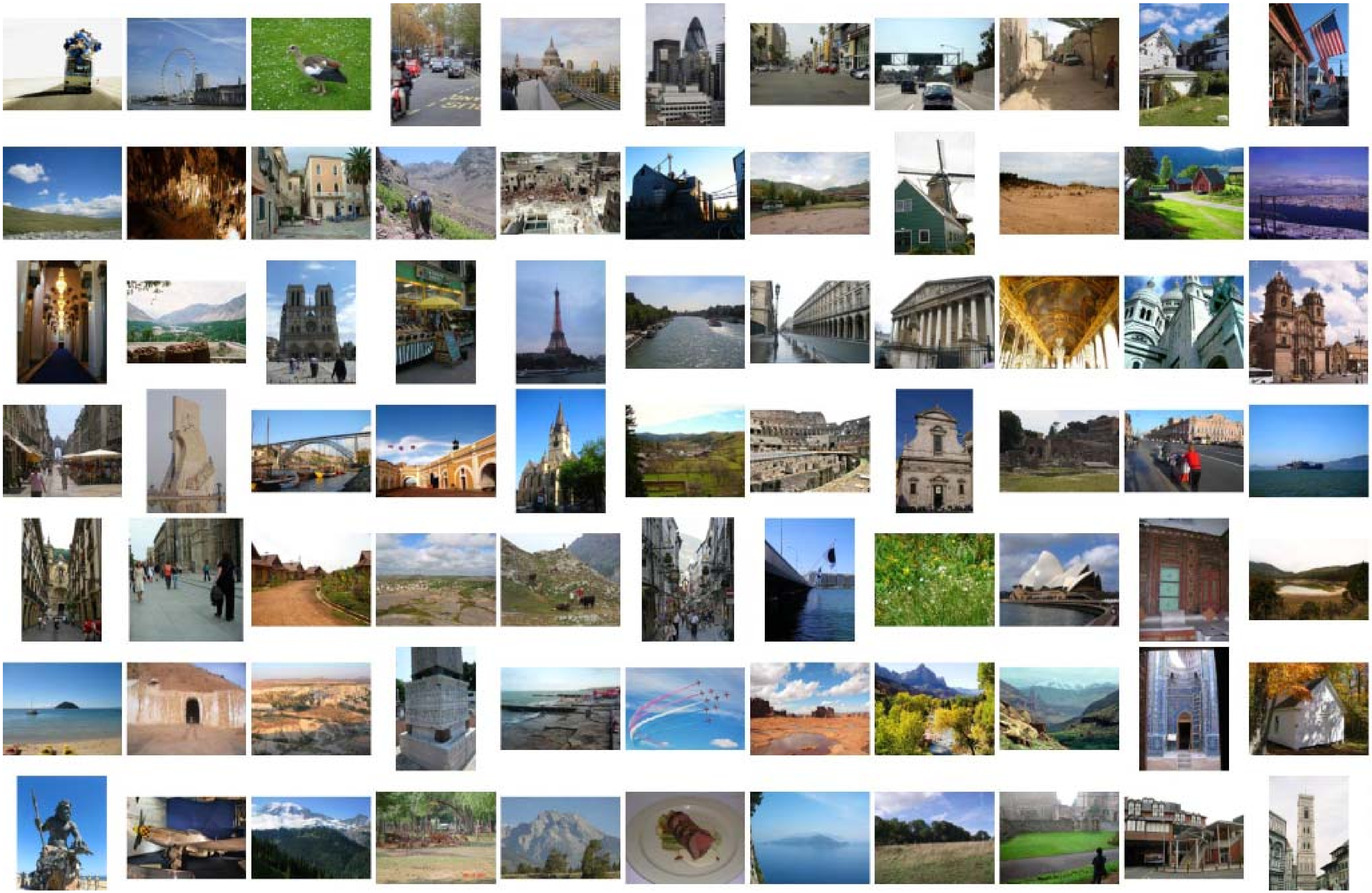
Photo Tourism and im2gps: 3D Reconstruction and Geolocation of Internet Photo Collections
Part II

Noah Snavely
Cornell

James Hays
CMU
MIT (Fall 2009)
Brown (Spring 2010 -)

Complexity of the Visual World



The Internet as a Data Source

- Social Networking Sites (e.g. Facebook, MySpace)
- Image Search Engines (e.g. Google, Bing)
- Photo Sharing Sites (e.g. Flickr, Picasa, Panoramio, photo.net, dpchallenge.com)
- Computer Vision Databases (e.g. CalTech 256, PASCAL VOC, LabelMe, Tiny Images, image-net.org, ESP game, Squigl, Matchin)

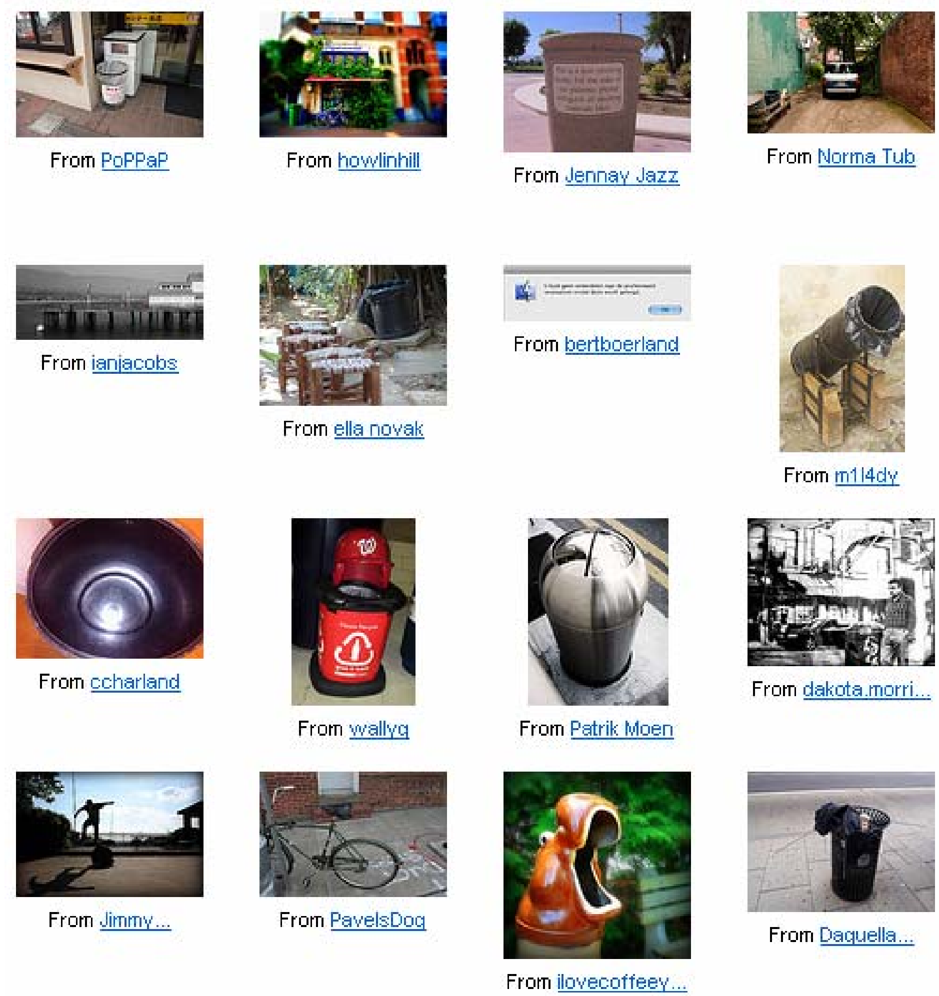
How Big is Flickr?

- As of June 19th, 2009
- Total content:
  - 3.6 billion photographs
  - 100+ million geotagged images
- *Public* content:
  - 1.3 billion photographs
  - 74 million geotagged images

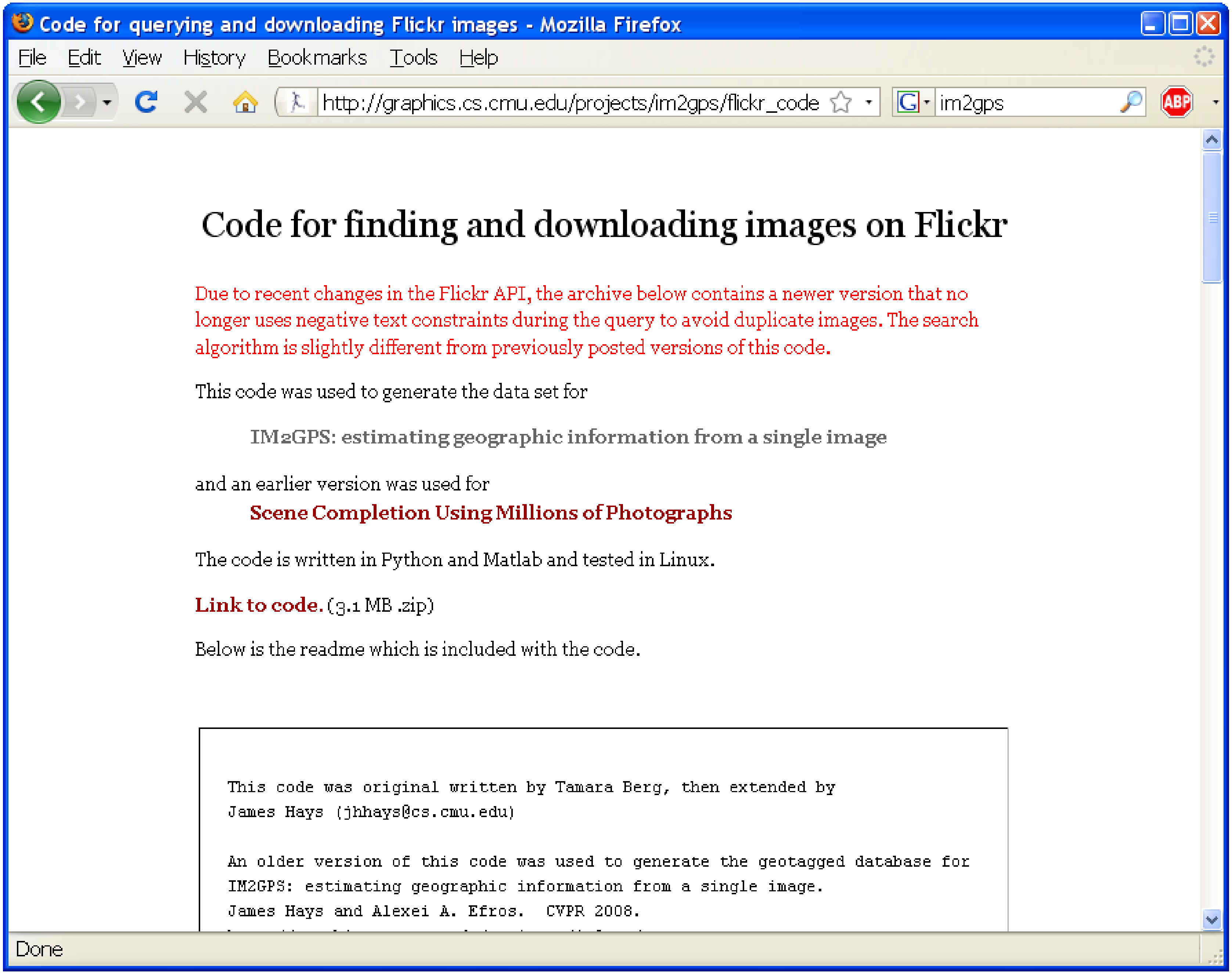
How Annotated is Flickr? (tag search)

- Party – 7,355,998
- Paris – 4,139,927
- Chair – 232,885
- Violin – 55,015
- Trashcan – 9,818

Trashcan Results



- http://www.flickr.com/search/?q=trashcan+NOT+party&m=tags&z=t&page=5



- http://graphics.cs.cmu.edu/projects/im2gps/flickr_code.html

http://www.flickr.com/services/api/flickr.photos.search.html

flickr.photos.search

Return a list of photos matching some criteria. Only photos visible to the calling user will be returned. To return private or semi-private photos, the caller must be authenticated with 'read' permissions, and have permission to view the photos. Unauthenticated calls will only return public photos.

Authentication

This method does not require authentication.

Arguments

api_key (Required)
: Your API application key. See here for more details.

user_id (Optional)
: The NSID of the user who's photo to search. If this parameter isn't passed then everybody's public photos will be searched. A value of "me" will search against the calling user's photos for authenticated calls.

tags (Optional)
: A comma-delimited list of tags. Photos with one or more of the tags listed will be returned.

tag_mode (Optional)
: Either 'any' for an OR combination of tags, or 'all' for an AND combination. Defaults to 'any' if not specified.

text (Optional)
: A free text search. Photos who's title, description or tags contain the text will be returned.

min_upload_date (Optional)
: Minimum upload date. Photos with an upload date greater than or equal to this value will be returned. The date should be in the form of a unix timestamp.

max_upload_date (Optional)
: Maximum upload date. Photos with an upload date less than or equal to this value will be returned. The date should be in the form of a unix timestamp.

min_taken_date (Optional)
: Minimum taken date. Photos with an taken date greater than or equal to this value will be returned. The date should be in the form of a mysql datetime.

max_taken_date (Optional)
: Maximum taken date. Photos with an taken date less than or equal to this value will be returned. The date should be in the form of a mysql datetime.

Undocumented
tag_mode: bool
e.g. tags = cat NOT tiger

http://www.flickr.com/services/api/flickr.photos.search.html

license (Optional)

The license id for photos (for possible values see the flickr.photos.licenses.getInfo method). Multiple licenses may be comma-separated.

sort (Optional)

The order in which to sort returned photos. Deafults to date-posted-desc (unless you are doing a radial geo query, in which case the default sorting is by ascending distance from the point specified). The possible values are: date-posted-asc, date-posted-desc, date-taken-asc, date-taken-desc, interestingness-desc, interestingness-asc, and relevance.

privacy_filter (Optional)

Return photos only matching a certain privacy level. This only applies when making an authenticated call to view photos you own. Valid values are:

- 1 public photos
- 2 private photos visible to friends
- 3 private photos visible to family
- 4 private photos visible to friends & family
- 5 completely private photos

bbox (Optional)

A comma-delimited list of 4 values defining the Bounding Box of the area that will be searched.

The 4 values represent the bottom-left corner of the box and the top-right corner, minimum_longitude, minimum_latitude, maximum_longitude, maximum_latitude.

Longitude has a range of -180 to 180 , latitude of -90 to 90. Defaults to -180, -90, 180, 90 if not specified.

Unlike standard photo queries, geo (or bounding box) queries will only return 250 results per page.

Geo queries require some sort of limiting agent in order to prevent the database from crying. This is basically like the check against "parameterless searches" for queries without a geo component.

A tag, for instance, is considered a limiting agent as are user defined min_date_taken and min_date_upload parameters — If no limiting factor is passed we return only photos added in the last 12 hours (though we may extend the limit in the future).

accuracy (Optional)

Recorded accuracy level of the location information. Current range is 1-16 :

- World level is 1
- Country is ~3
- Region is ~6
- City is ~11
- Street is ~16

Defaults to maximum value if not specified.

safe_search (Optional)

Safe search setting:

- 1 for safe.
- 2 for moderate

http://www.flickr.com/services/api/flickr.photos.search.html

media (Optional)

Filter results by media type. Possible values are `all` (default), `photos` or `videos`

has_geo (Optional)

Any photo that has been geotagged, or if the value is "0" any photo that has *not* been geotagged.

Geo queries require some sort of limiting agent in order to prevent the database from crying. This is basically like the check against "parameterless searches" for queries without a geo component.

A tag, for instance, is considered a limiting agent as are user defined min_date_taken and min_date_upload parameters &emdash; If no limiting factor is passed we return only photos added in the last 12 hours (though we may extend the limit in the future).

geo_context (Optional)

Geo context is a numeric value representing the photo's geotagginess beyond latitude and longitude. For example, you may wish to search for photos that were taken "indoors" or "outdoors".

The current list of context IDs is :

- **0**, not defined.
- **1**, indoors.
- **2**, outdoors.

Geo queries require some sort of limiting agent in order to prevent the database from crying. This is basically like the check against "parameterless searches" for queries without a geo component.

A tag, for instance, is considered a limiting agent as are user defined min_date_taken and min_date_upload parameters &emdash; If no limiting factor is passed we return only photos added in the last 12 hours (though we may extend the limit in the future).

lat (Optional)

A valid latitude, in decimal format, for doing radial geo queries.

Geo queries require some sort of limiting agent in order to prevent the database from crying. This is basically like the check against "parameterless searches" for queries without a geo component.

A tag, for instance, is considered a limiting agent as are user defined min_date_taken and min_date_upload parameters &emdash; If no limiting factor is passed we return only photos added in the last 12 hours (though we may extend the limit in the future).

lon (Optional)

A valid longitude, in decimal format, for doing radial geo queries.

Geo queries require some sort of limiting agent in order to prevent the database from crying. This is basically like the check against "parameterless searches" for queries without a geo component.

http://www.flickr.com/services/api/flickr.photos.search.html

Geo queries require some sort of limiting agent in order to prevent the database from crying. This is basically like the check against "parameterless searches" for queries without a geo component.

A tag, for instance, is considered a limiting agent as are user defined min_date_taken and min_date_upload parameters &emdash; If no limiting factor is passed we return only photos added in the last 12 hours (though we may extend the limit in the future).

radius (Optional)
: A valid radius used for geo queries, greater than zero and less than 20 miles (or 32 kilometers), for use with point-based geo queries. The default value is 5 (km).

radius_units (Optional)
: The unit of measure when doing radial geo queries. Valid options are "mi" (miles) and "km" (kilometers). The default is "km".

is_commons (Optional)
: Limit the scope of the search to only photos that are part of the Flickr Commons project. Default is false.

extras (Optional)
: A comma-delimited list of extra information to fetch for each returned record. Currently supported fields are: `license`, `date_upload`, `date_taken`, `owner_name`, `icon_server`, `original_format`, `last_update`, `geo`, `tags`, `machine_tags`, `o_dims`, `views`, `media`.

per_page (Optional)
: Number of photos to return per page. If this argument is omitted, it defaults to 100. The maximum allowed value is 500.

page (Optional)
: The page of results to return. If this argument is omitted, it defaults to 1.

Example Response

This method returns the standard photo list xml:

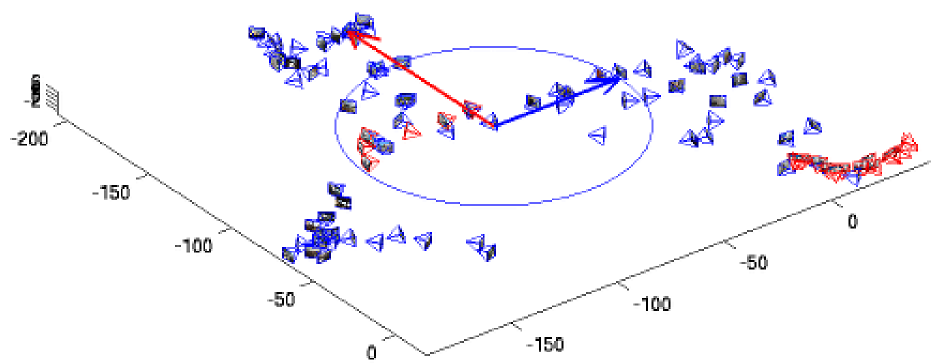
```
<photos page="2" pages="89" perpage="10" total="881">
        <photo id="2636" owner="47058503995@N01"
                secret="a123456" server="2" title="test_04"
                ispublic="1" isfriend="0" isfamily="0" />
        <photo id="2635" owner="47058503995@N01"
                secret="b123456" server="2" title="test_03"
                ispublic="0" isfriend="1" isfamily="1" />
        <photo id="2633" owner="47058503995@N01"
                secret="c123456" server="2" title="test_01"
                ispublic="1" isfriend="0" isfamily="0" />
        <photo id="2610" owner="12037949754@N01"
                secret="d123456" server="2" title="00_tall"
                ispublic="1" isfriend="0" isfamily="0" />
</photos>
```

To map `<photo>` elements to urls, please read the url documentation.

Global Image Geolocation

Background: Image Localization

- *Where Am I?* Contest. ICCV 2005. Organized by Richard Szeliski.
- *Nokia Challenge: Where was this Photo Taken, and How?* Multimedia Grand Challenge 2009.



Background: Image Localization

- Zhang and Kosecka. *Image Based Localization in Urban Environments*. I3DPVT 2006.
- Philbin, Chum, Isard, Sivic, and Zisserman*. Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases.* *C*VPR 2008

Global Image Geolocation

- Hays and Efros. *Im2gps: estimating geographic information from a single image*. CVPR 2008.
- Quack, Leibe, Van Gool. *World-Scale Mining of Objects and Events from Community Photo Collections*. CIVR 2008
- Crandall, Backstrom, Huttenlocher, and Kleinberg. *Mapping the World’s Photos*. WWW 2009.

im2gps

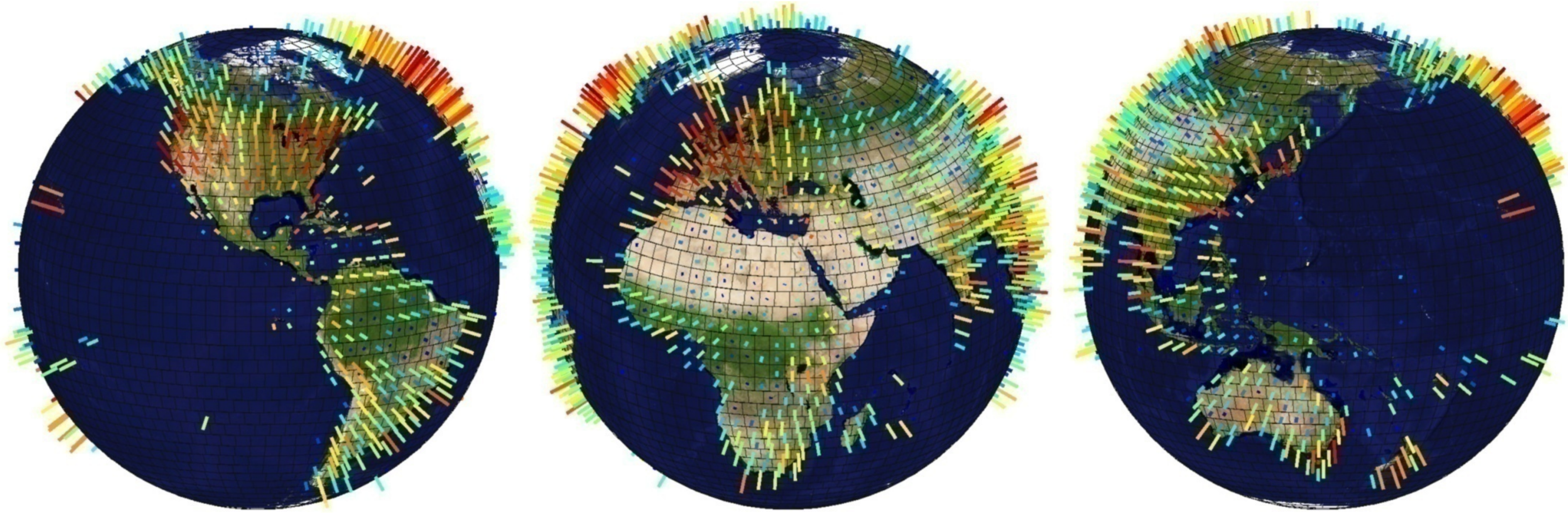
- Big Question: is geolocation just instance-level, landmark recognition or can you reason about location and geography from non-specific scene properties?

Where is this?

Where is this?

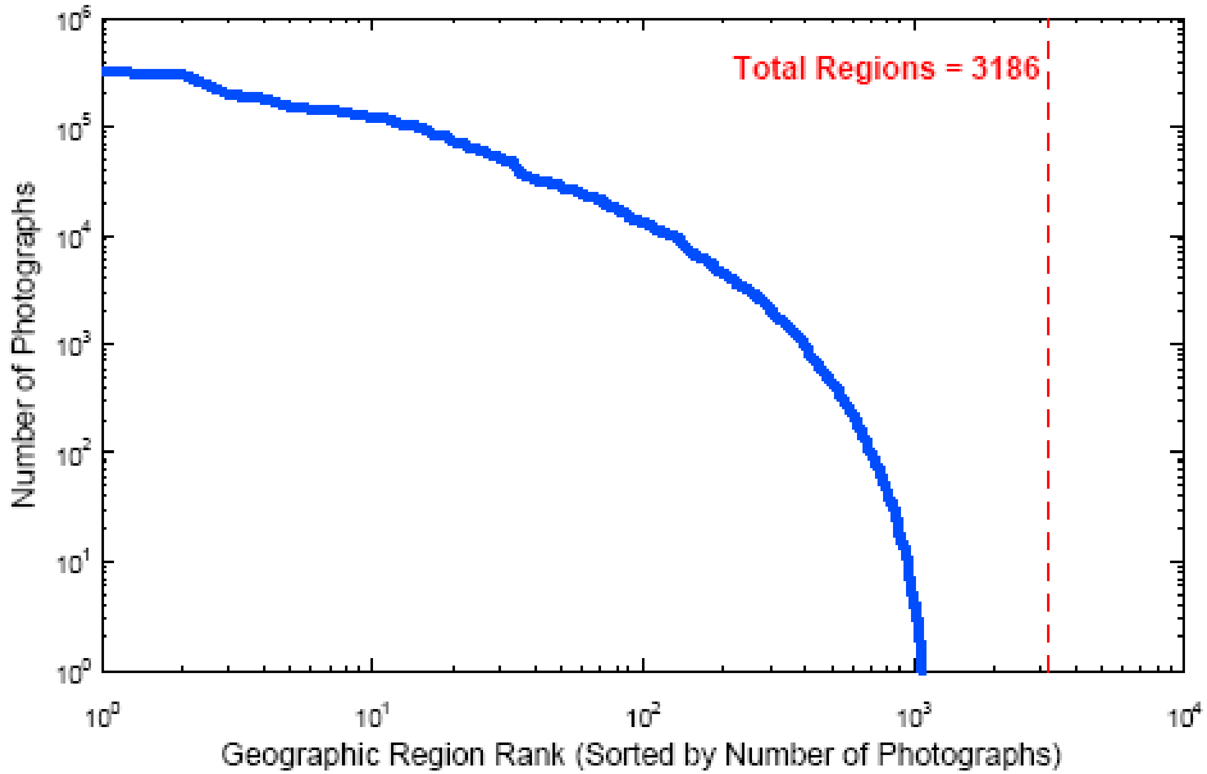
Where is this?

im2gps Geographic Photo Density



6.4 mil. photos by 110K photographers.
1 TB of visual data.
Photographs had at least one place keyword.
Photos average ~1 content descriptive keyword.

Photo Region Frequencies



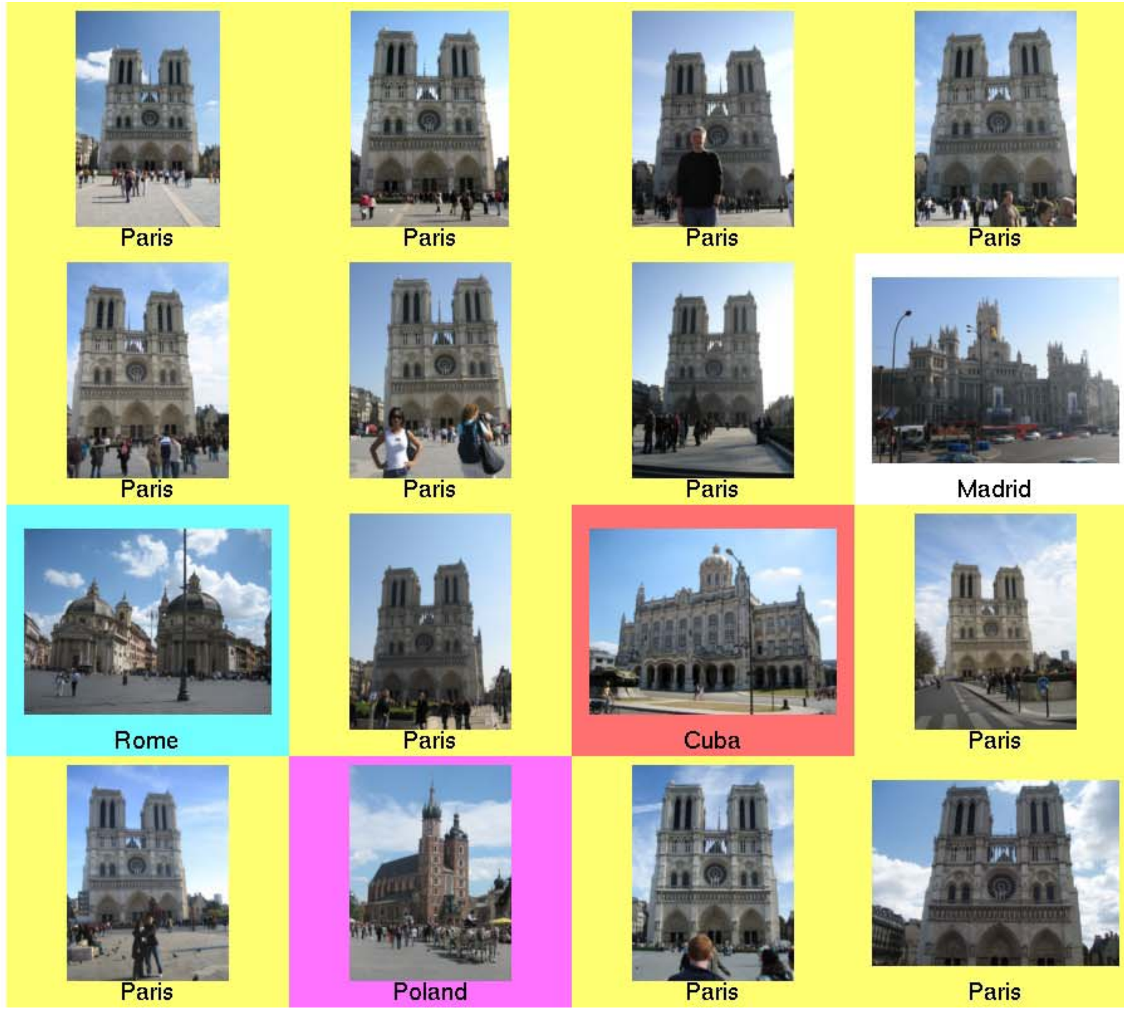
Im2gps Image Features

- **Gist descriptor** – 5x5 spatial resolution, 4 scales, 8 orientations. Code
- **Tiny Color Image** – 5x5 and 16x16 spatial resolutions.
- **Color Histogram** – L*A*B* 4x14x14 histograms.
- **Texton Histogram** – 512 entry, filter bank based. Code
- **Line Features** – Histograms of straight line lengths and angles.
- **Geometric Context** – 8x8 probability of geometric class (e.g. Ground, Sky, Vertical, Porous). Code
- Histograms are compared with Chi Squared measure, other features with L1 distance.

Scene Matches

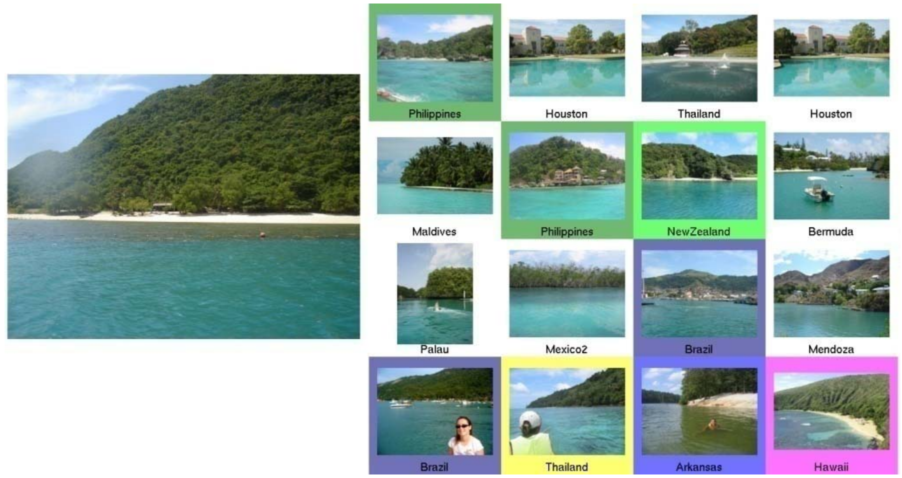

Madrid
england
France
Paris
Croatia
heidelberg
Macau
Malta
Cairo
Italy
Italy
Italy
Latvia
europe
Barcelona
Austria

Scene Matches

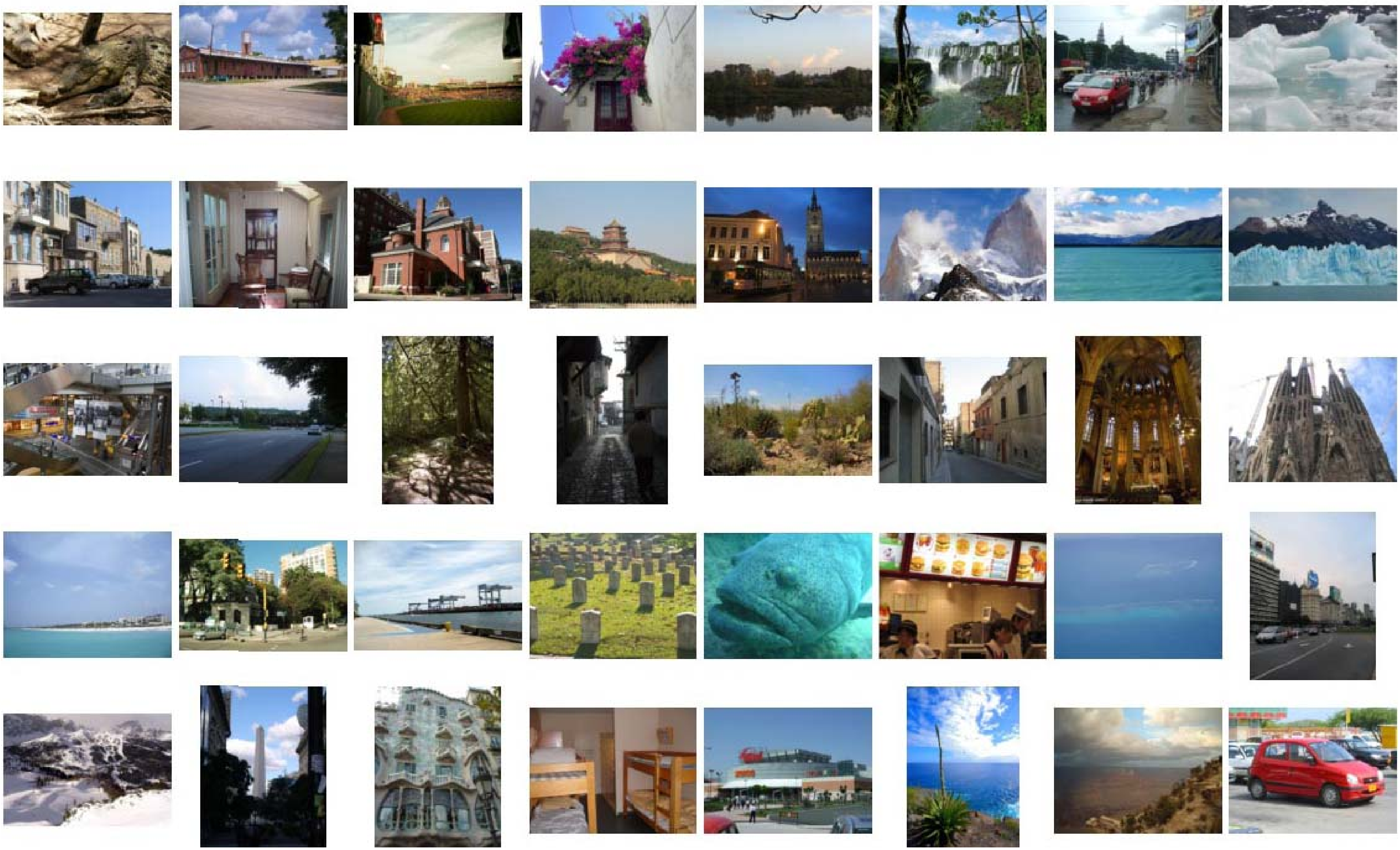



Paris
Paris
Paris
Paris
Paris
Paris
Paris
Madrid
Rome
Paris
Cuba
Paris
Paris
Poland
Paris
Paris

Scene Matches

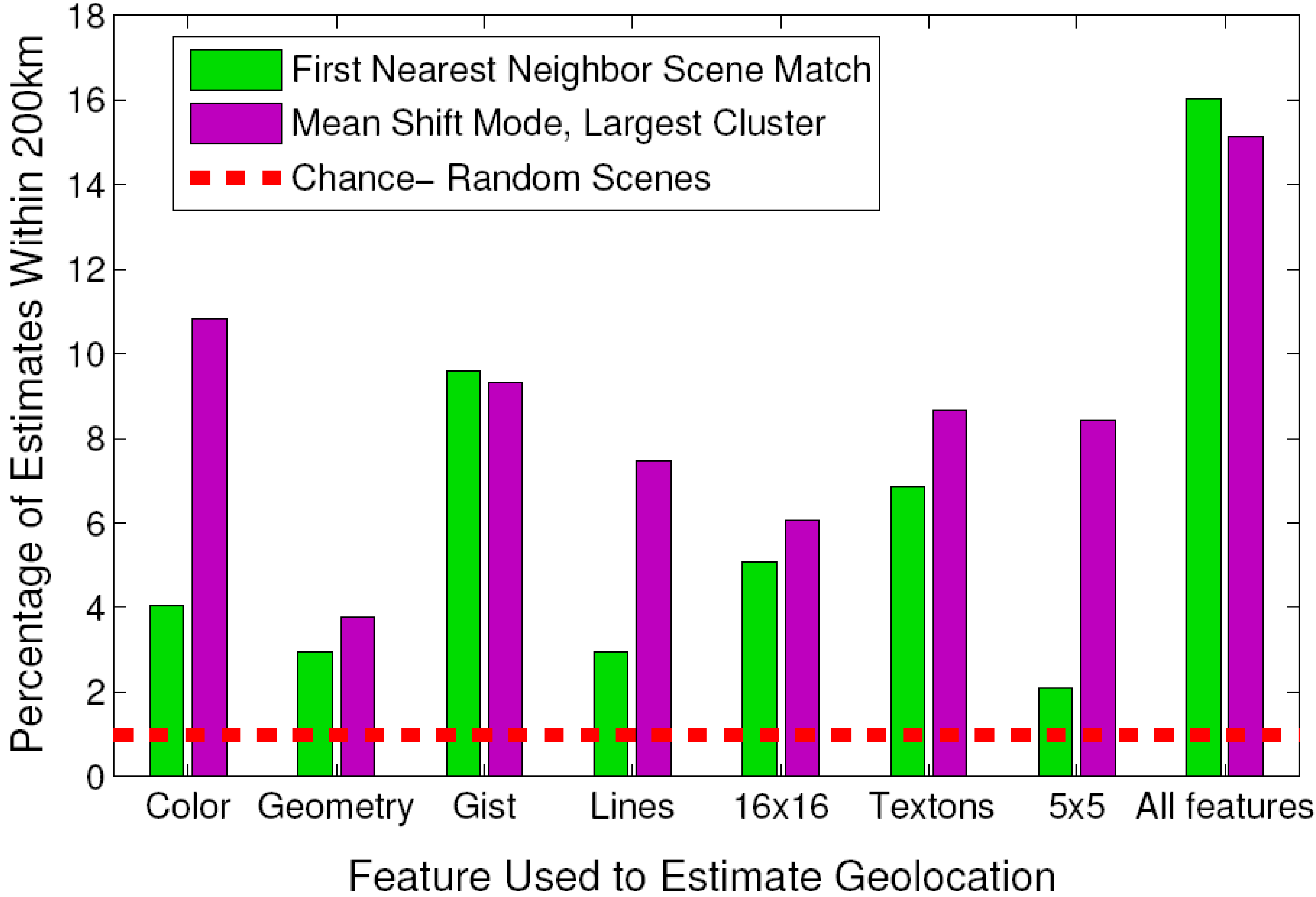

Philippines
Houston
Thailand
Houston
Maldives
Philippines
NewZealand
Bermuda
Palau
Mexico2
Brazil
Mendoza
Brazil
Thailand
Arkansas
Hawaii

Quantitative Evaluation Test Set

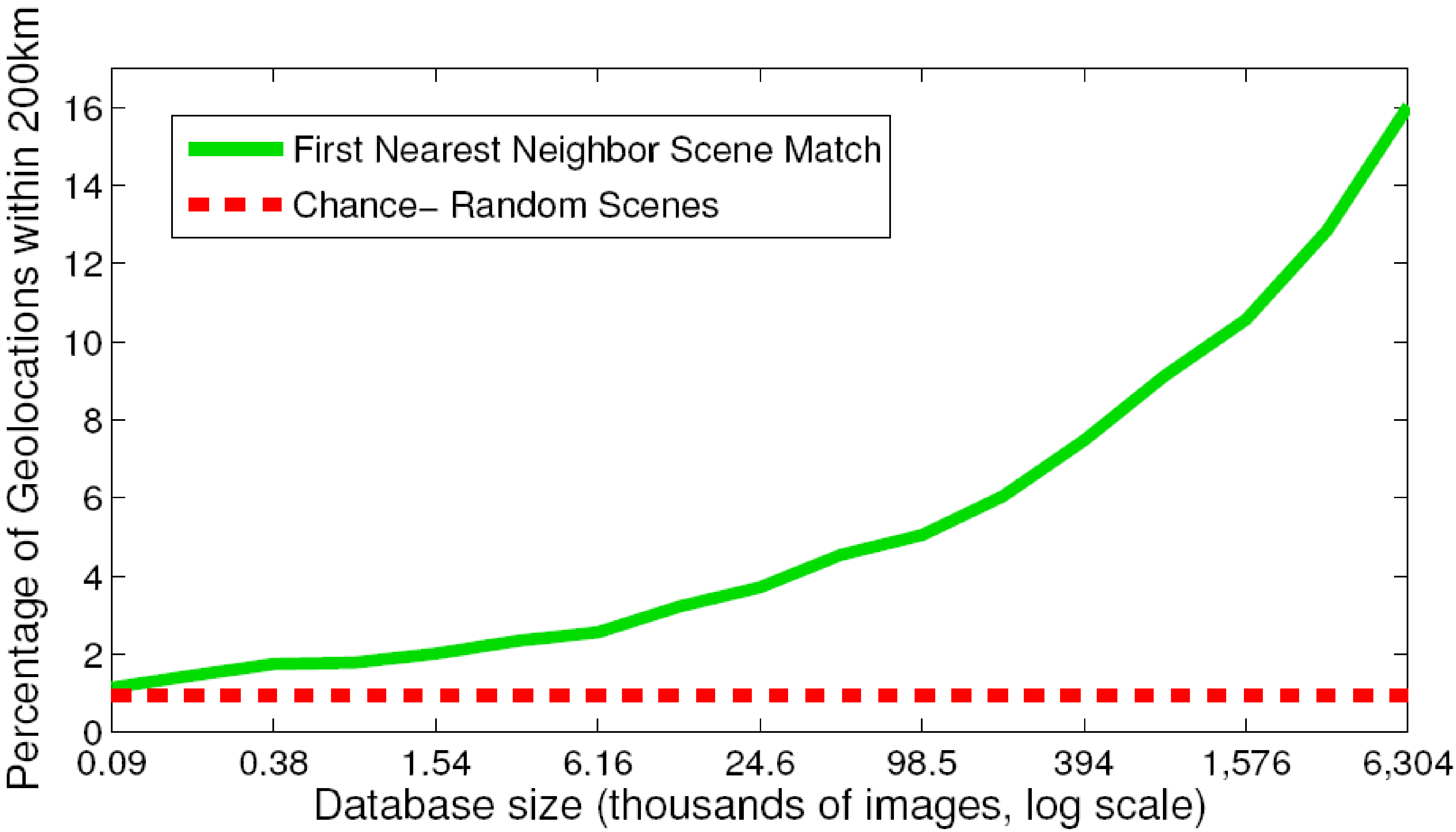


...

Feature Performance



The Importance of Data



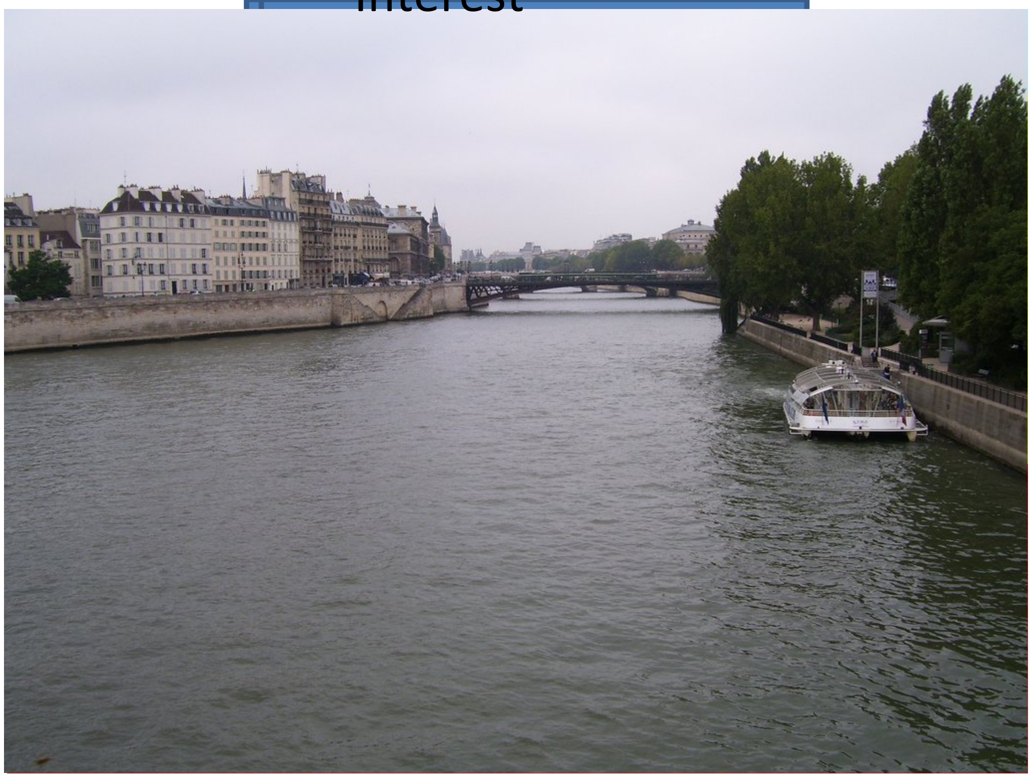
im2gps Feature Problems

- im2gps features aren’t very robust to minor scene differences (e.g. translation, cropping, illumination change).
- im2gps features aren’t ideal for instance-level landmark recognition.

im2gps Feature Problems

Scene Geometry Derived Features

Gist at Geometric Region of Interest



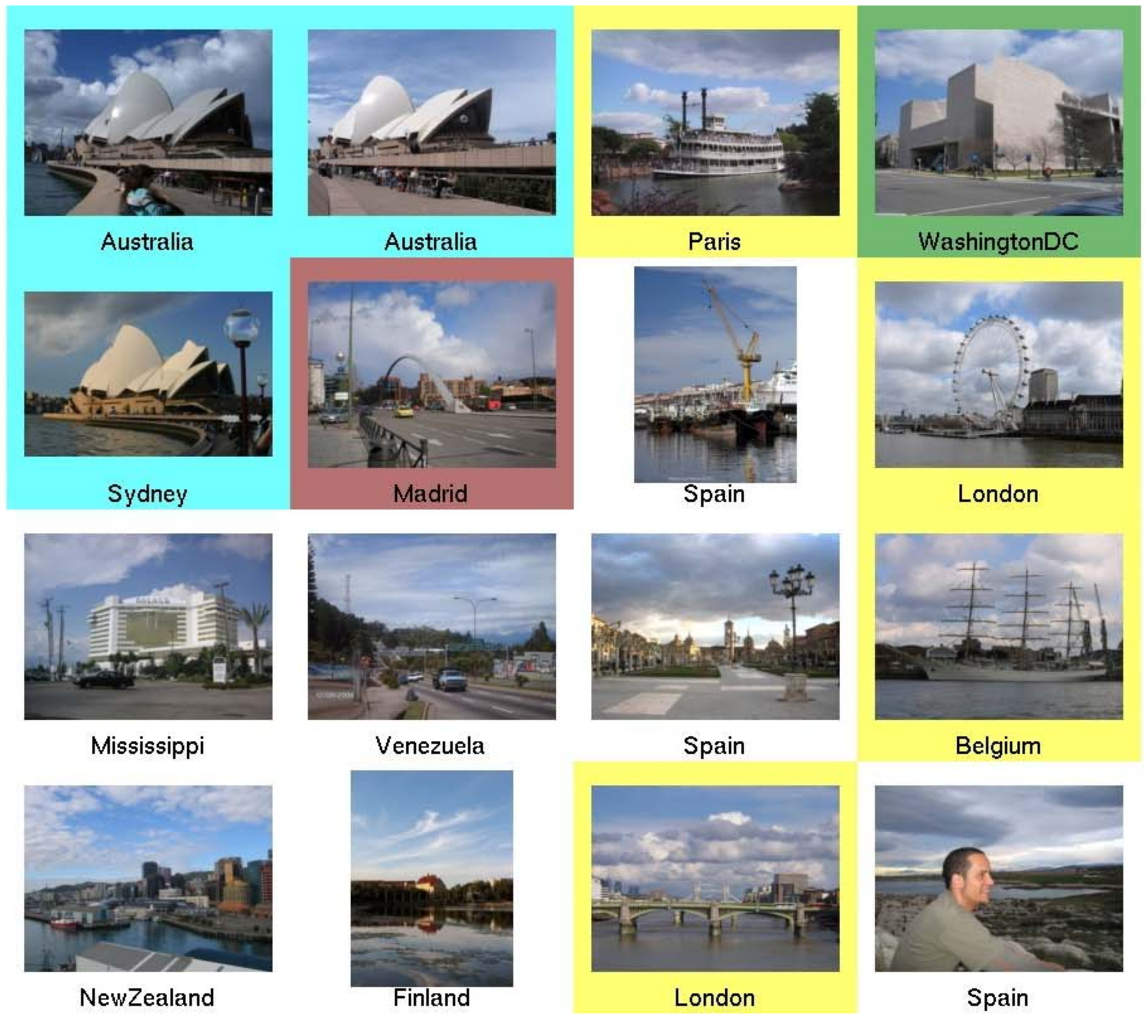
Sky: Color and Texton Histograms

Porous: Color and Texton Histograms

Vertical: Color and Texton Histograms

Ground: Color and Texton Histograms

im2gps Feature Problems

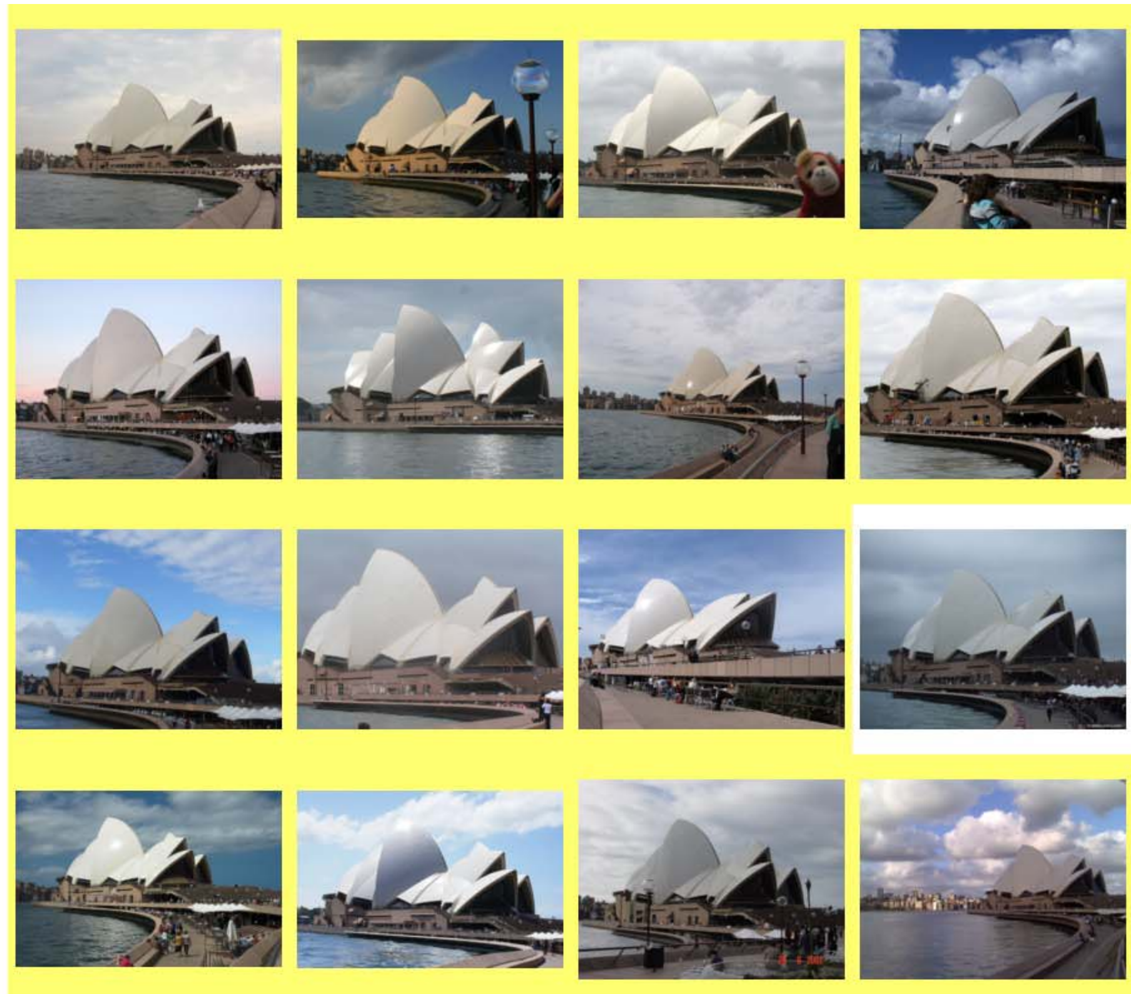


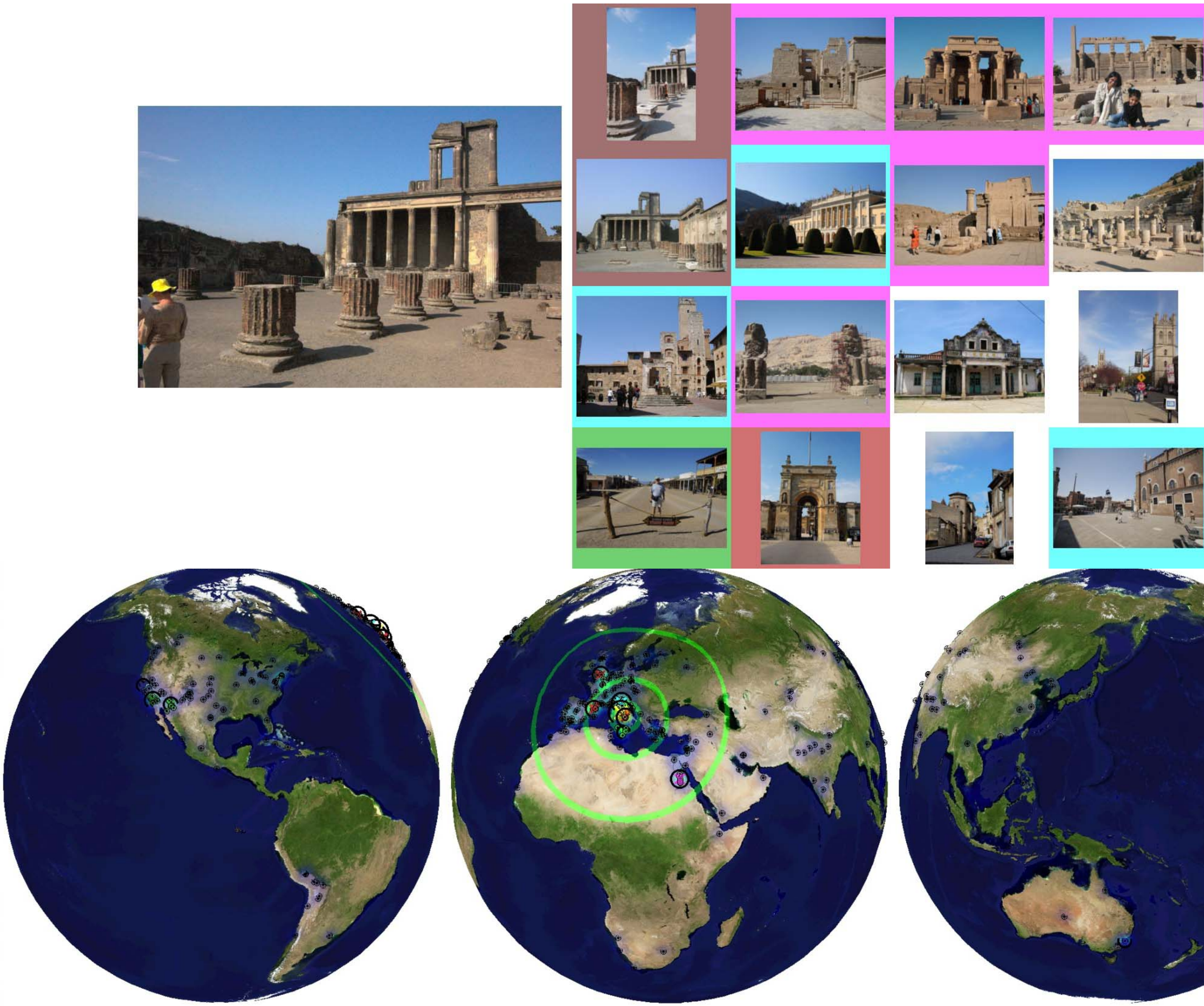
Large Vocabulary Bag of SIFTs

- Inspired by Philbin, Chum, Isard, Sivic, and Zisserman. *Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases.* CVPR 2008.
- 1,000 and 50,000 entry histograms for SIFTS found at MSER and Hessian Affine interest points, with 5-way soft assignment.
- No spatial verification, no inverted file.

New Feature Summary

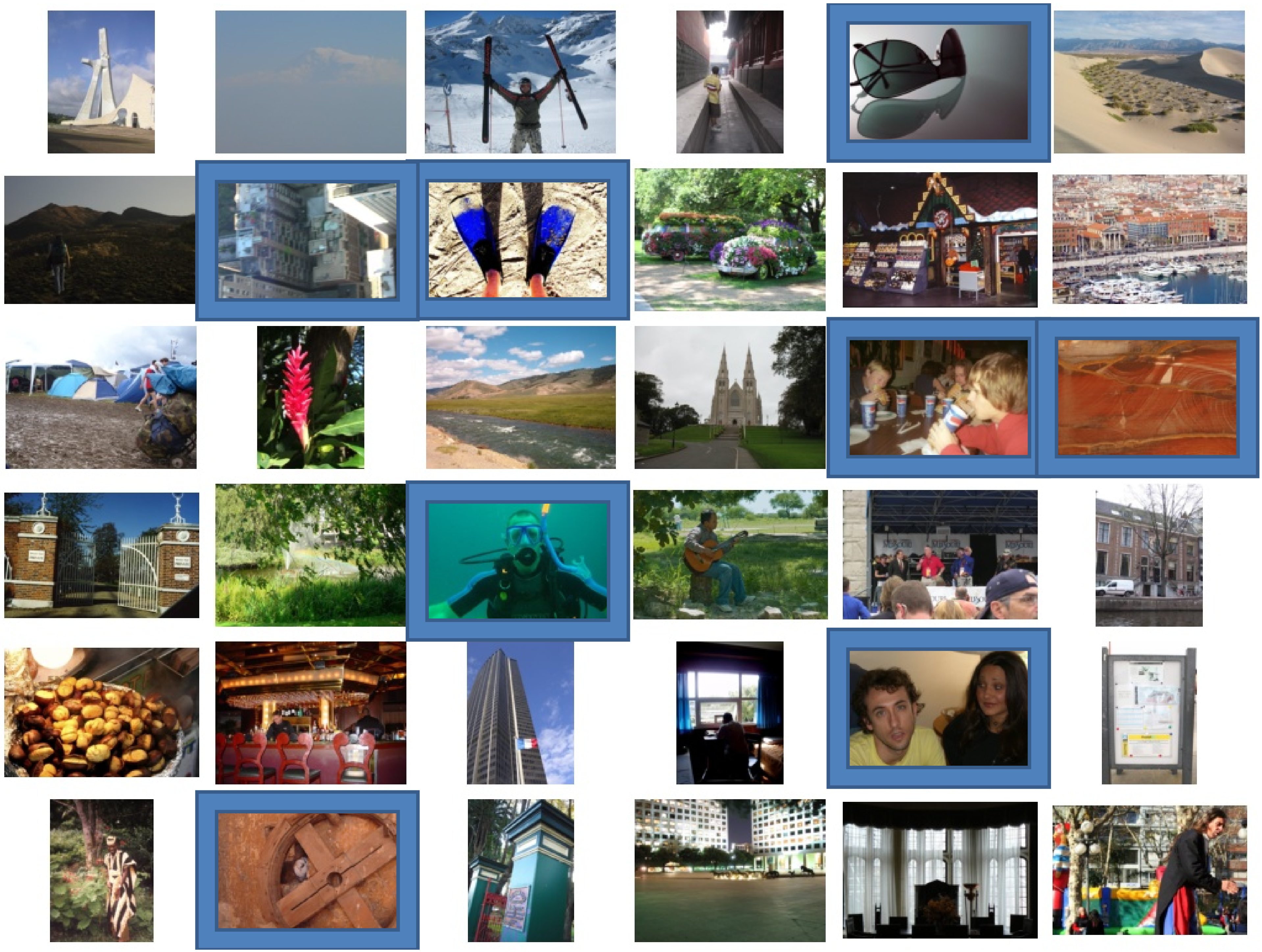
- Gist centered at geometric region of interest.
- Geometry modulated texton and color histograms.
- Bags of SIFT.
- We end up with ~30 elementary features.
- About 8 CPU years to compute.





New Test Set

- 2000 random, geotagged Internet Images.
- Most images are extremely difficult or impossible to geolocate.

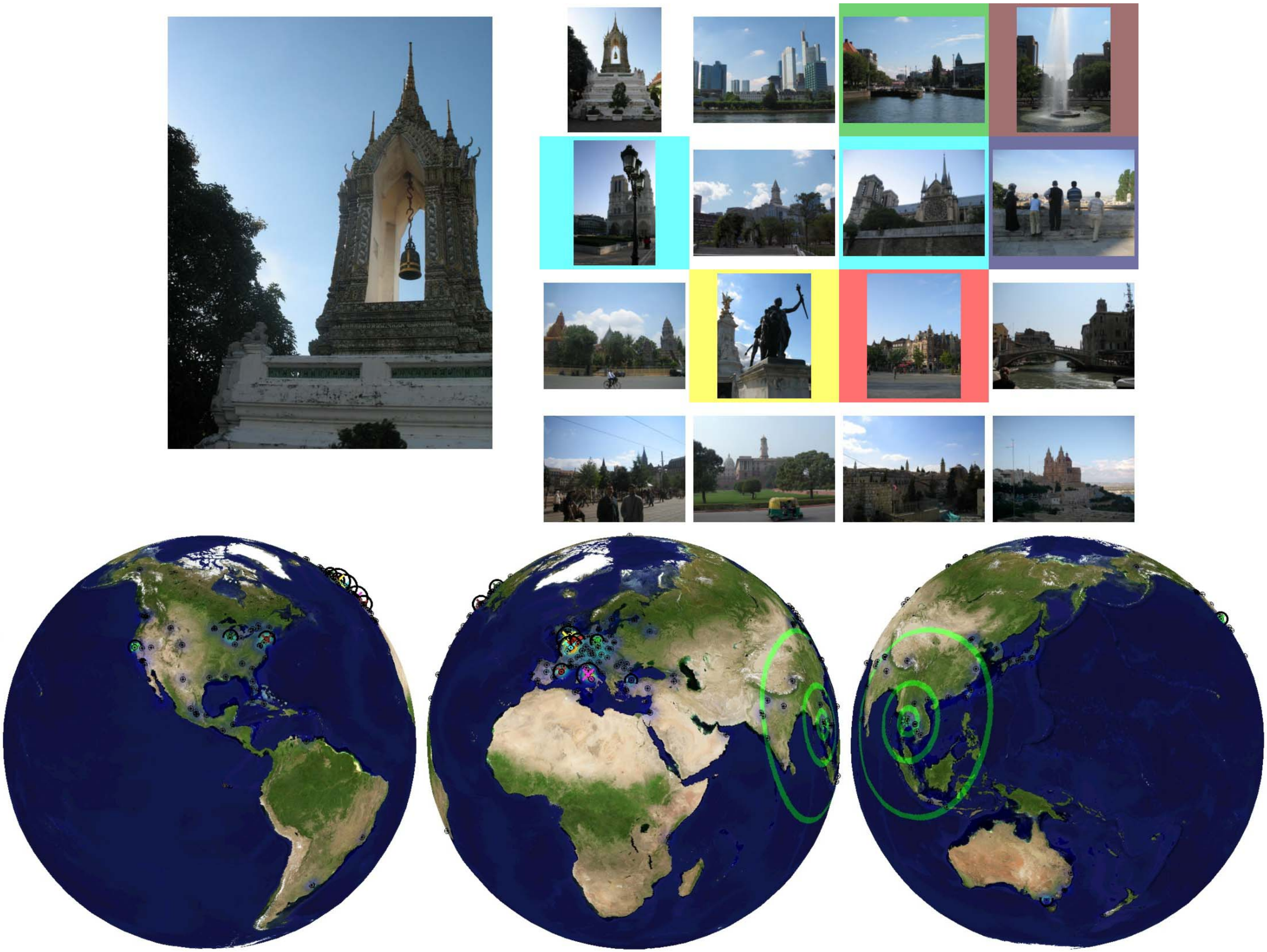


Quantitative Results

Geolocation Strategy	*Accuracy*
• Im2gps features, 1nn	6.35%
• Sift features, 1nn	8.05%
• All new features, 1nn	8.85%
• All new features, Largest Cluster	7.00%

Learning

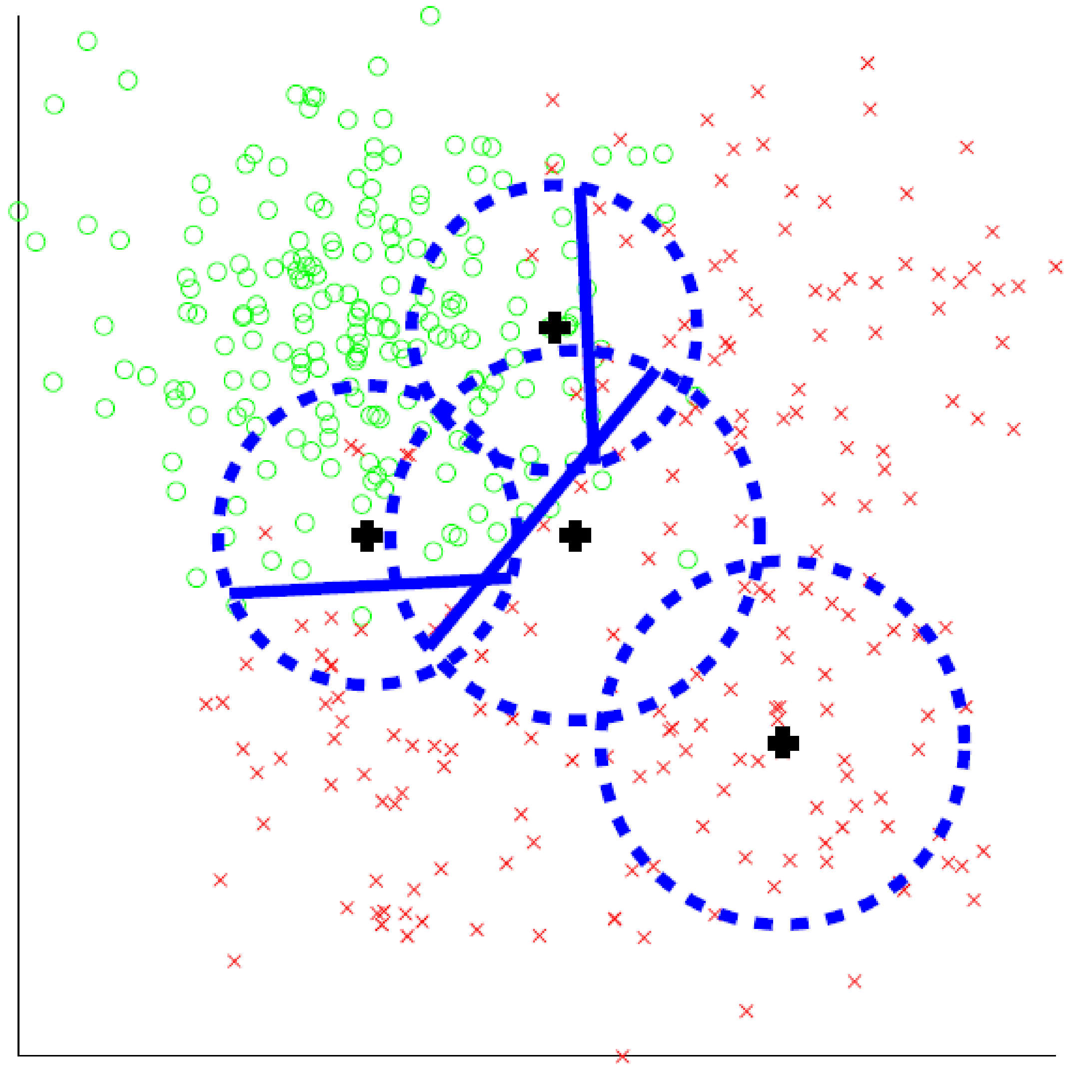
- Can we use machine learning to improve geolocation accuracy?
- Why would we expect learning to work better than 1 nearest neighbor?



Learning Issues

- We have 6+ million images, with 100,000 feature dimensions, so naïve application of many learning algorithms is intractable.
- What are our classes? Isn’t geolocation more naturally a regression problem?

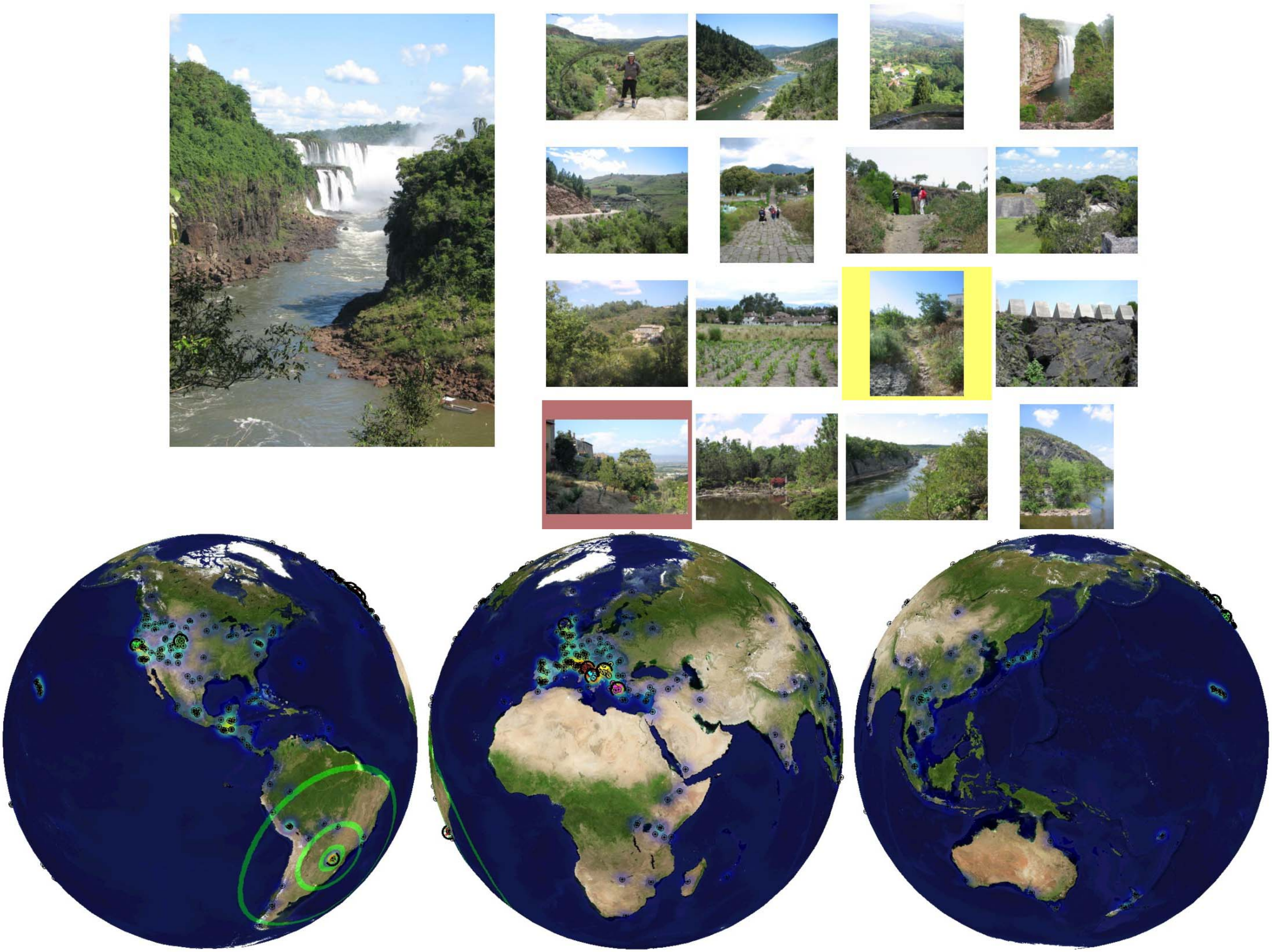
“Lazy Learning” - SVM-KNN

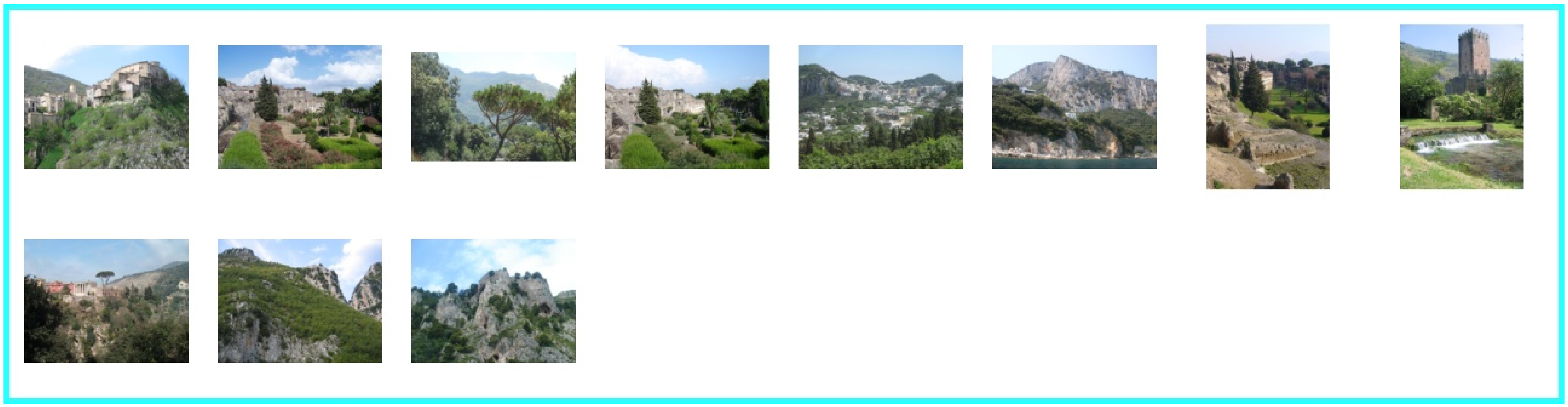


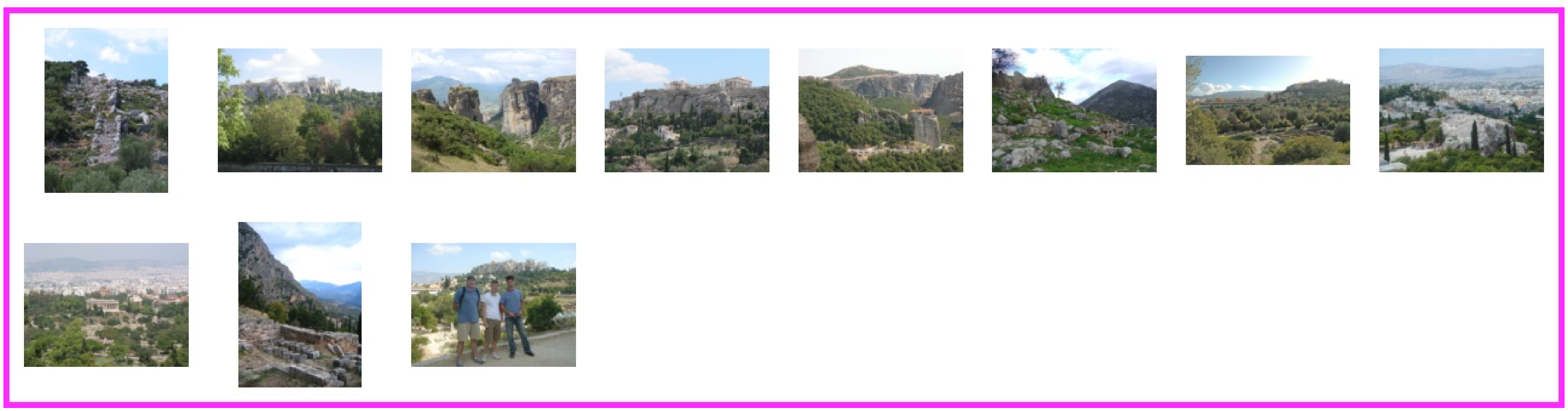
Zhang, Berg, Maire, Malik, CVPR 2006

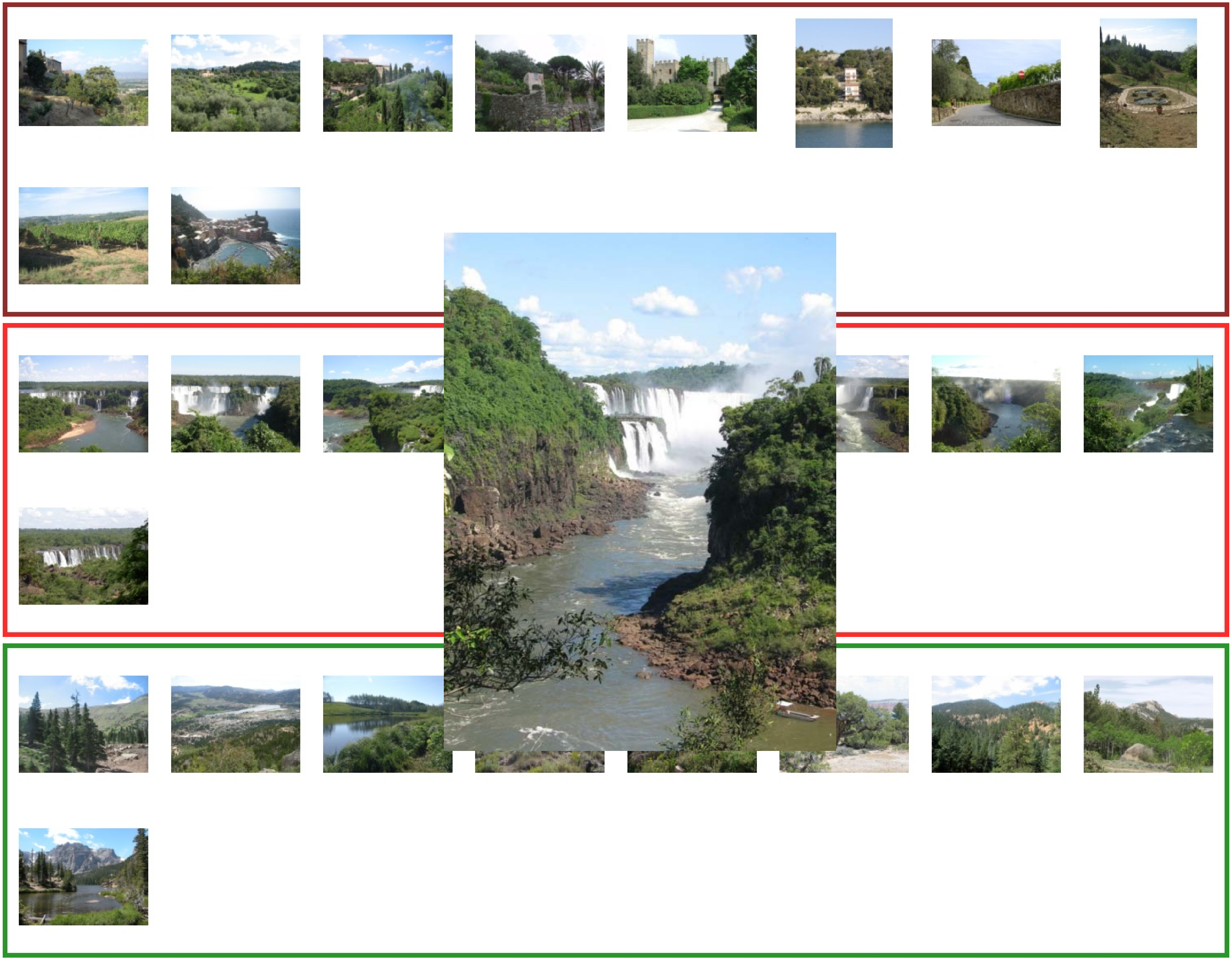
Lazy Learning

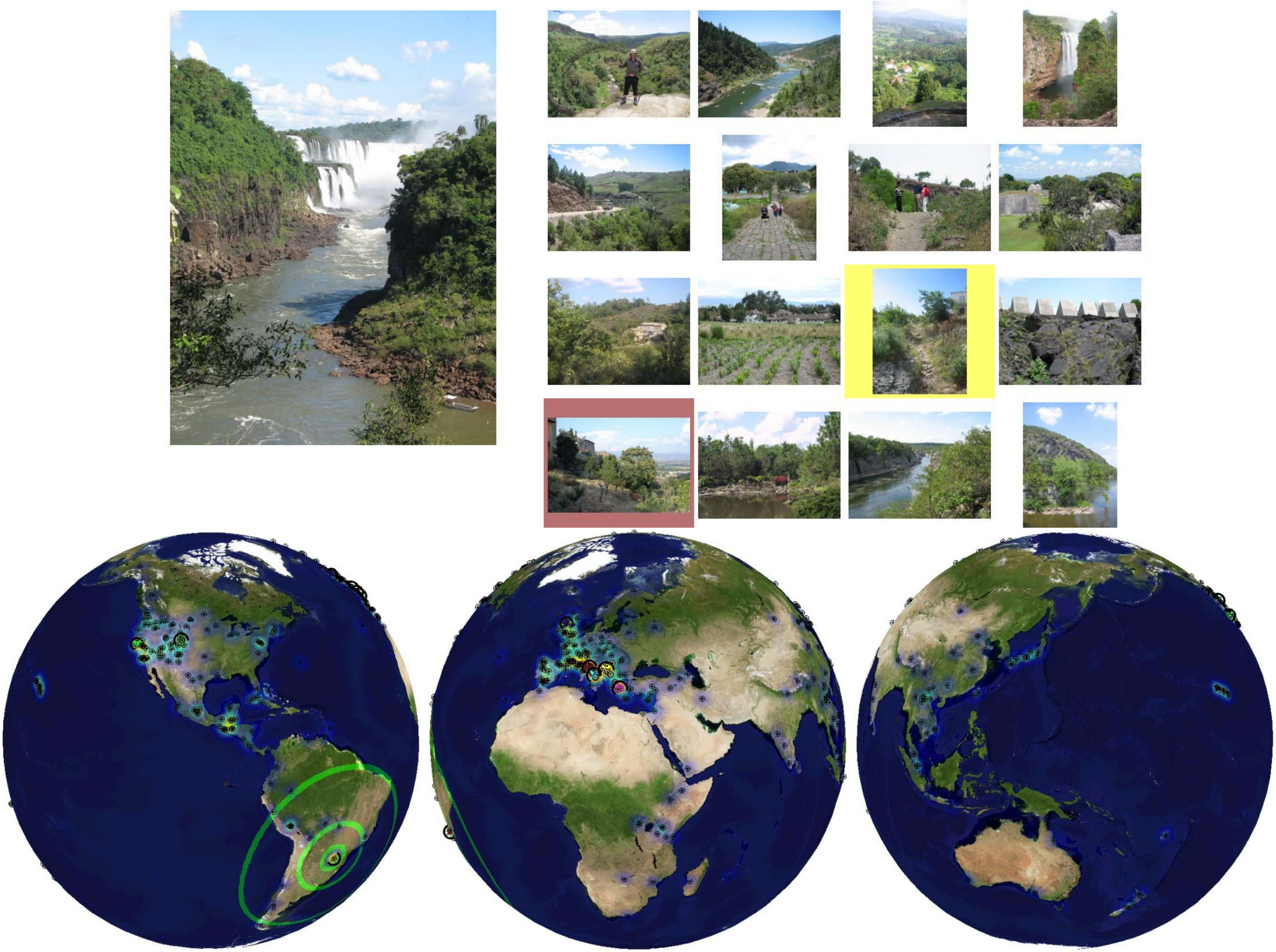
- The mean shift clusters will be our classes (e.g. London class, Seattle class, Kenya class).
- This leaves us an N-way classification problem. For each cluster, we find the non-linear decision boundary which maximally separates those cluster images from all other nearest neighbors.
- We use a kernelized version of our chi squared and L1 distances among the top 400 nearest neighbors





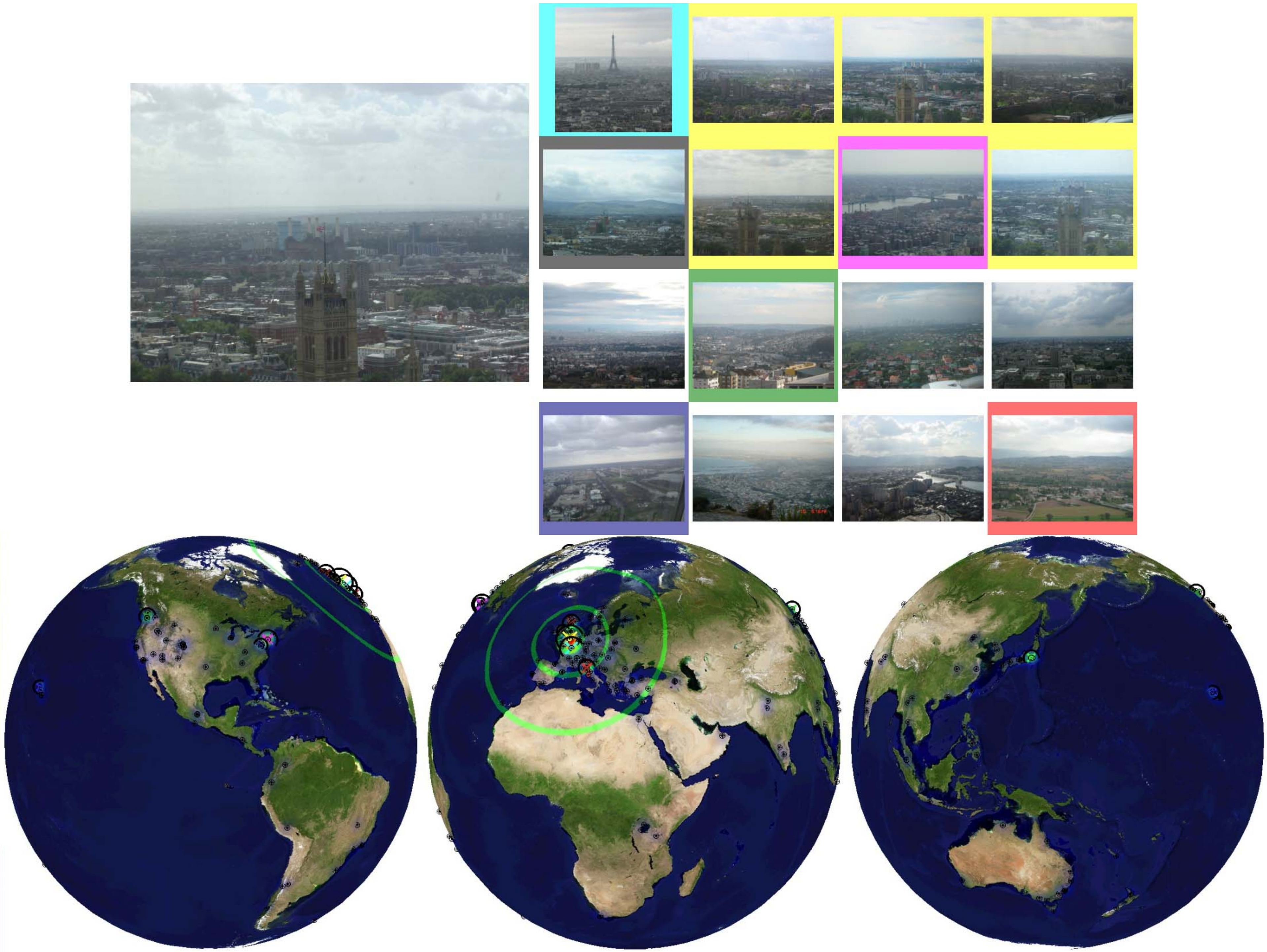


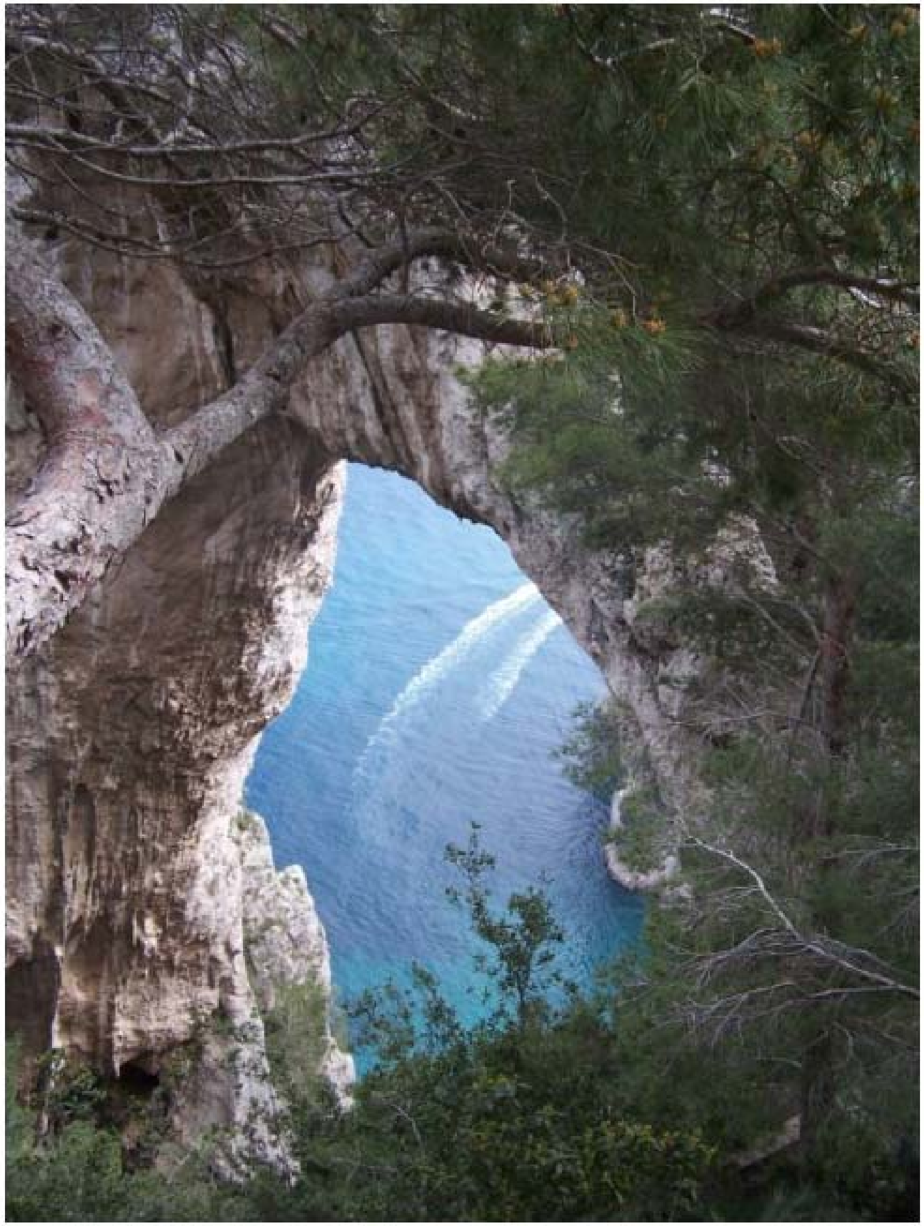


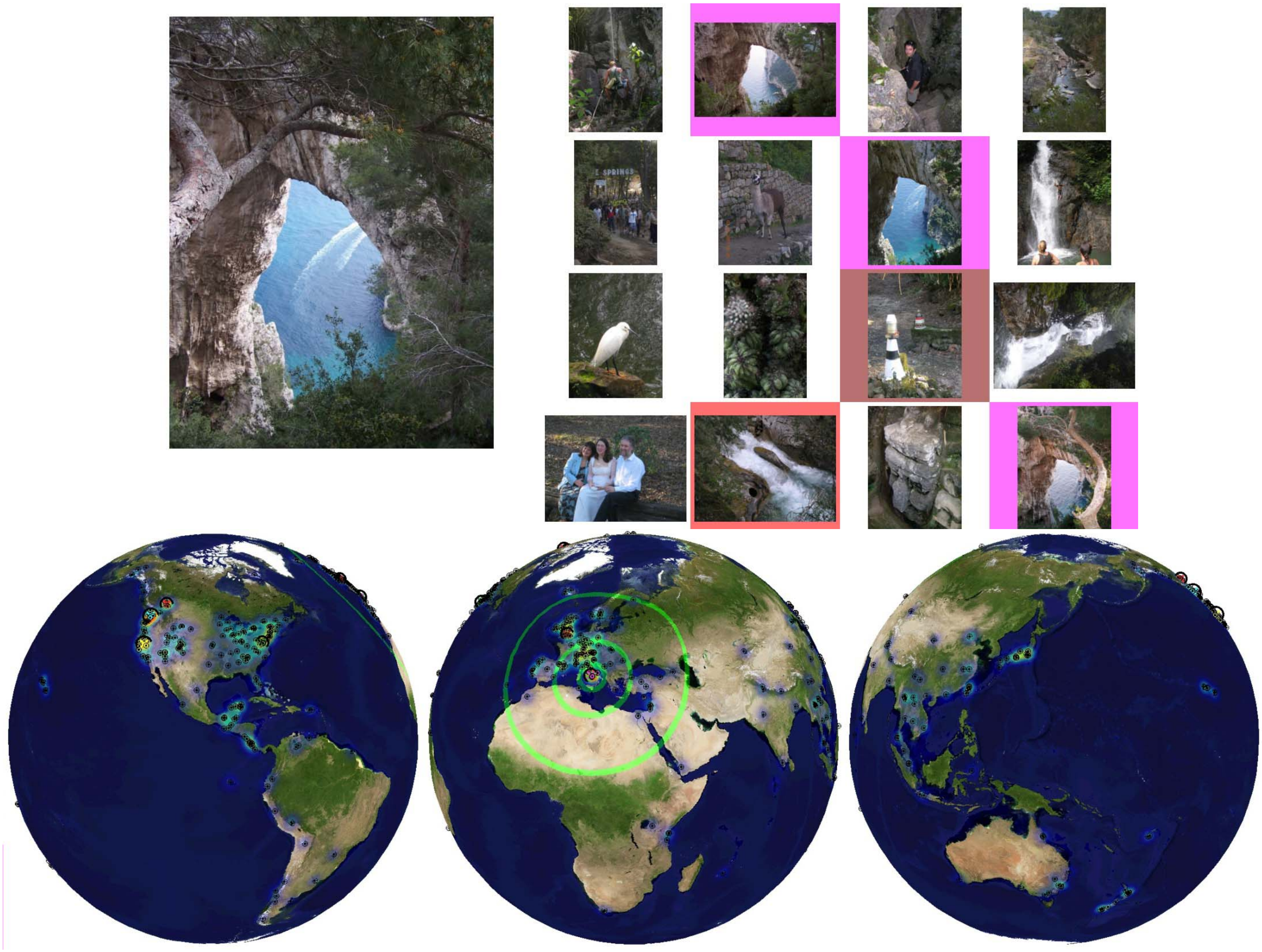


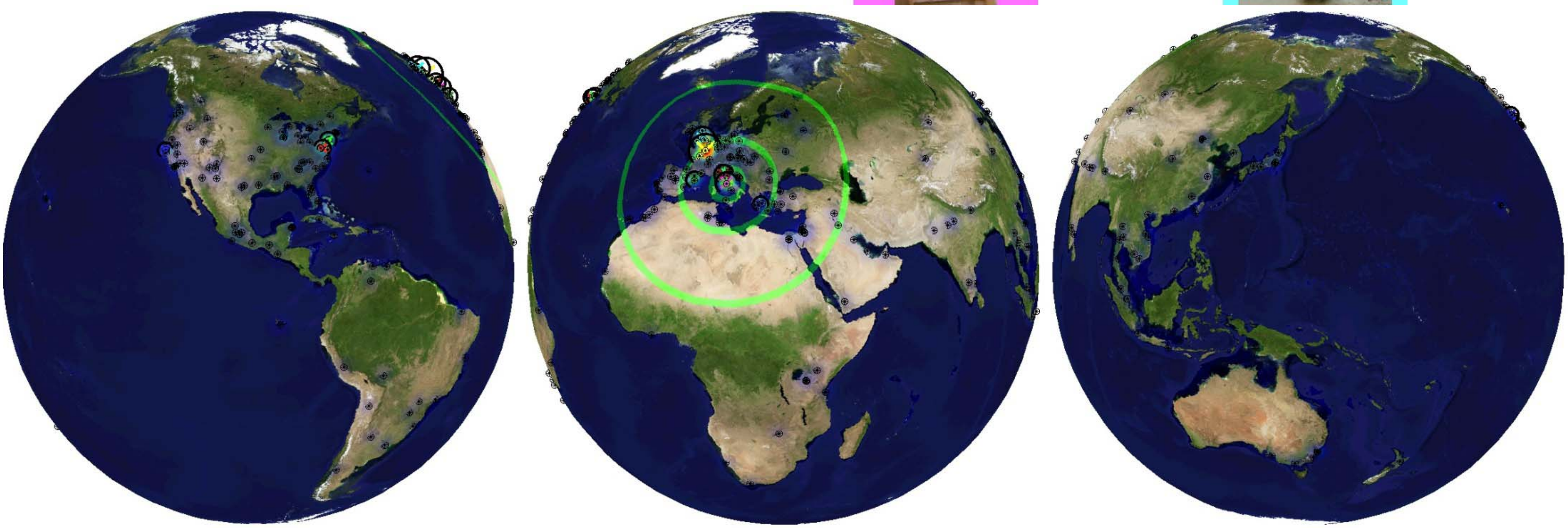
Lazy Learning

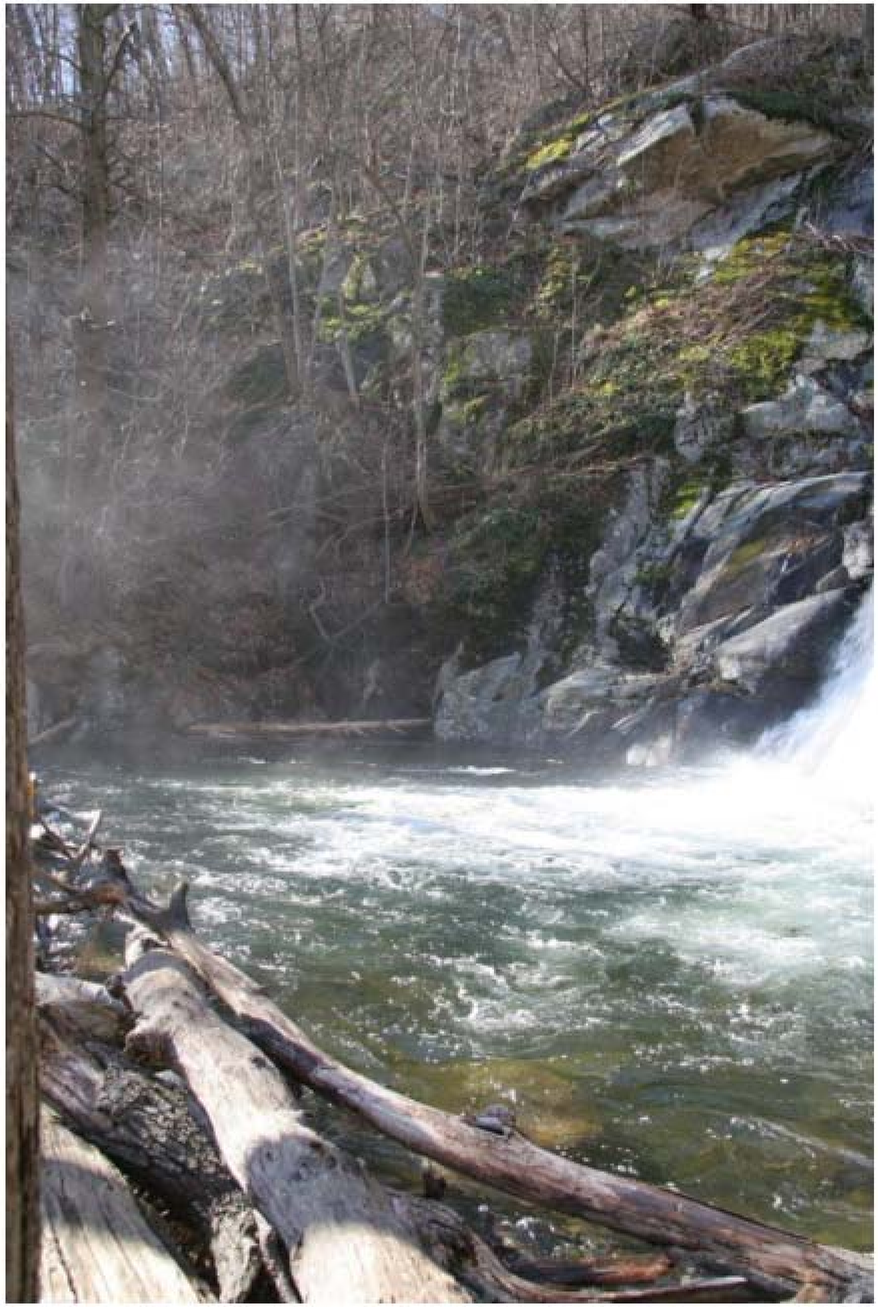
- Avoids computational expense of all-pairs distance computation for SVM training.
- Relies on the nearest neighbor search to retrieve reasonable candidates.

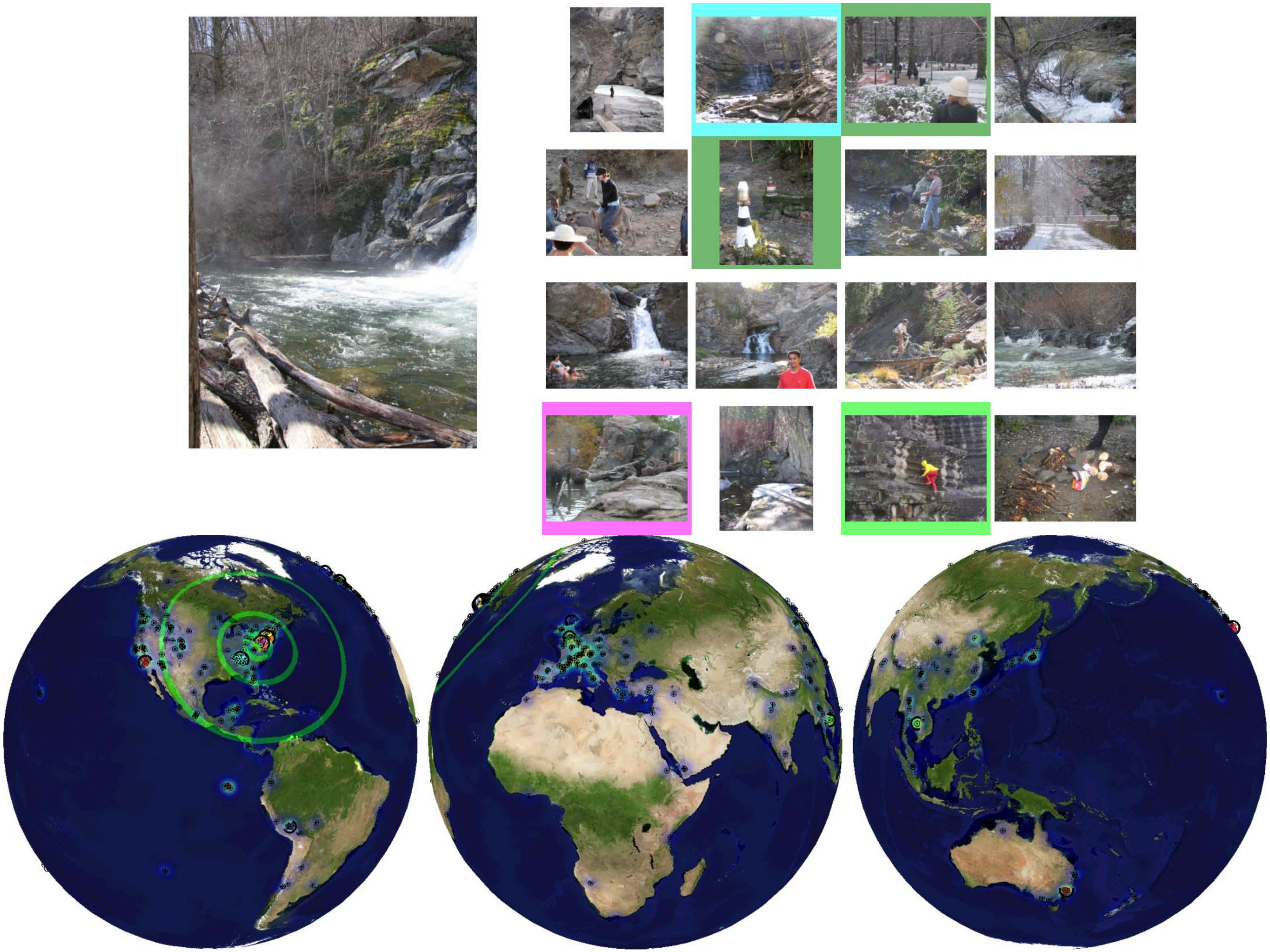


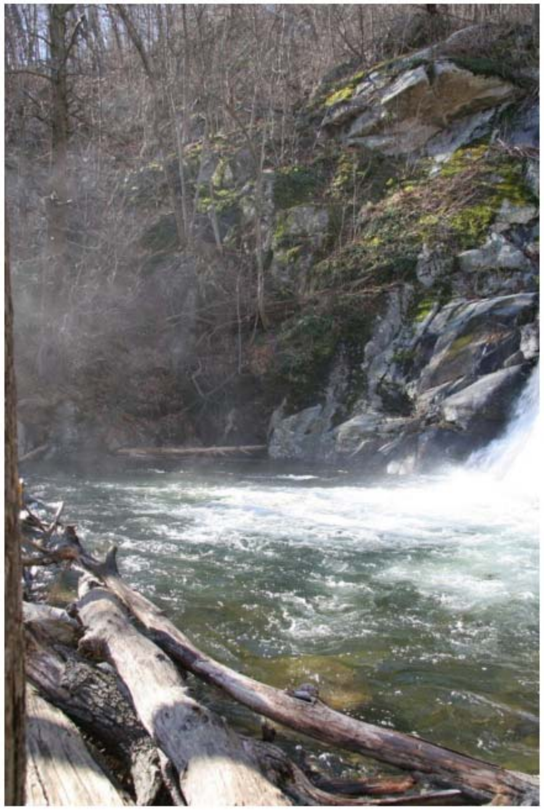












Quantitative Results

Geolocation Strategy	*Accuracy*
• Im2gps features, 1nn	6.35%
• Sift features, 1nn	8.05%
• All new features, 1nn	8.85%
• All new features, Largest Cluster	7.00%
• All new features, Lazy SVM	12.50%

Performance at Different Scales

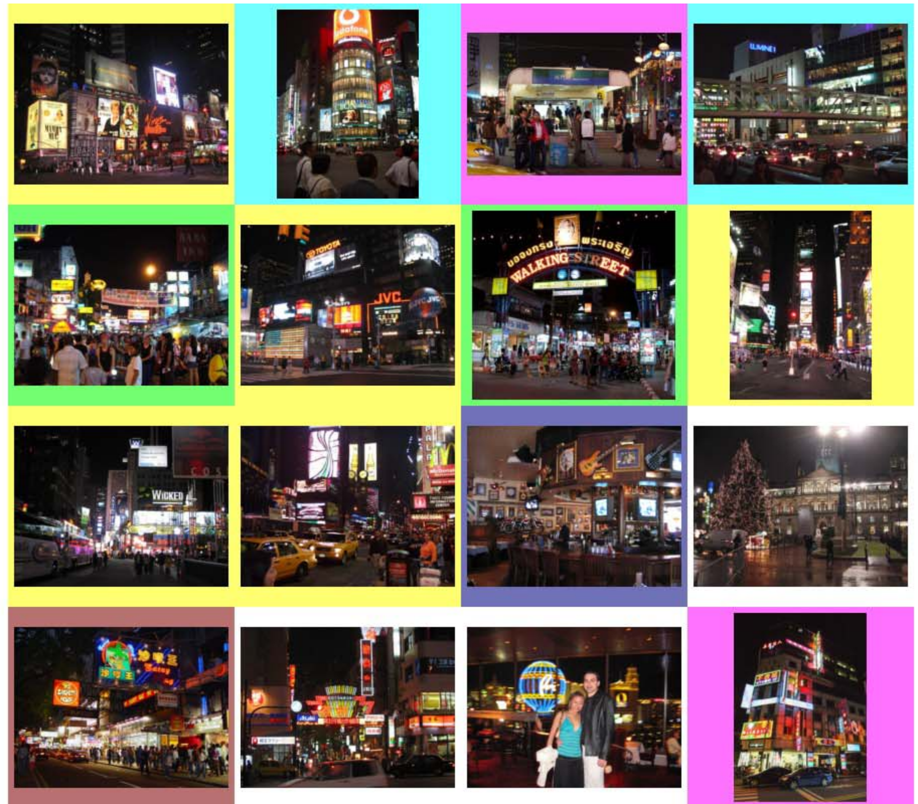
Distance Threshold	*Accuracy*
• 50km	9.15%
• 200km	12.50%
• 400km	15.95%
• 1200km	27.20%
• 3000km	38.05%

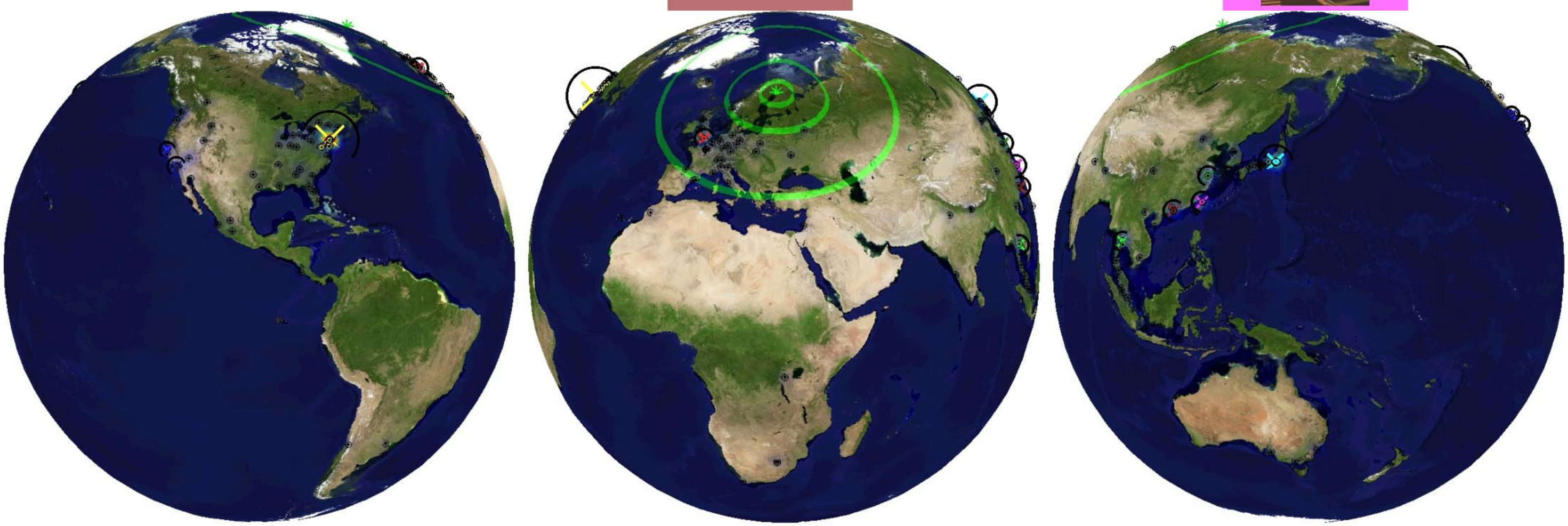
Remaining issues

- I’ve defined an overcomplete feature set. How many of these features are really necessary?
- Better parameter fitting on validation set
  - Min cluster size
  - Number of nearest neighbors
  - Mean shift bandwidth
  - What type of kernel should I use
- Geolocation is very rarely correct in sparsely photographed areas.

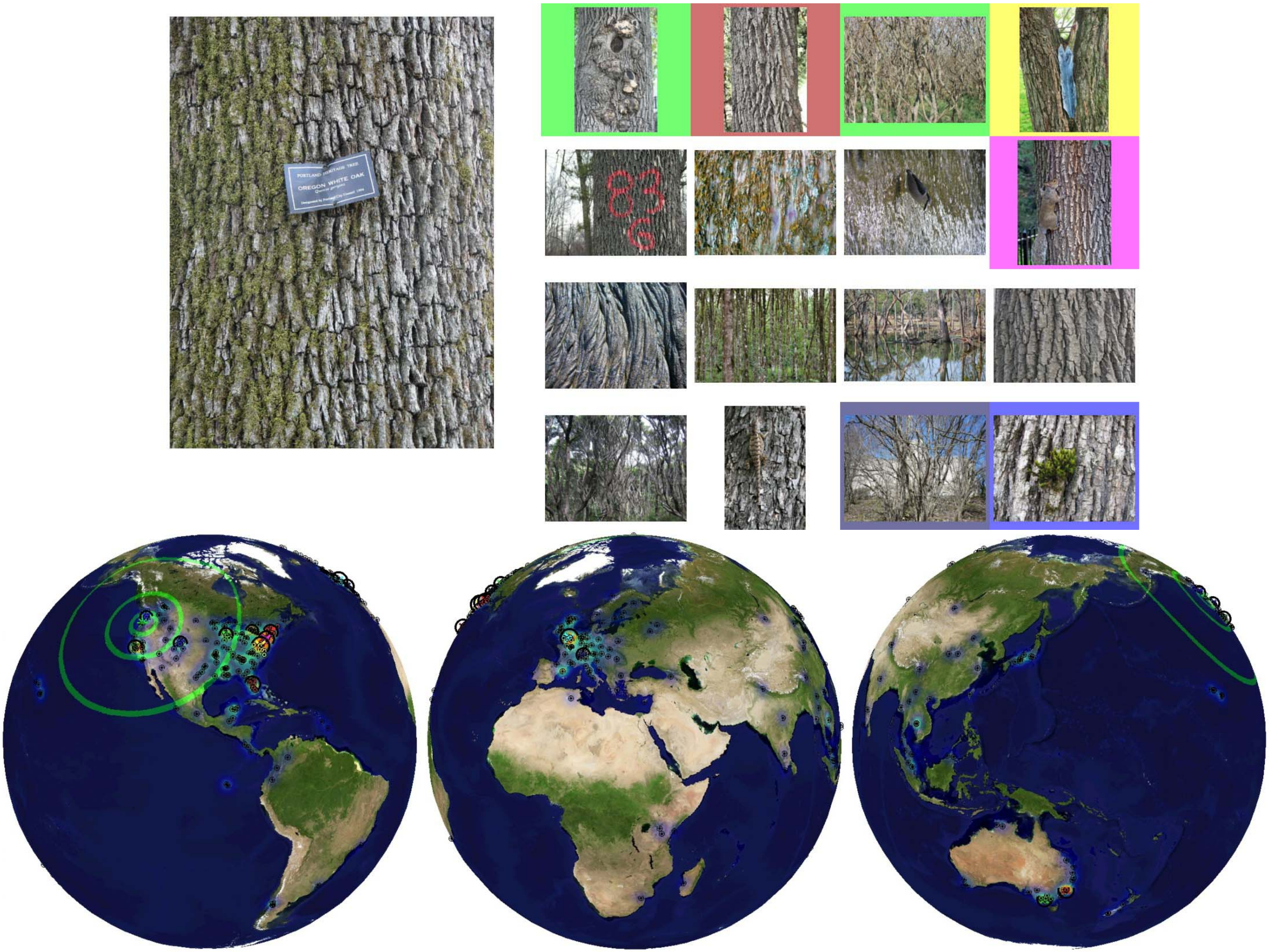

SANTA'S
GIFT SHOP


SANTA'S GIFT SHOP

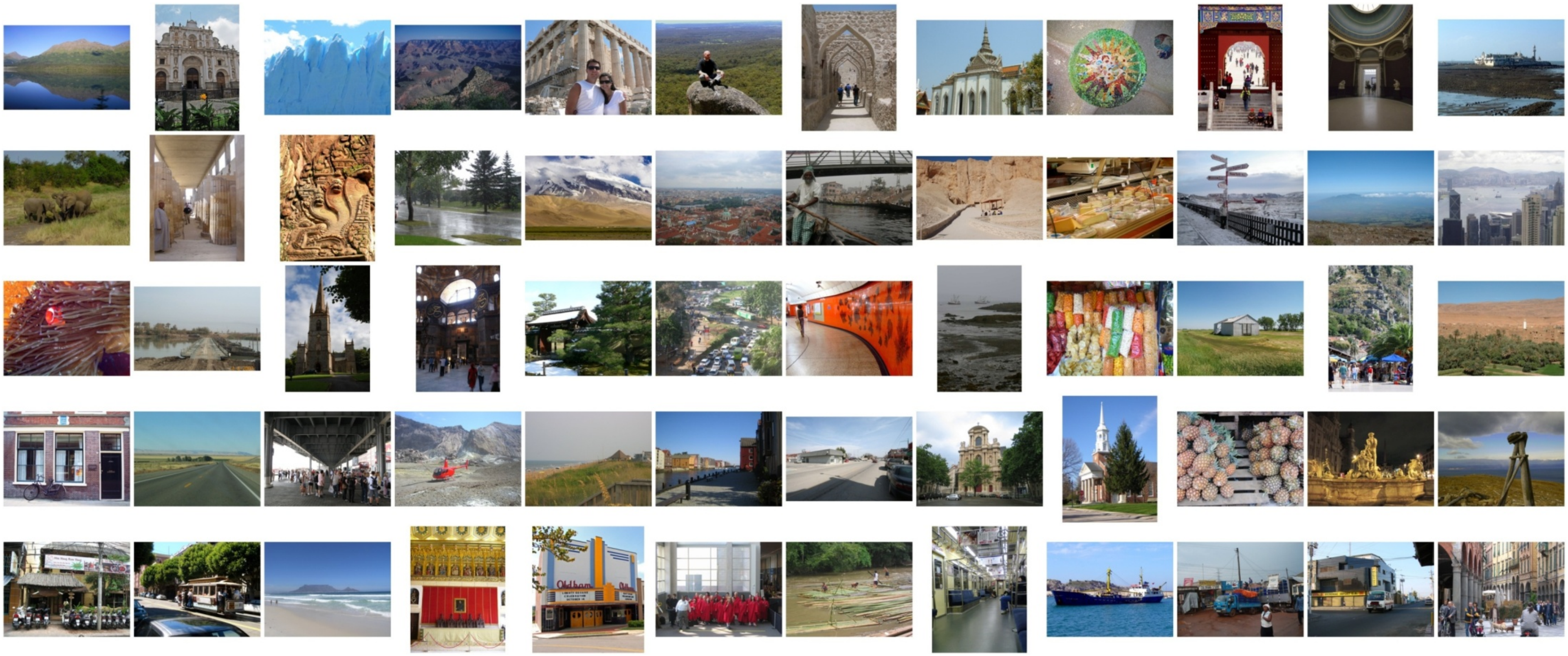

TOYOTA
JVC
WALKING STREET
WICKED




PORTLAND HERITAGE TREE
OREGON WHITE OAK


PORTLAND HERITAGE TREE
OREGON WHITE OAK
83
6

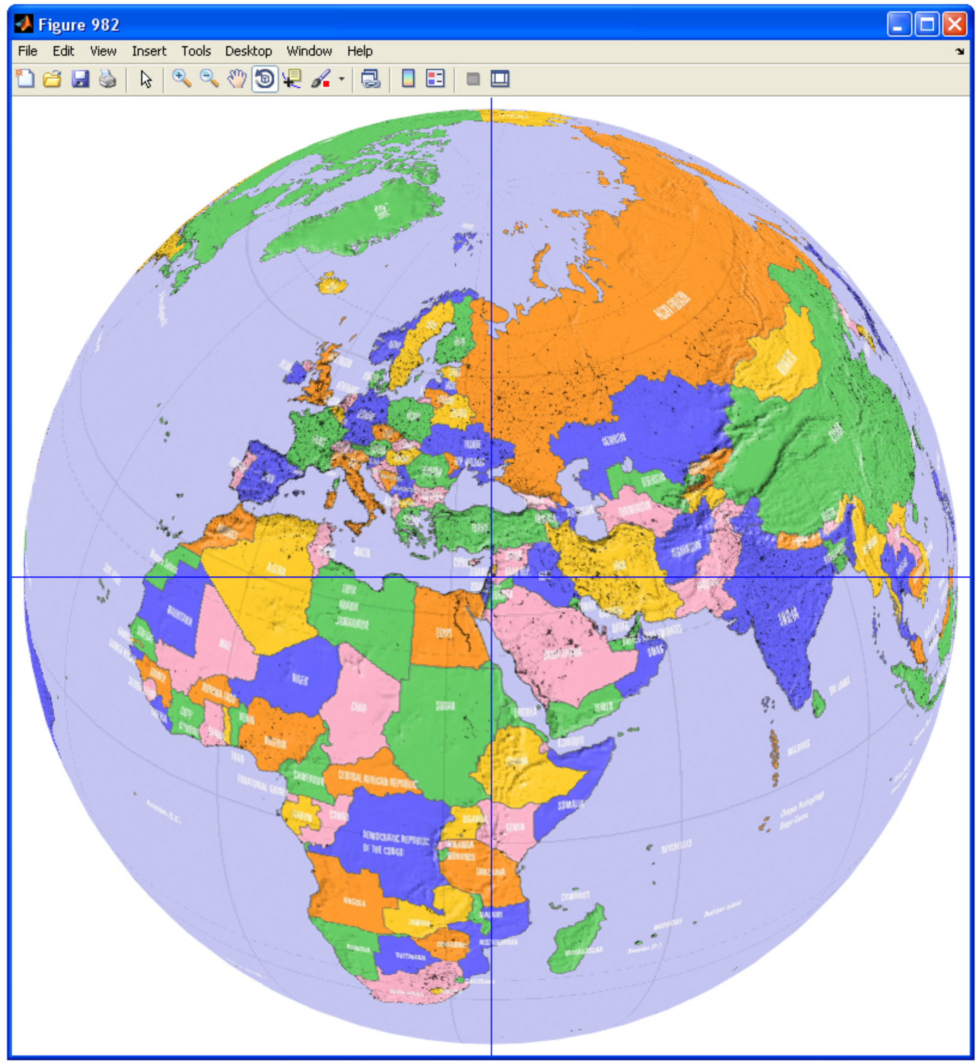
Human Geolocation of Images



Test set: 64 images pruned from a random sample

Hays and Efros. Where in the World? Human and Computer Geolocation of Images. VSS 2009.

20 participants were shown 64 photos and asked to guess the location where each photo was taken.

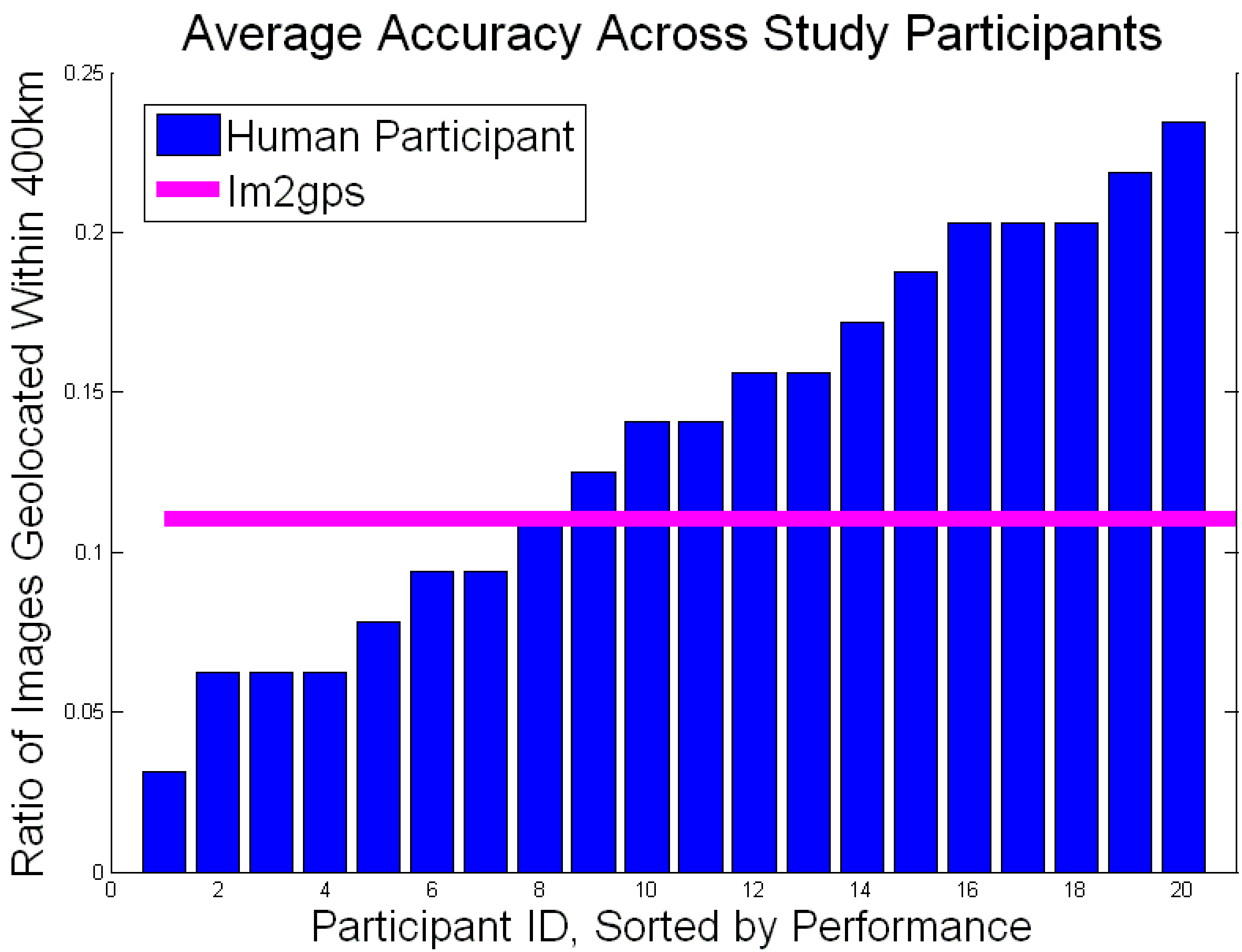


The first set of 32 images allow unlimited viewing. The final 32 images are flashed for only 100 milliseconds.

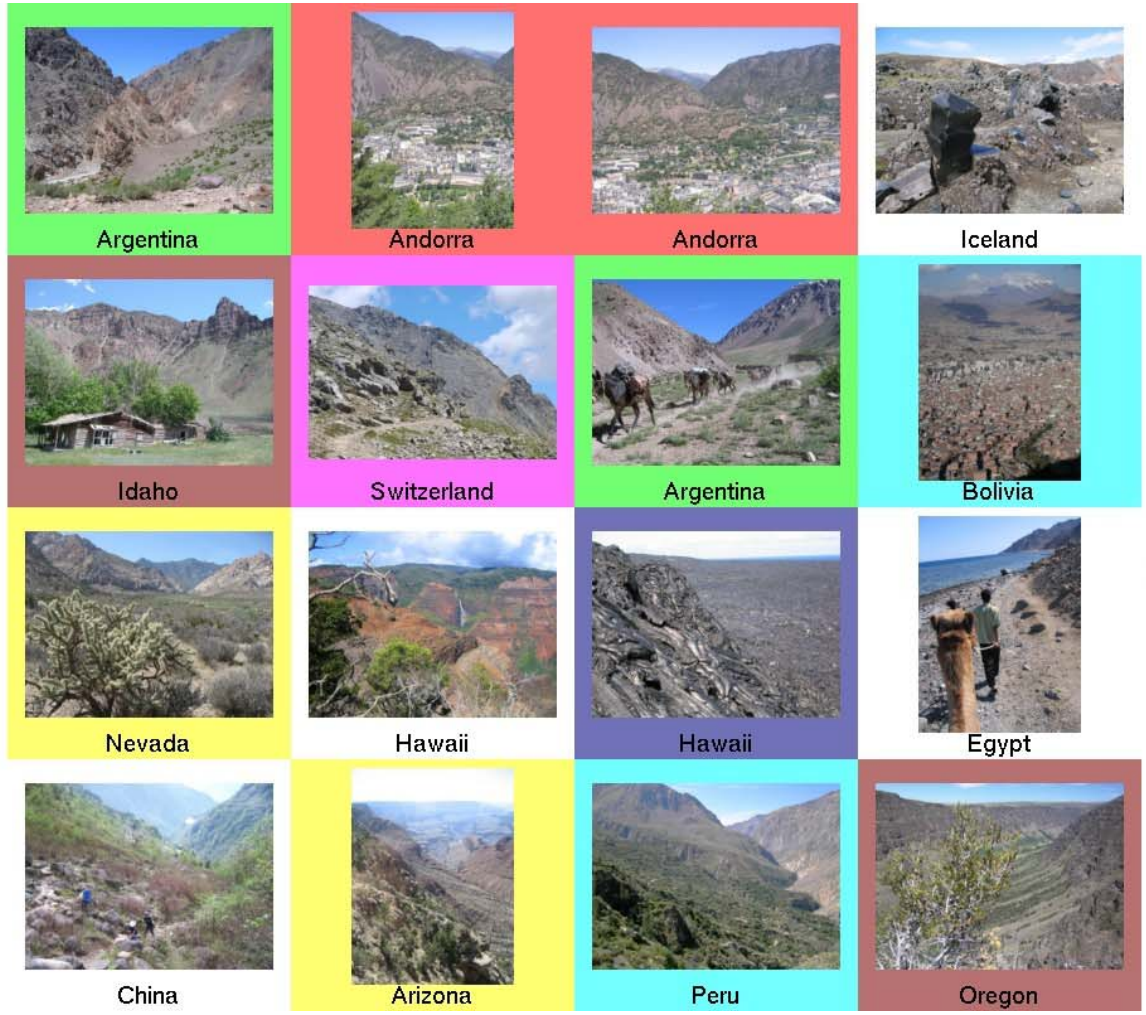
Countdown 100ms Blank

The image sets alternate viewing conditions. No learning or fatigue effects were observed.

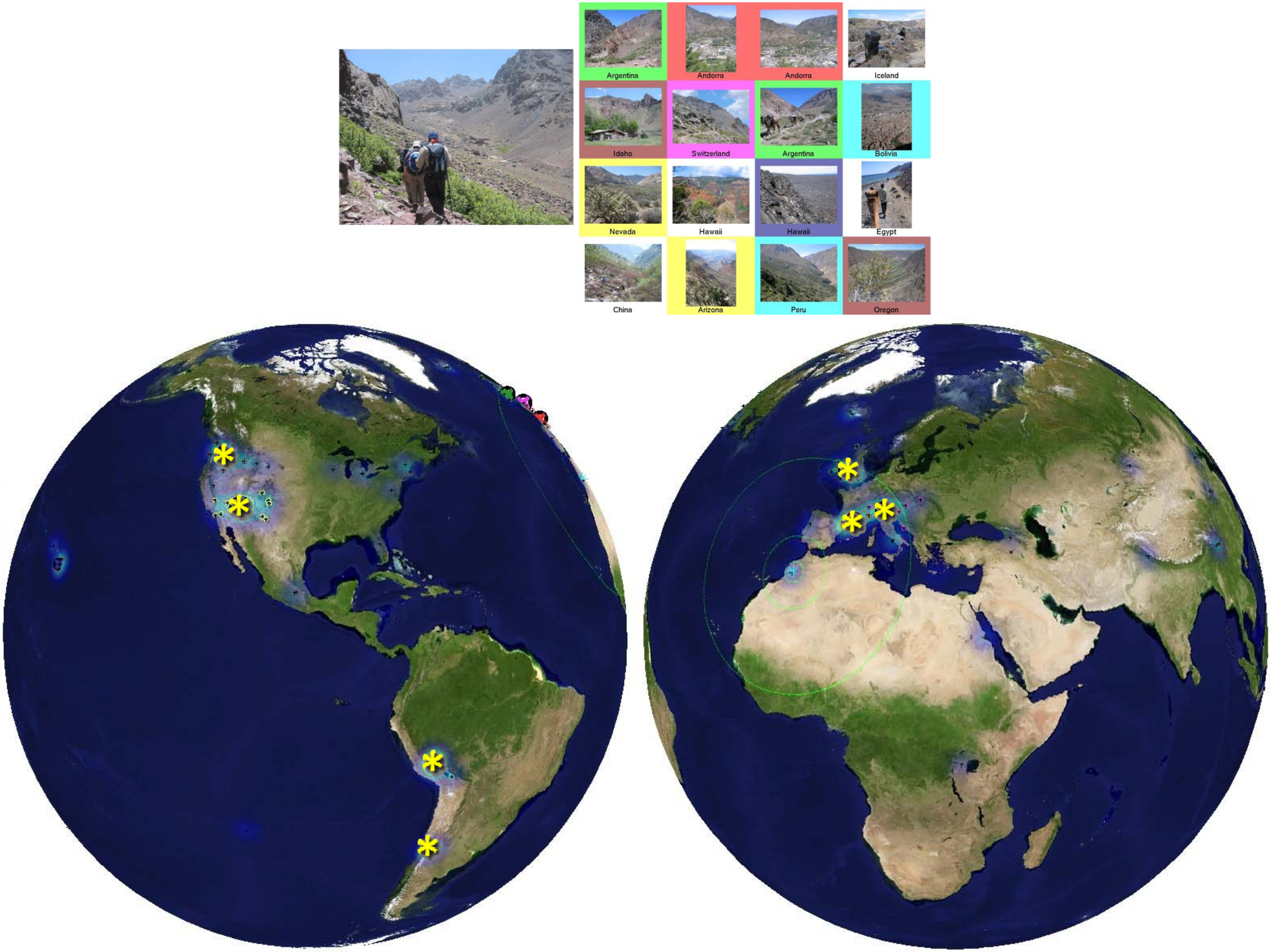
New Features and Learning 31%



Estimating Geographic Properties

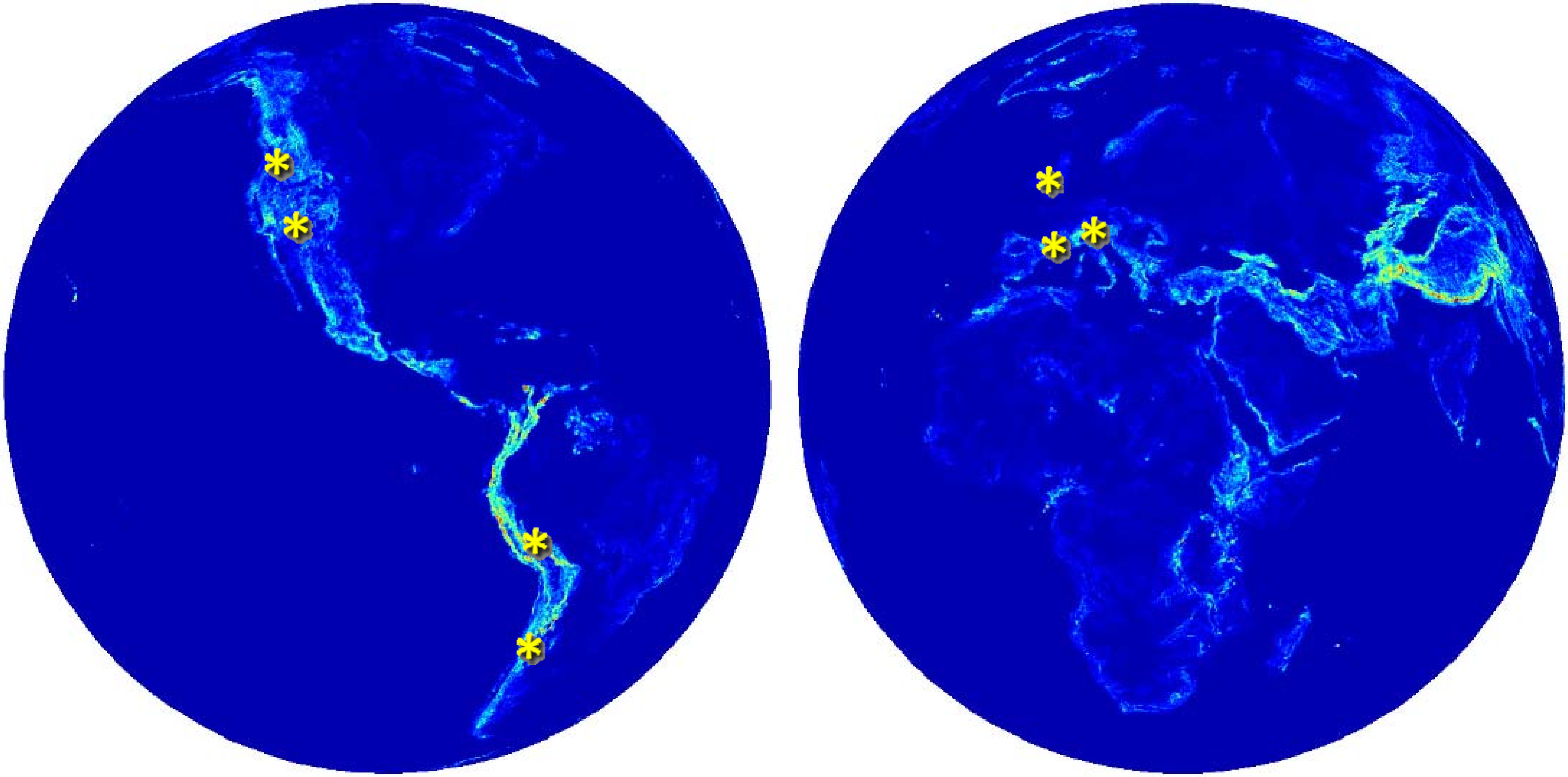


Argentina
Andorra
Andorra
Iceland
Idaho
Switzerland
Argentina
Bolivia
Nevada
Hawaii
Hawaii
Egypt
China
Arizona
Peru
Oregon

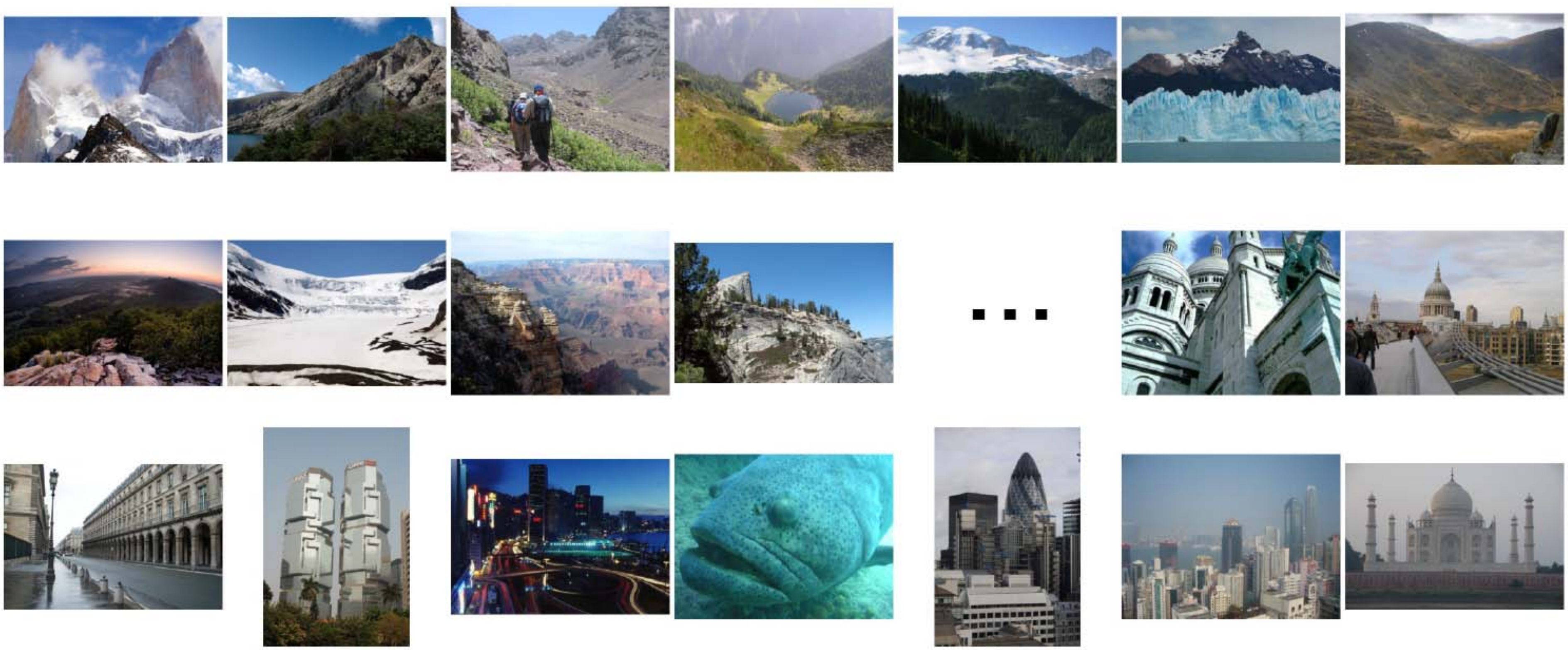


Argentina
Andorra
Andorra
Iceland
Idaho
Switzerland
Argentina
Bolivia
Nevada
Hawaii
Hawaii
Egypt
China
Arizona
Peru
Oregon

112 m / km



Elevation gradient magnitude ranking



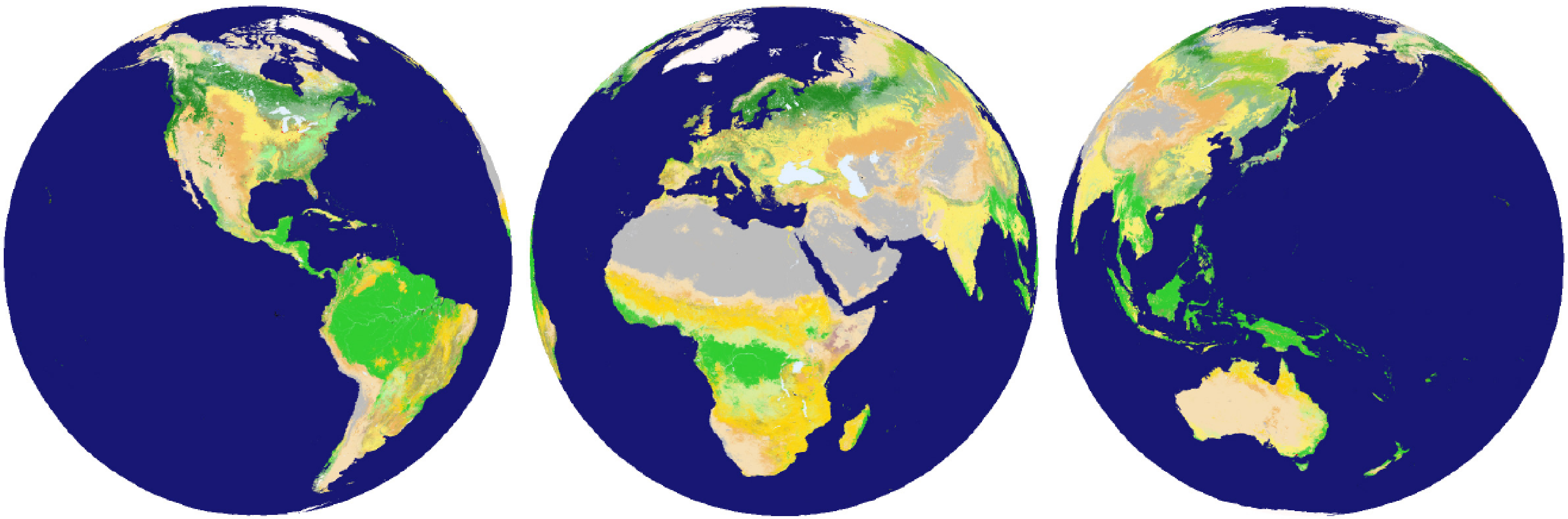
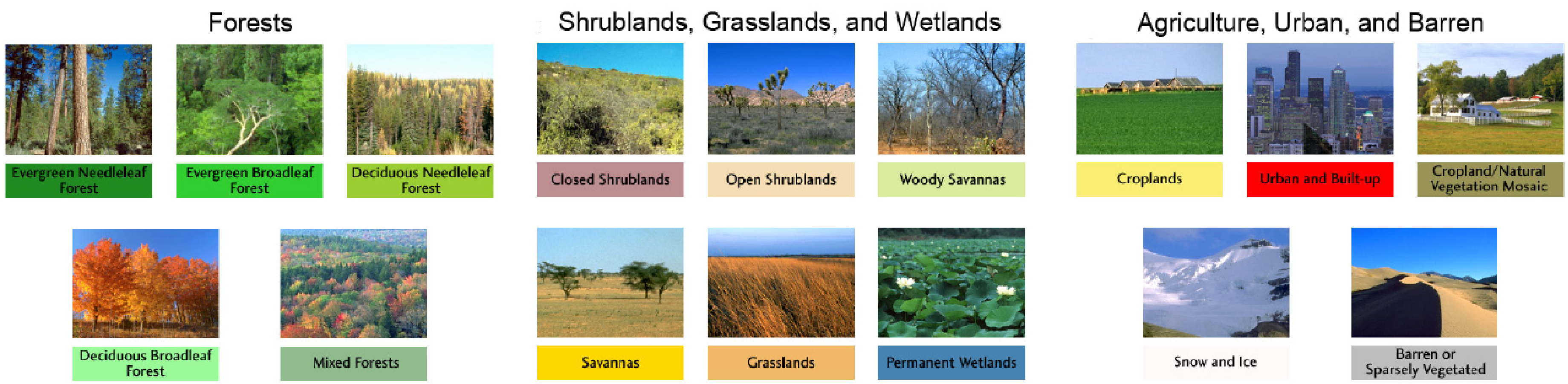


Figure 4. Global land cover classification map.



Forest

Geographic and Keyword Context for Object Detection

- Divvala, Hoiem, Hays, Efros, Herbert. *An Empirical Study of Context in Object Detection.* CVPR 2009. **Poster Session 3. Tuesday, June 23**

Goal: Better Detection using Scene Context

Dining Table

Context from Scene Match Properties

Scene Matches

• • •

Associated **Tags**/**Geo**

Zoo …
Bear

Urban …
High Pop

Ruins …
Maya

Forest …
Low Pop

2007 …
Riodejaneiro

Urban…
High Pop

Features for "Dining Table" Classifier

Keywords

Table: 0
Food: 0
Café: 1
Animals: 3
Nature: 4
Monument: 3
Hiking: 3
Verdict: Unlikely

Geography

Photo Density: 0.5
Pop Density: 0.6
Forest: 0.4
RainForest: 0.2
Verdict: Not Sure

Geographic context for object detection

- With the entire system (which I haven't discussed) we raise average precision from .182 to .194
- VOC 2008: Best performance in 6 of 20 categories, second best in 6 categories.

Geographic context for object detection

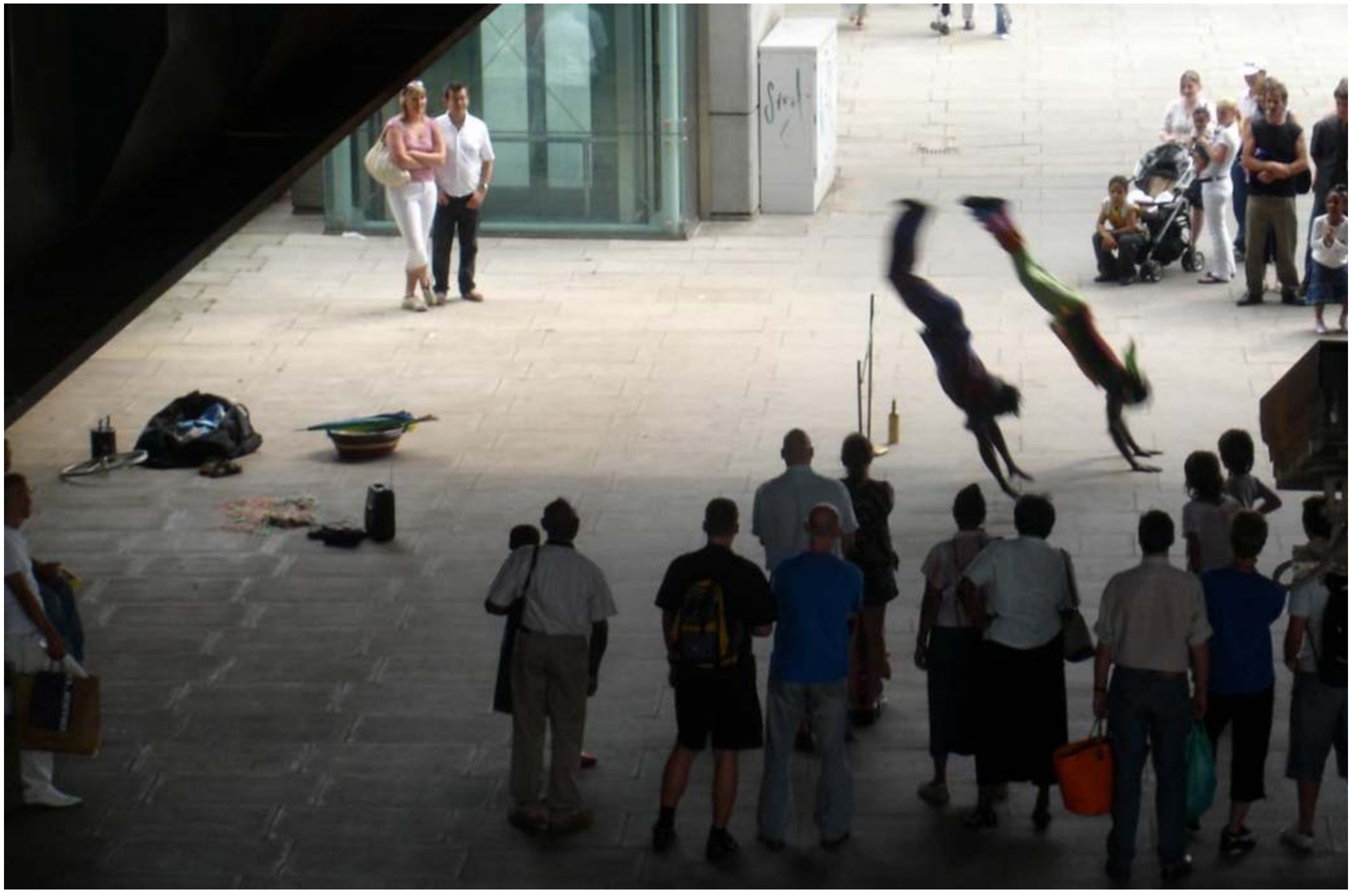
- Geography properties alone get an average precision of ~.15 on Pascal VOC 2008 Object Classification Challenge.
- Keyword histograms alone get an average precision of ~.26.

- Divvala, Hoiem, Hays, Efros, Herbert. *An Empirical Study of Context in Object Detection.* CVPR 2009. **Poster Session 3. Tuesday, June 23**

Image Sequence Geolocation

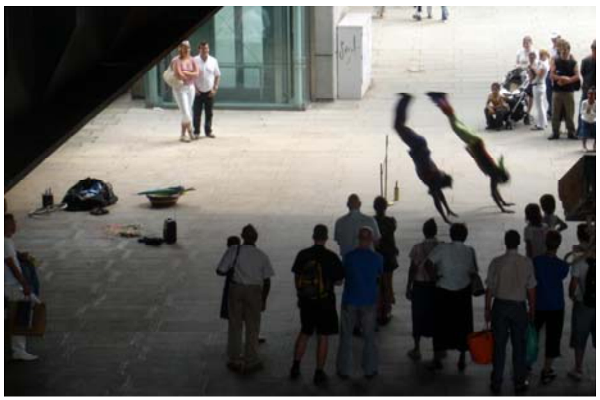
Kalogerakis, Vesselova, Hays, Efros, Hertzmann.
Image Sequence Geolocation with Human Travel Priors.
ICCV 2009

Where is This?



Where is This?

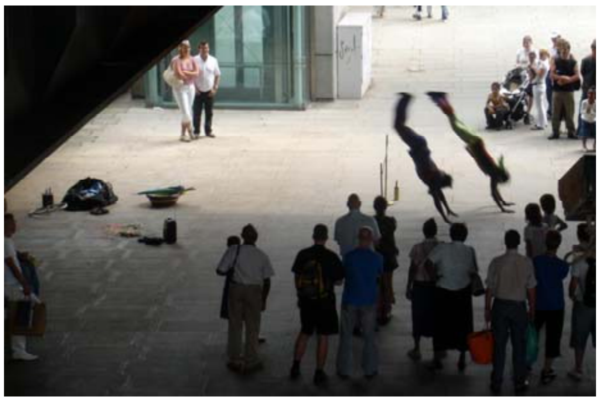
Where are These?



15:14,
June 18th, 2006

16:31,
June 18th, 2006

Where are These?



15:14,
June 18th, 2006

16:31,
June 18th, 2006

17:24,
June 19th, 2006

Image Sequence Geolocation with Human Travel Priors

From a single-image accuracy of 10% we reach 58% by letting geolocation estimates reinforce each other *temporally*.