)
  - Treat the set of entire historical tags used by a user as her profile, and build a language model U to model it.
  - Two choices:
    - Equal weight for every tag (P)
    - Different weight for different tag based on taken time of an image (WP)
- Measure similarity between the user profile and each location
  - Use KL divergence

$$D_{KL}(U \parallel L) = \sum_i U(i) \log \frac{U(i)}{L(i)}$$

Proposed Method
—Combining Image Tags and User Profile

- Unified Framework (NB+P, NB+WP)
  - Combine the evidence of image tags and user profile

$$S = (1 - b^{-n}) S_{tag} + b^{-n} S_{user}$$

$$S_{user} = S_{D_{KL}(U\|L)} = 1 - \frac{D_{KL}(U \parallel L) - \min D_{KL}}{\max D_{KL} - \min D_{KL}}$$

$$S_{tag} = S_{P(l|T)} = \frac{P(l \mid T) - \min P}{\max P - \min P}$$

Experiments

- Dataset:
  - Training set: 3,491,429 images uploaded before March 1st 2011
  - Tuning set: 10,000 images uploaded after March 1st 2011
  - Testing set: 10,000 images uploaded after March 1st 2011
- Metric:
  - Accuracy (Acc)
  - Fraction of images predicted correctly within k-cell distance(Acc@k)
  - Fraction of images predicted correcly among top-k locations (Top-k)

Experiment Results

	Acc	Acc@1	Acc@2	Acc@3	Top-2	Top-3	Top-4
			$1KM$				
kNN	0.0521	0.0848	0.1054	0.1194	0.0733	0.0852	0.0922
LM	0.0576	0.0958	0.1173	0.1297	0.0815	0.0935	0.1029
NB	0.0598	0.0989	0.1217	0.1354	0.0811	0.0947	01033
NB+RT	0.0572	0.0946	0.1197	0.1351	0.0771	0.0906	0.0995
NB+CT	0.0634	0.1065	**0.1352**	0.1521	**0.0866**	**0.1011**	0.1123
P	0.0121	0.0270	0.0438	0.0599	0.0187	0.0236	0.0286
WP	0.0389	0.0737	0.0960	0.1178	0.565	0.696	0.775
NB+P	0.0577	0.0968	0.1230	0.1430	0.0800	0.0934	0.1046
NB+WP	**0.0637 (+22.3%)**	**0.1067**	0.1334	**0.1546**	0.0858	0.1007	**0.1138**
			$10KM$				
kNN	0.1420	0.1809	0.1953	0.2032	0.1774	0.1904	0.1997
LM	0.1550	0.2014	0.2187	0.2263	0.1996	0.2135	0.2228
NB	0.1581	0.2058	0.2245	0.2312	0.2010	0.2218	0.2368
NB+RT	0.1602	0.1941	0.1985	0.2005	0.2084	0.2311	0.2447
NB+CT	0.1794	0.2145	0.2193	0.2212	**0.2323**	**0.2575**	**0.2722**
P	0.0710	0.1214	0.1557	0.1697	0.1004	0.1164	0.1280
WP	0.1309	0.1982	0.2328	0.2493	0.1734	0.1996	0.2132
NB+P	0.1685	0.2301	0.2589	0.2717	0.2138	0.2365	0.2503
NB+WP	**0.1811 (+27.5%)**	**0.2442**	**0.2721**	**0.2851**	0.2285	0.2515	0.2669
			$100KM$				
kNN	0.2101	0.2362	0.2510	0.2609	0.2437	0.2565	0.2634
LM	0.2442	0.2766	0.2976	0.3081	0.2758	0.2898	0.3006
NB	0.2463	0.2818	0.3056	0.3231	0.2977	0.3300	0.3490
NB+RT	0.2839	0.3349	0.3644	0.3849	0.3399	0.3685	0.3854
NB+CT	0.3056	0.3585	0.3912	0.4151	0.3596	0.3870	0.4068
P	0.2196	0.2828	0.3209	0.3456	0.2729	0.3044	0.3311
WP	0.2978	0.3643	0.4020	0.4283	0.3560	0.3893	0.4185
NB+P	0.3023	0.3633	0.3973	0.4189	0.3640	0.3967	0.4214
NB+WP	**0.3277 (+56.0%)**	**0.3899**	**0.4254**	**0.4491**	**0.3856**	**0.4193**	**0.4439**

- Combining image tags and user profile together is able to achieve much better result
- The taken time stamps of historical images play an important role in user profile

Conclusion

- We focus on the problem of suggesting geolocations for social images.
- For the first time we analyze the user uploading patterns, geotagging behaviors and the relationship between taken-time gap of two images and their distance.
- We proposed a unified framework for geolocation suggestion.
- Experimental results show the effectiveness of the proposed method.

Q & A

Thanks!
Any questions?