Mapping the World's Photos: Collective Perception

Daniel Huttenlocher

Joint work Lars Backstrom, David Crandall, Jon Kleinberg and Yunpeng Li

Cornell University
Faculty of Computing and Information Science

Representing the World Around Us

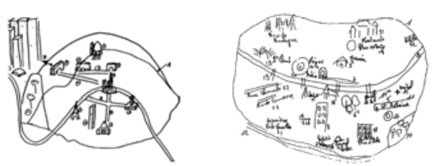
A city consists of streets, squares and buildings that exist in objective, geographic space. But there is also a psychological representation of the city that each inhabitant carries around in his head.

The capacity to form such a representation of the overall structure of the city depends not only on the individual but on the city as well, and the degree to which it is imagible. A highly imagible city does not mean that every point is equally identifiable. Rather, there are clearly identifiable focal points throughout the city which are interconnected and thus form a coherent picture.

[Milgram72]

Collective Perception and Mental Maps

A city is a social fact. We would all agree to that. But we need to add an important corollary: the perception of a city is also a social fact, and as such needs to be studied in its collective as well as its individual aspect. It is not only what *exists* but what is *highlighted* by the community that acquires salience in the mind of the person. A city is as much a collective representation as it is an assemblage of streets, squares, and buildings.



[Milgram76]

Experiments: Hand-Drawn Maps

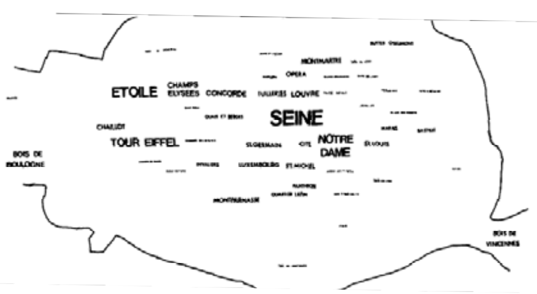
- 218 subjects each draw map of Paris
- Total of 4132 elements in maps
- Hand code elements
- Tabulate commonly occurring ones

[Milgram76]

TABLE 8.1 THE FIFTY ELEMENTS MOST FREQUENTLY INCLUDED IN THE HANDDRAWN MAPS OF PARIS

Rank	Name of element	Percent of maps in which this element appears	Rank	Name of element	Percent of maps in which this element appears
1.	Seine	84.3	26.	Bastille	22.1
2.	Limites de Paris	81.5	27.	Quartier Latin	20.7
3.	Etoile, Arc de Triomphe	61.9	28.	Panthéon	20.7
4.	Notre Dame	55.5	29.	Place des Vosges	18.4
5.	Tour Eiffel	54.6	30.	Gare de Lyon	18.4
6.	Bois de Boulogne	49.1	31.	Champ de Mars	17.9
7.	Louvre	45.4	32.	Madeleine	17.9
8.	Concorde	45.4	33.	Parc Monceau	17.0
9.	Champs Elysées	40.4	34.	Parc de Montsouris	16.6
10.	Jardin du Luxembourg	38.5	35.	Gare St. Lazare	16.6
11.	Bois de Vincennes	38.1	36.	Jardin des Plantes	16.1
12.	Gare et Tour Montp.	35.3	37.	Gare de l'Est	15.6
13.	Île de la Cité	33.9	38.	Palais Royale	15.2
14.	Tuileries	33.5	39.	Gare du Nord	14.7
15.	Butte Montmartre	32.1	40.	Place de la République	14.3
16.	Chaillot, Trocadero	32.1	41.	Gare d'Austerlitz	13.8
17.	Île de St. Louis	31.7	42.	Père Lachaise	12.9
18.	St. Germain	31.2	43.	Porte, Place d'Italie	12.4
19.	Opéra	30.7	44.	Place de la Nation	12.0
20.	Boulevard St. Michel	30.1	45.	Chambre des Députés	11.5
21.	Invalides	29.8	46.	École Militaire	11.5
22.	Marais	26.2	47.	Les Halles	10.1
23.	Buttes Chaumont	24.4	48.	Grand, Petit Palais	9.7
24.	Sacre Coeur	23.4	49.	La Défense	9.7
25.	Quais, Berges	22.5	50.	Grands Boulevards	9.2

Map of Top Ranked Elements



[Milgram76]

Collective Perception in Internet Age

- Billions of publicly available photos online
  - Most with tags – only somewhat descriptive
  - Hundreds of millions with geo location
    - Will grow quickly with new devices
- Large-scale data about the world – extract shared mental maps
  - From scale of a single city to the globe
  - From hundreds of people to hundreds of thousands or millions
  - From explicit experimental settings to everyday activities

Photo Sharing Web Sites

- Rich metadata
  - Tags, geo-location, photographer
  - Camera data: time/date stamp, focal length, shutter speed, camera model, …
  - Relationships between users and photos: favorites, contact lists, …

flickr facebook Picasa Web Albums photobucket Kodak Gallery FOTOLOG

Analogy to Web Search

- Techniques for organizing collections of Web documents exploit both link structure and content analysis [Page99] [Kleinberg99]
  - Collective understanding, "votes" on importance
- Photo sharing sites also have connective structure provided by many people
  - Photos taken nearby in space (and time)
  - Stream of photos by given photographer
  - Contacts, friendships between photographers
- Combine with text and image content

Structure in Photo Collections

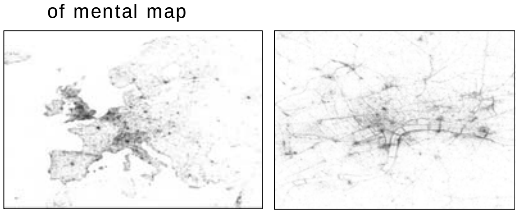
- Clustering/modeling using geo-tags, text tags, image features, social network [Ahern07] [Golder08] [Jaffe06] [Kennedy08] [Lerman07] [Marlow06] [Quack08]
- Building and annotating maps [Grabler08] [Kennedy08] [Google Sketchup3d]
- Geometric structure [Schaffalitzky02] [Snavely06,07] [Microsoft Photosynth]

Geo Tagging

- Photos tagged with geographic info – latitude and longitude
  - GUI, GPS and radio
- Photos taken nearby often related but far from guaranteed – e.g., Independence Hall

Latent Structure in Geo Tags

- Restrict number of photos per photographer
- Spatial distribution reflects relatedness
  - Use to find and characterize important elements of mental map



Outline of Remainder of Talk

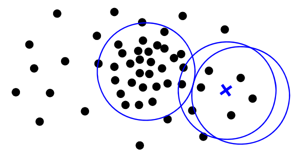
- Automatically finding and describing important places – "compact structure"
  - Geolocation, text and image content
- Application: automatically generated maps
  - "Collective perception"
  - Highlight and characterize important elements
- Modeling locations and classifying spatial location of unlabeled images
  - Many locations, large training and test sets, temporal photostream
- Summary and discussion

Finding Important Locations

- Natural scales of interest ("octaves")
  - 100km city/metro area, 10km town, 1km neighborhood, 100m landmark
- Want to discover locations automatically at one or more spatial scales
  - Think of geo-tags as samples from unknown distribution whose modes we want to estimate at certain scales
- Mean-shift procedure for mode estimation
  - Fixed-scale clustering, rather than k-means or agglomerative methods

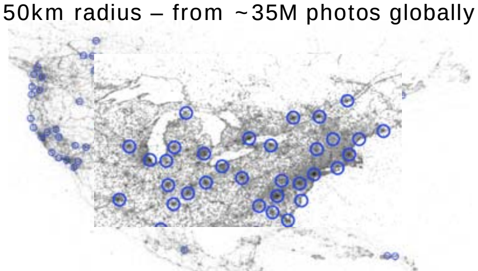
Mean Shift Clustering

- Simple non-parametric procedure for estimating peaks in distribution [Comaniciu02]
  1. initialize kernel (e.g., disc) to some position
  2. compute centroid of samples inside the disc
  3. move center of disc to centroid
  4. stop if converged, otherwise go to step 2



Sample Clustering Result

- Top 100 clusters in North America at 50km radius – from ~35M photos globally



Representative Text Tags

- Text tags that are characteristic of a given spatial region
  - Score tags according to likelihood in region versus baseline occurrence

$$\frac{P(\text{photo } p \text{ has tag } t \mid p \text{ inside region})}{P(\text{photo } p \text{ has tag } t)}$$

  - Limit any single user's contribution in a region
  - Consider tags that occur for at least some fraction of photos in region (e.g., 5%)
  - Similar approaches in [Ahern07] [Kennedy08]
- Top scoring tags ordered by likelihood

Tags for Top 100km Radius Clusters

Rank	Users	Photos	Most distinctive tags
1	20138	726693	manhattan nyc newyorkcity newyork
2	16541	700108	london england uk
3	15316	707604	sanfrancisco california
4	10004	457873	losangeles california
5	9563	320423	paris france
6	6905	349931	washingtondc dc washington
7	6754	310579	chicago illinois
8	6663	343940	seattle washington
9	5375	249257	boston massachusetts
10	5185	192230	sandiego california
11	4910	154523	amsterdam holland netherlands
12	4817	138594	rome roma italy italia
13	4564	144449	barcelona spain
14	4398	141786	berlin germany
15	4346	141931	monterey santacruz california

Clusters at Multiple Geo Scales

- Cities and metropolitan areas form natural peaks at 100km radius
  - From large areas like London, Paris and LA to small areas such as Ithaca and Iowa City
- Landmarks often correspond to peaks at approximately 100m radius
  - Buildings such as St. Paul's Cathedral, places such as Rockefeller Plaza or Trafalgar Square
- Spatial hierarchy
  - Use landmark peaks within a city peak to describe the city (similarly for neighborhoods)

Top Landmarks (City and Global)

|| Top landmark | 2nd landmark | 3rd landmark | 4th landmark | 5th landmark |

	Top landmark	2nd landmark	3rd landmark
Earth	eiffel	trafalgarsquare	tatemodern
1. **newyorkcity**	empirestatebuilding	timessquare	rockefeller
2. **london**	trafalgarsquare	tatemodern	bigben
3. **sanfrancisco**	coittower	pier39	unionsquare
4. **paris**	eiffel	notredame	louvre
5. **losangeles**	disneyland	hollywood	gettymuseum
6. **chicago**	cloudgate	chicagoriver	hancock
7. **washingtondc**	washingtonmonument	wwii	lincolnmemorial
8. **seattle**	spaceneedle	market	seattlepubliclibrary
9. **rome**	colosseum	vaticano	pantheon
10. **amsterdam**	dam	westerkerk	nieuwmarkt
11. **boston**	fenwaypark	trinitychurch	faneuilhall
12. **barcelona**	sagradafamilia	parcguell	boqueria
13. **sandiego**	balboapark	sandiegozoo	ussmidway
14. **berlin**	brandenburgertor	reichstag	potsdamerplatz
15. **lasvegas**	paris	newyorknewyork	bellagio

Saliency of a City's Landmarks

- Simple measure $\frac{\text{total \# of photos in top 10 landmarks of city } c}{\text{\# of photos in } c}$

Most salient

58.2	agra tajmahal
49.4	córdoba cordoba
46.4	dubrovnik croatia
45.7	salamanca españa
44.2	blackrockcity burningman
42.0	ljubljana slovenia
38.5	corpuschristi texas
34.6	montsaintmichel saintmalo
33.5	grandcanyon grand
32.8	deathvalley death
28.0	romë roma
27.9	trogir split

Least salient

6.1	desmoines iowa
6.1	minneapolis minnesota
6.0	fremantle perth
6.0	bern suisse
5.9	rochester ny
5.9	brisbane queensland
5.9	frankfurt germany
5.8	brest finistère
5.8	amsterdam holland
5.7	newcastle durham
3.7	graubünden schweiz
3.4	taipei taiwan

Representative Images

- Finding visual characterizations of clusters
  - Harder than selecting high likelihood text tags
  - Similar images primarily when taken at nearly the same place – 100m scale
    - Though some characteristic images at city scale too such as NYC yellow cabs, London buses
  - Similar images are generally a relatively small percentage of all images in a spatial cluster
    - E.g., random photos of Independence Hall vs. canonical view such as full facade

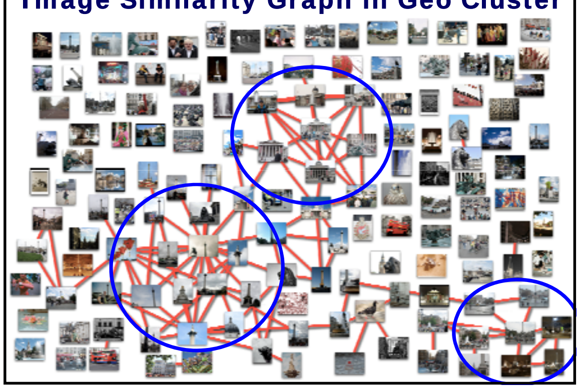
Representative Images (2)

- Related work on clustering textual and visual features [Kennedy08]
  - Using 100k photos of San Francisco and hand-selected landmarks, not that scalable
  - Others have used mix of content and geo, we argue for separating

Representative Images (3)

- Highly-photographed thing in geo cluster
  - Each photo is "vote" for importance
- Build an image similarity graph
  - Measure similarity between pairs of photos using local interest point descriptors
  - Nodes represent images, edge weights represent similarities
- Find highly-connected components in the image similarity graph
  - Using spectral clustering (e.g., [Shi00])
- Select high degree node in component

Image Similarity Graph in Geo Cluster



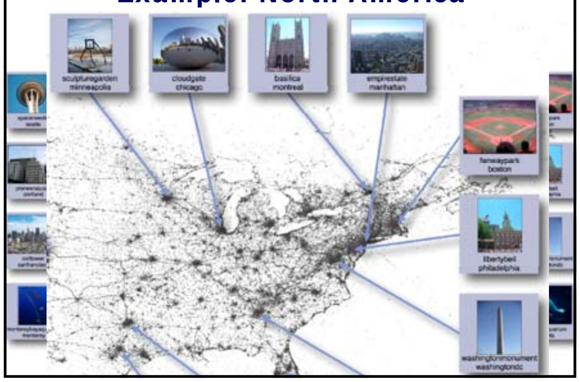
Measuring Image Similarity

- Use SIFT locally invariant interest point descriptors [Lowe04]
  - Points that are stable across image transformations (e.g. corners)
  - Compute invariant descriptor for each interest point
  - ~1000 interest points per image, 128-dimensional descriptors
- To compare 2 images, count “matching” points – descriptors highly similar

Creating Shared Mental Maps

- We now have automatic techniques for
  - Finding highly-photographed spatial regions, at multiple scales
  - Finding representative textual tags
  - Finding representative images at landmark scale
- Use to create labeled maps of “what’s important” completely automatically
  - City and landmark scales (100km and 100m)
  - From ~35M geo-tagged photos on Flickr, downloaded via API, medium res. (~500 x 350)
- Computation on 50-node Hadoop cluster

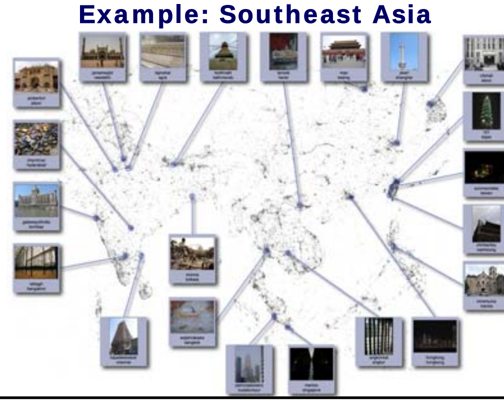
Example: North America



Example: Europe

Example: South America

Example: Southeast Asia



Example: UK and Ireland

Example: Landmarks in Manhattan

Example: Landmarks in Paris

Example: Landmarks in DC

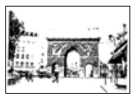
Example: Landmarks in London

Inferring Spatial Location

- Inverse problem: inferring location given images (possibly also text tags)
- [Milgram76] studied how people do
  - Where place photos in their “mental map”
- [Hays08] geo-locate images from visual features – estimate lat-long
  - Nearest-neighbor search on “training” dataset of 6 million images
    - Localize 16% of photos within 200km
    - Small test set of 237 hand-selected images
  - Similar approach in [Tsai05] for 1k images and 10 landmarks

Location: Landmark Classification

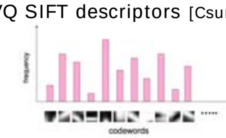
- Our approach is motivated by idea of mental map – saliency and importance
  - Localize key places rather than trying to place any image in lat-long coordinates
- Consider small numbers of identifiable locations in a given city and in the world

[Milgram76]

Classifying Landmarks

- Given a photo known to be taken at one of several landmarks, identify correct one
  - Using svm_multiclass [Tsochantaridis05]
- Textual and visual features based on vector space models
  - Each text tag with >3 occurrences a dimension
  - Codebook of 1-10k VQ SIFT descriptors [Csurka04]

Classification Experiments

- Learn n landmarks, classify disjoint test set
  - Between 10 and 500 landmarks
  - At least hundreds of training and test images per landmark
  - One person's photos only in training or in test
- Landmark recognition more general than specific object recognition (e.g., Trafalgar)
- Random baseline of 1/n
  - Restrict to same number of photos for each landmark in given experiment for comparison
  - Similarly significant if use true unequal counts

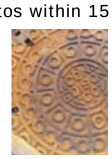
Landmark Classification Results

		Single images		
Categories	**Baseline**	visual	textual	combined
Top 10 landmarks	10.00	53.39	69.25	80.11
Landmarks 200-209	10.00	49.02	79.47	85.91
Landmarks 400-409	10.00	40.20	78.37	82.50
Top 20 landmarks	5.00	44.54	57.61	69.29
Landmarks 200-219	5.00	38.57	71.13	78.67
Landmarks 400-419	5.00	27.93	71.56	75.82
Top 50 landmarks	2.00	35.97	52.52	63.45
Landmarks 200-249	2.00	27.45	65.62	72.63
Landmarks 400-449	2.00	21.70	64.91	69.77
Top 100 landmarks	1.00	27.19	50.44	60.77
Top 200 landmarks	0.50	17.87	47.02	55.29
Top 500 landmarks	0.20	9.21	40.58	44.96

Photo Sequences

- Photos nearby in time for a particular photographer
  - Highly related location but often quite different image content (and text tags)
  - Exploit to improve classification results
    - Include features from photos within 15 minutes





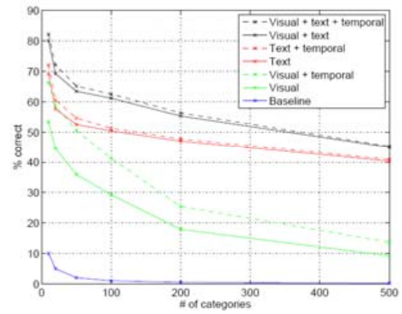

Structured Output for Sequences

- Classify sequence of photos in terms of what landmarks taken in succession
  - Use neighbors as context for given photo, i.e., score single photo not entire sequence
- Use svm_struct
  - For predicting structured outputs, reduces to svm_multiclass for length 1 sequences
  - Viterbi-style decoding/learning
- Strength of temporal relations based on time and distance (known for training)

Temporal Classification Results

Categories	Baseline	Single images visual	Single images textual	Single images combined	Photo streams visual	Photo streams textual	Photo streams combined
Top 10 landmarks	10.00	53.39	69.25	80.11	66.35	72.10	82.22
Landmarks 200-209	10.00	49.02	79.47	85.91	57.95	79.49	86.81
Landmarks 400-409	10.00	40.20	78.37	82.50	48.90	78.68	83.23
Top 20 landmarks	5.00	44.54	57.61	69.29	58.67	60.56	72.10
Landmarks 200-219	5.00	38.57	71.13	78.67	49.70	72.10	80.02
Landmarks 400-419	5.00	27.93	71.56	75.82	34.65	72.70	76.28
Top 50 landmarks	2.00	35.97	52.52	63.45	50.57	54.64	65.16
Landmarks 200-249	2.00	27.45	65.62	72.63	37.22	67.26	74.09
Landmarks 400-449	2.00	21.70	64.91	69.77	29.65	66.90	71.62
Top 100 landmarks	1.00	27.19	50.44	60.77	41.29	51.32	62.56
Top 200 landmarks	0.50	17.87	47.02	55.29	25.44	47.73	56.30
Top 500 landmarks	0.20	9.21	40.58	44.96	13.68	41.02	45.28

Landmark Classification Results



Larger VQ Codebooks

- VQ SIFT descriptors not necessarily good features for such a task
  - Continued improvement with bigger codebook
- Clustering billions of features into tens of thousands of clusters so far prohibitive
  - Though not at classification time

# of categories	Single images 1,000	2,000	5,000	10,000
10	44.68	48.43	53.39	54.51
20	35.73	38.40	44.54	46.10
50	24.47	30.35	35.97	37.58
100	16.90	20.54	27.19	29.29

Temporal Paths

Summary

- Photo sharing sites reveal information about collective perception of world
- We study how to exploit this
  - Automatically organize large photo collections
  - Discover interesting things about the world and about human behavior
- Automatically extract hotspots and labels
  - Find spatial clusters at different scales
  - Extract textual and visual representations clusters
- Localize and model popular landmarks

Questions

- D. Crandall, L. Backstrom, D. Huttenlocher and J. Kleinberg. Mapping the World's Photos. WWW09.
- D. Crandall, Y. Li and D. Huttenlocher. Landmark Classification in Large-Scale Image Collections. ICCV09.