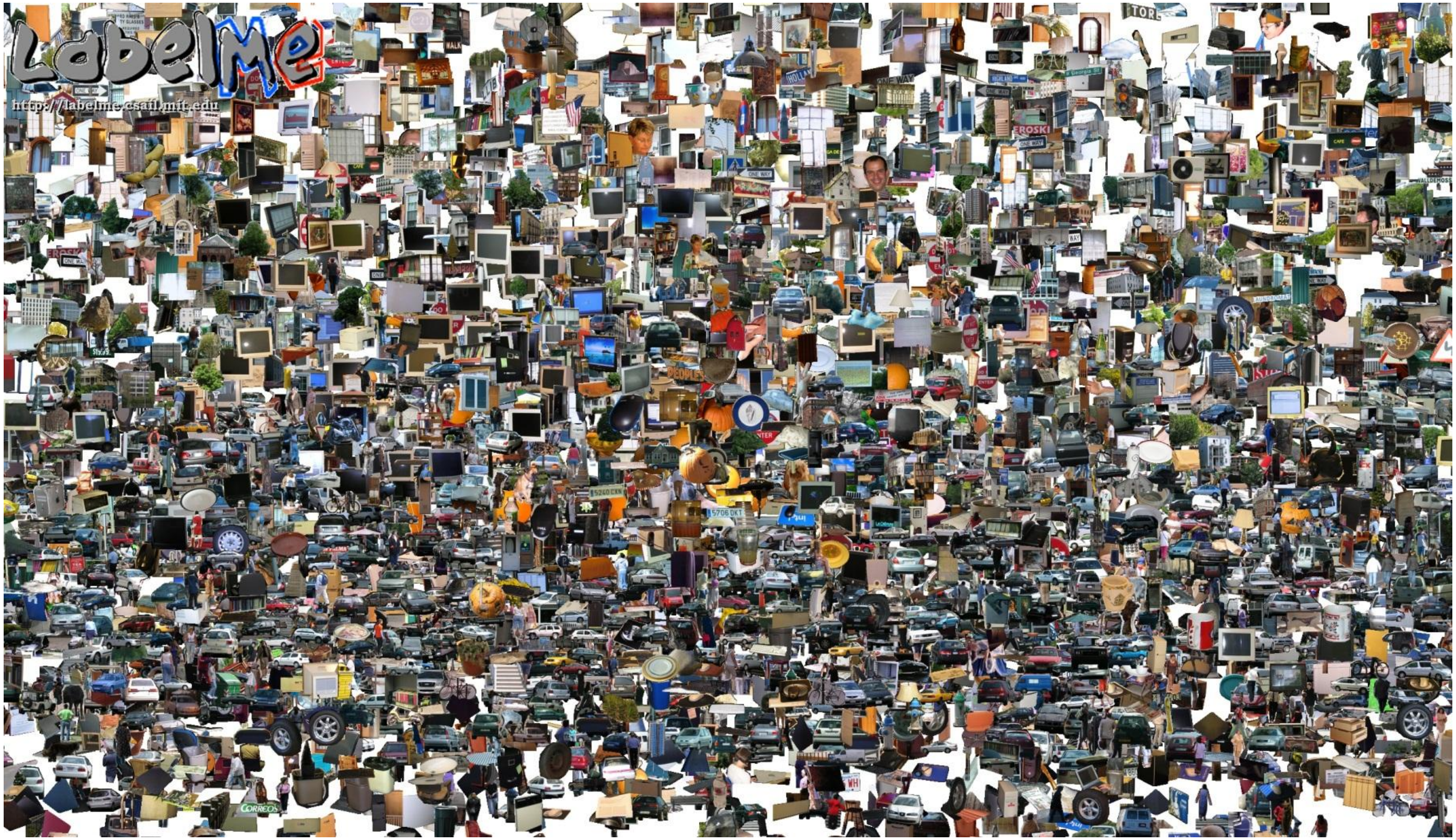
Previous Lecture

- Action recognition in videos and images

Next Lecture

- Guest talk by Pedro Felzenszwalb on Object Detection

Opportunities of Scale



Computer Vision

James Hays, Brown

Many slides from James Hays, Alyosha Efros, and Derek Hoiem

Graphic from Antonio Torralba

Today’s class

- Opportunities of Scale: Data-driven methods
  - Scene completion
  - Im2gps
  - Recognition via Tiny Images
  - More recognition by association

Google and massive data-driven algorithms

A.I. for the postmodern world:

– all questions have already been answered…many times, in many ways

– Google is dumb, the "intelligence" is <u>in the data</u>

Google Translate

Chinese Room, John Searle (1980)

If a machine can convincingly simulate an intelligent conversation, does it necessarily understand? In the experiment, Searle imagines himself in a room, acting as a computer by manually executing a program that convincingly simulates the behavior of a native Chinese speaker.

Most of the discussion consists of attempts to refute it. "The overwhelming majority," notes *BBS* editor Stevan Harnad,“ still think that the Chinese Room Argument is dead wrong." The sheer volume of the literature that has grown up around it inspired Pat Hayes to quip that the field of cognitive science ought to be redefined as "the ongoing research program of showing Searle's Chinese Room Argument to be false.

Big Idea

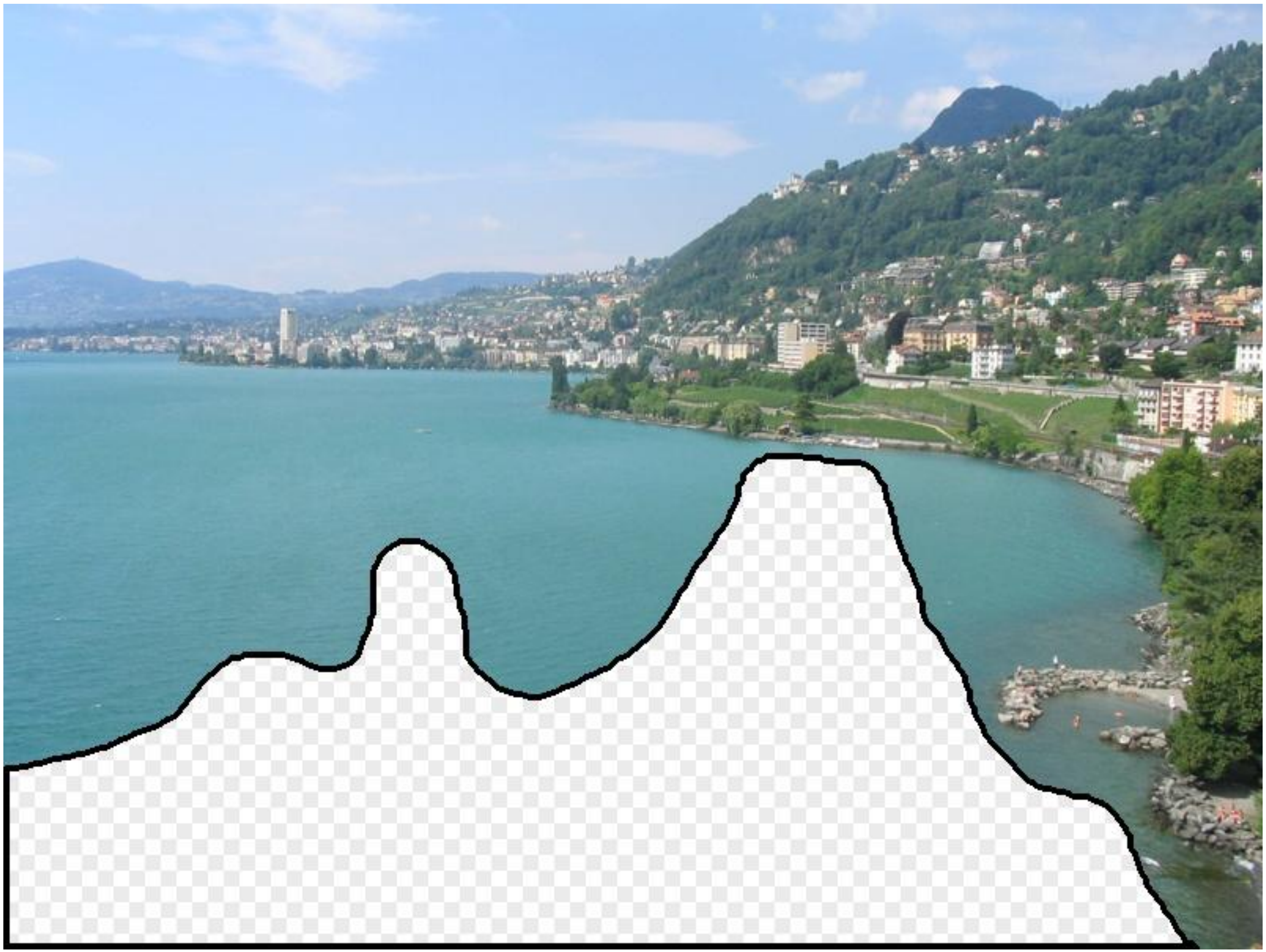
- What if invariance / generalization isn't actually the core difficulty of computer vision?
- What if we can perform high level reasoning with brute-force, data-driven algorithms?

Image Completion Example

[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]

http://graphics.cs.cmu.edu/projects/scene-completion/

What should the missing region contain?



Which is the original?

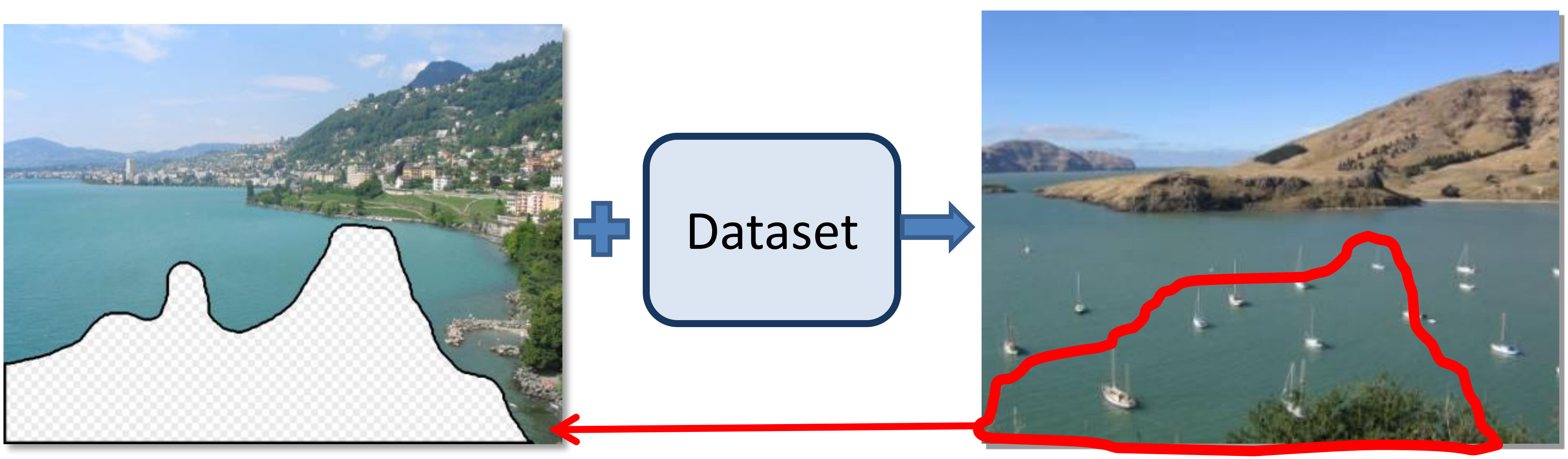
(a)

(b)

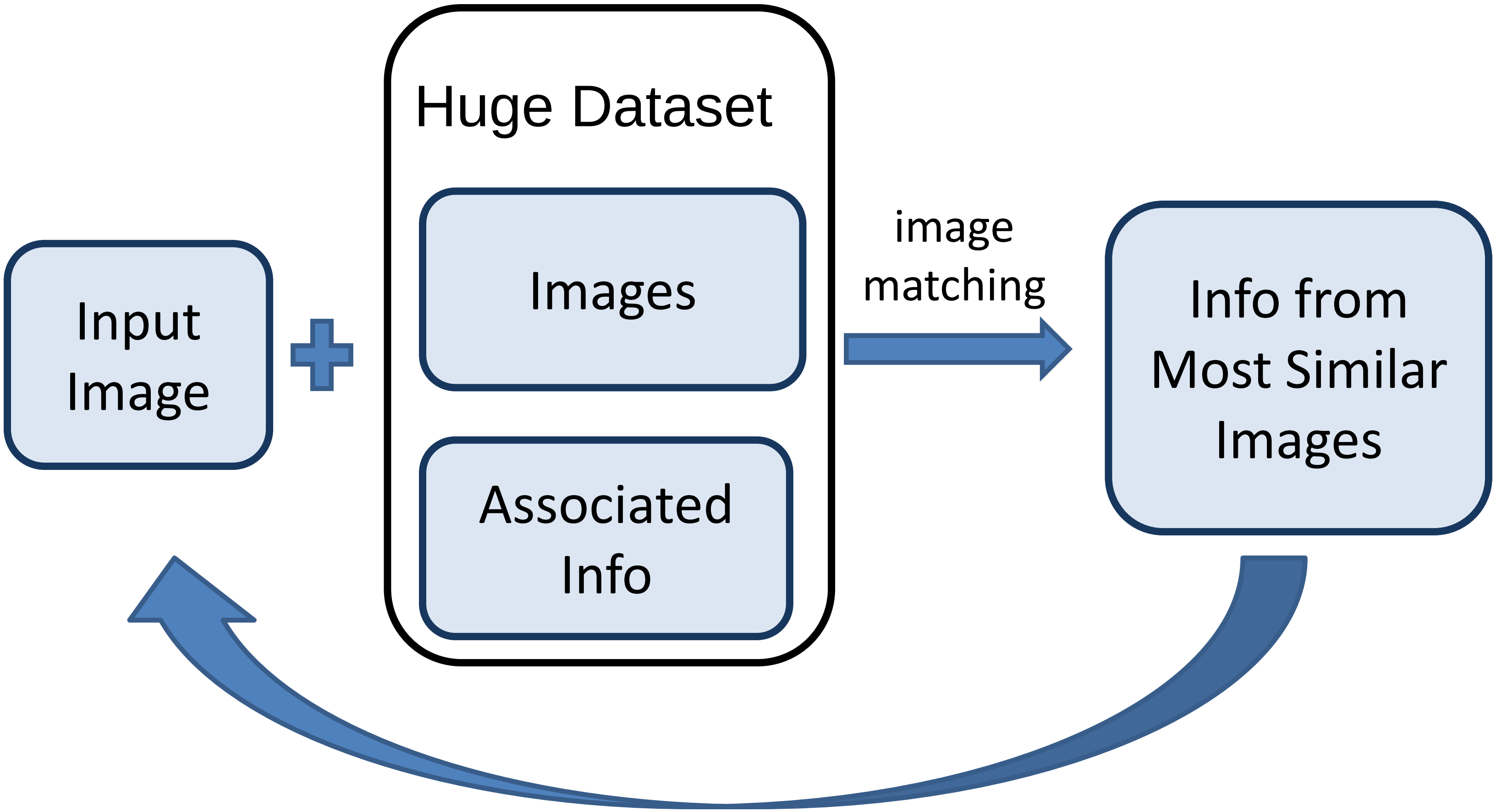
(c)

How it works

- Find a similar image from a large dataset
- Blend a region from that image into the hole

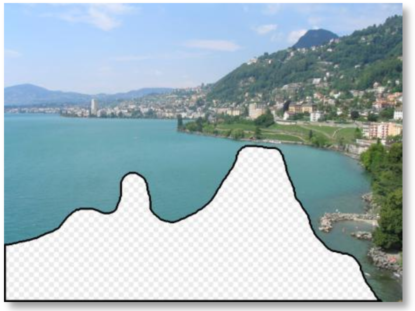


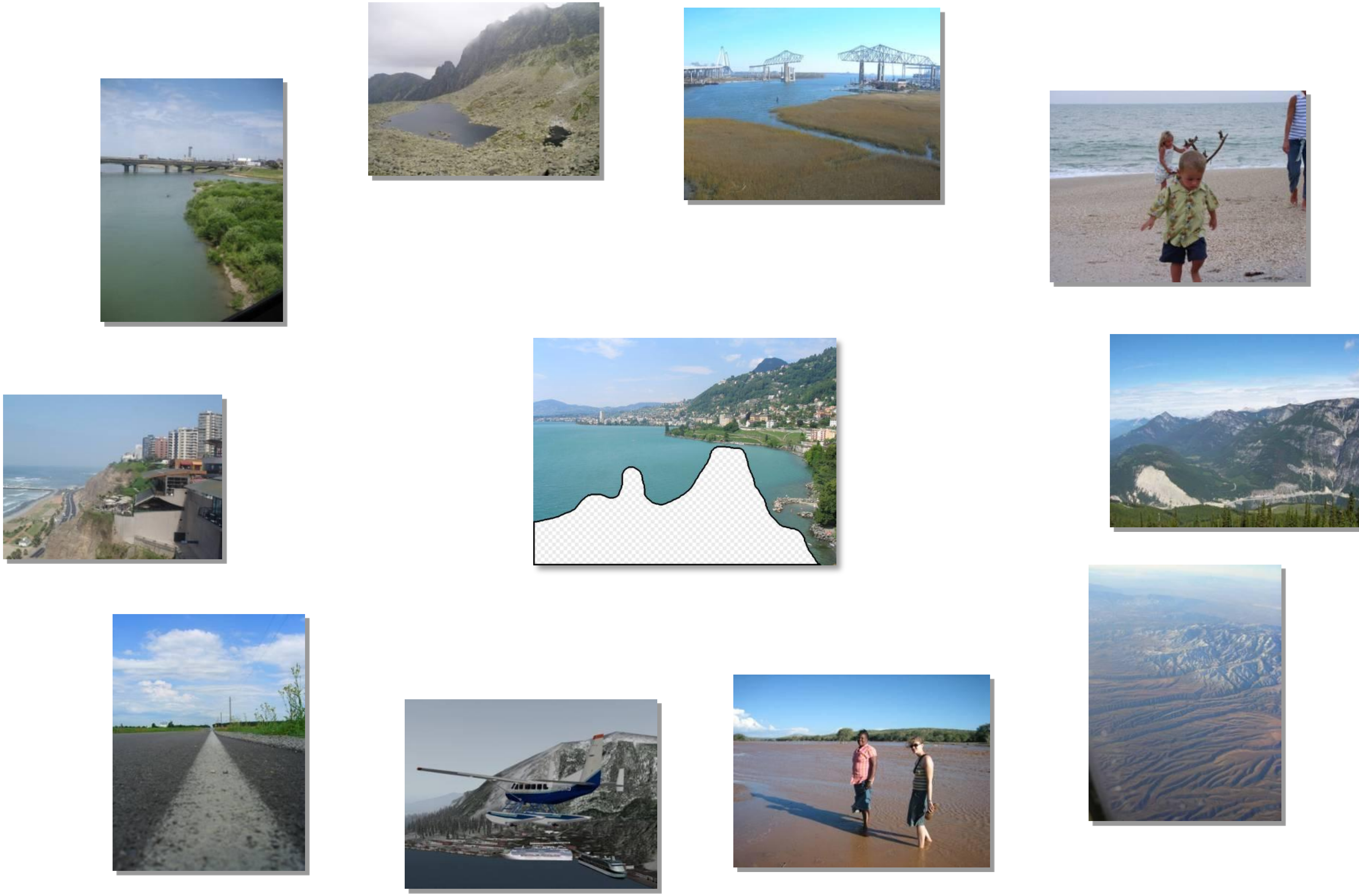
General Principal



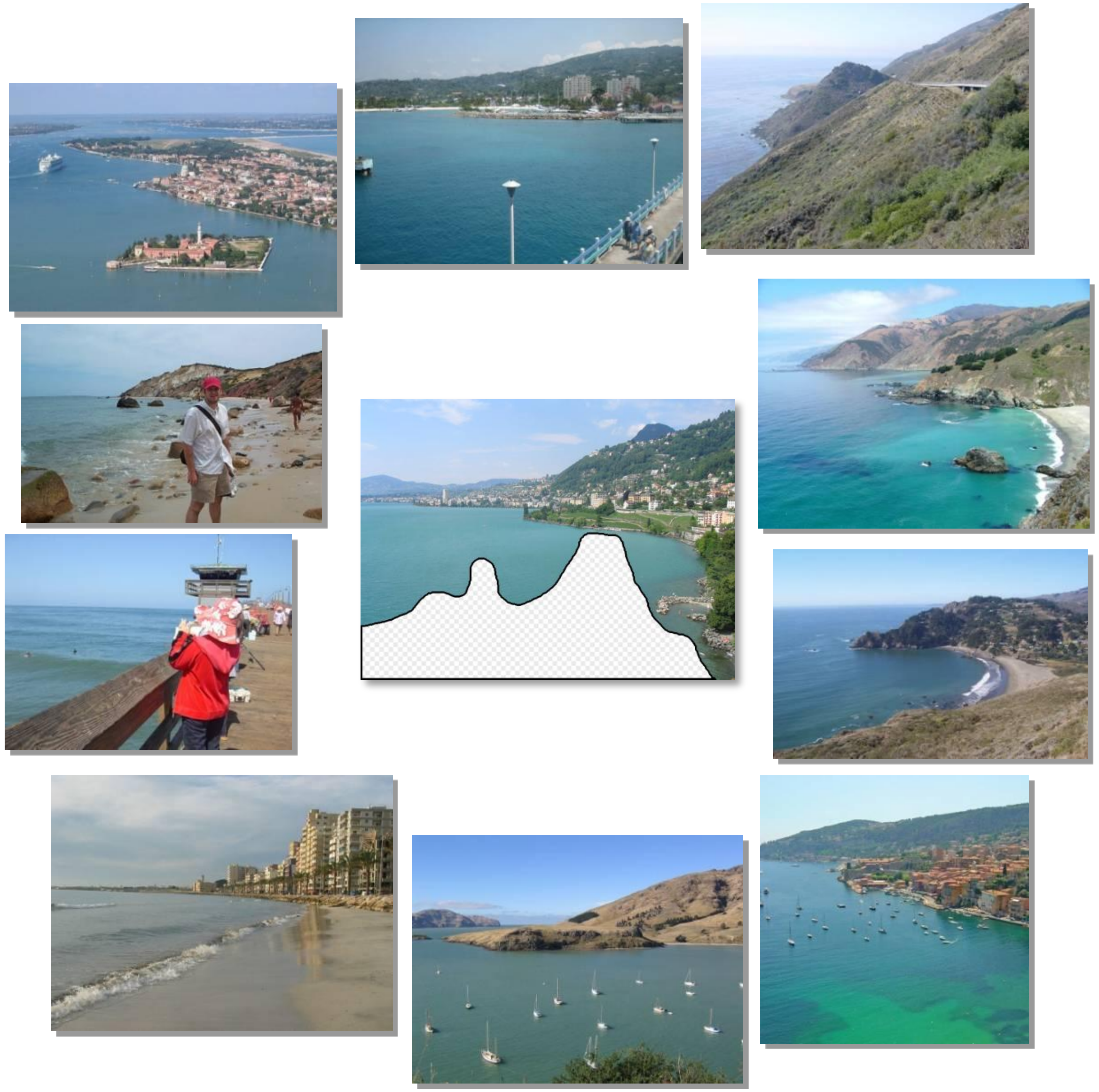
Trick: If you have enough images, the dataset will contain very similar images that you can find with simple maching methods.

How many images is enough?





Nearest neighbors from a collection of 20 thousand images



Nearest neighbors from a collection of 2 million images

Image Data on the Internet

- Flickr (as of Sept. 19th, 2010)
  - 5 billion photographs
  - 100+ million geotagged images
- Imageshack (as of 2009)
  - 20 billion
- Facebook (as of 2009)
  - 15 billion

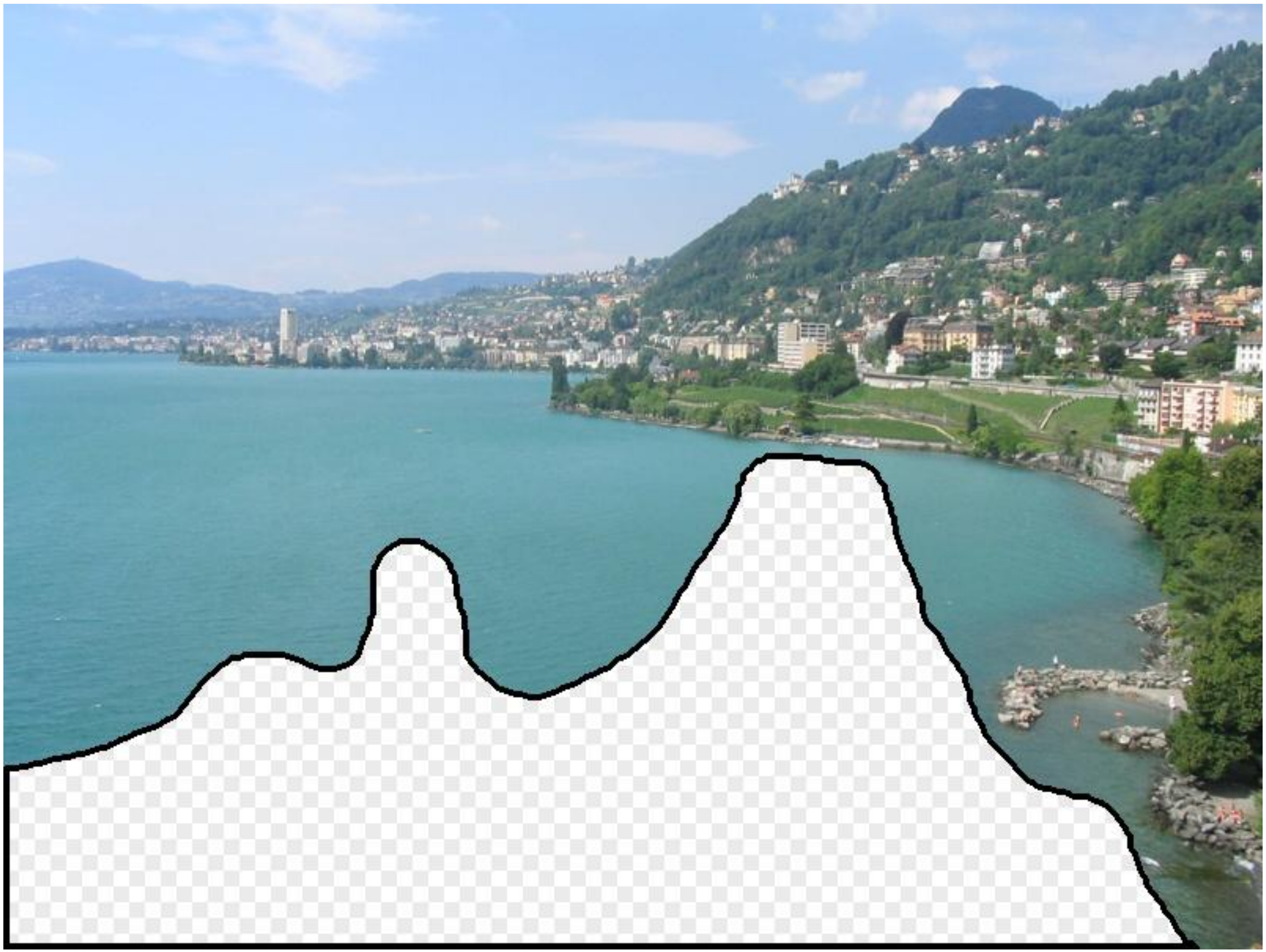
http://royal.pingdom.com/2010/01/22/internet-2009-in-numbers/

Image completion: how it works

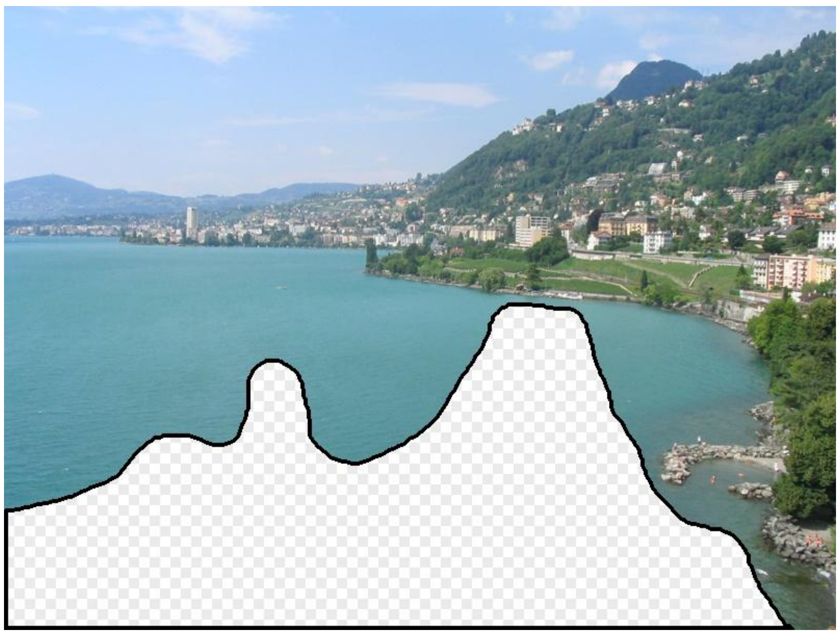
[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]

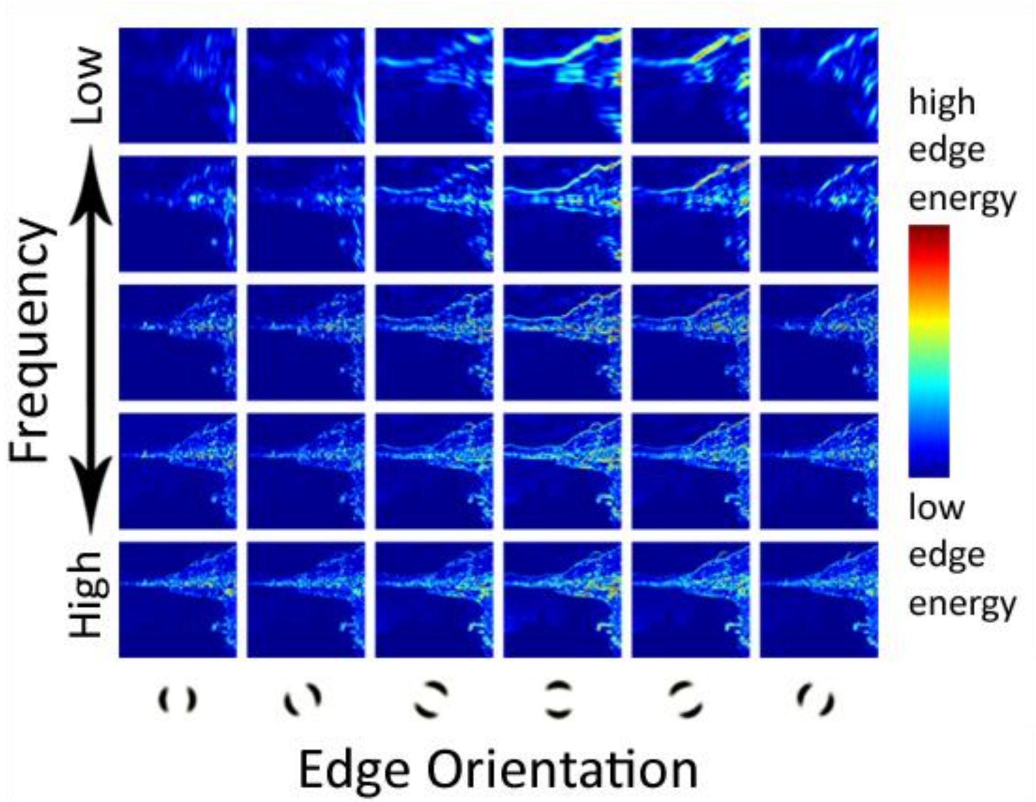
The Algorithm

Scene Matching

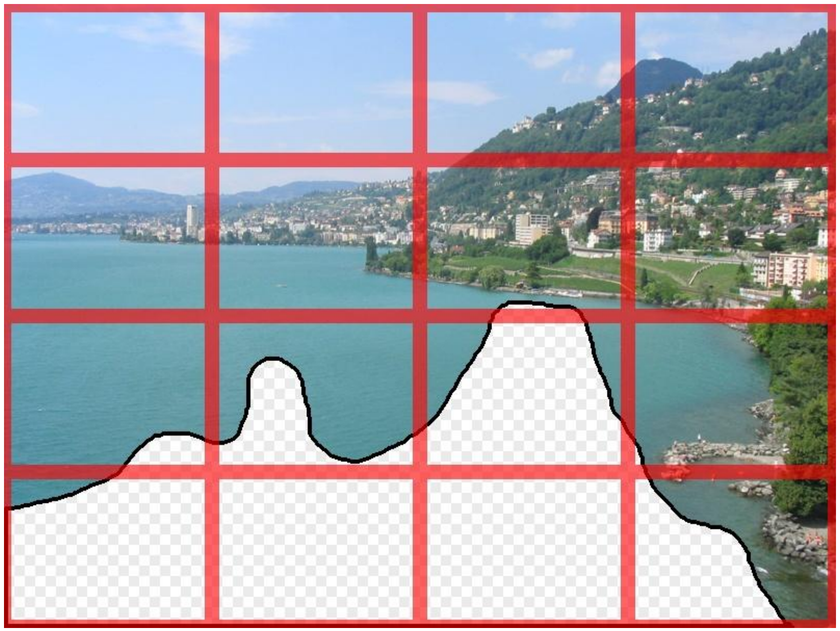


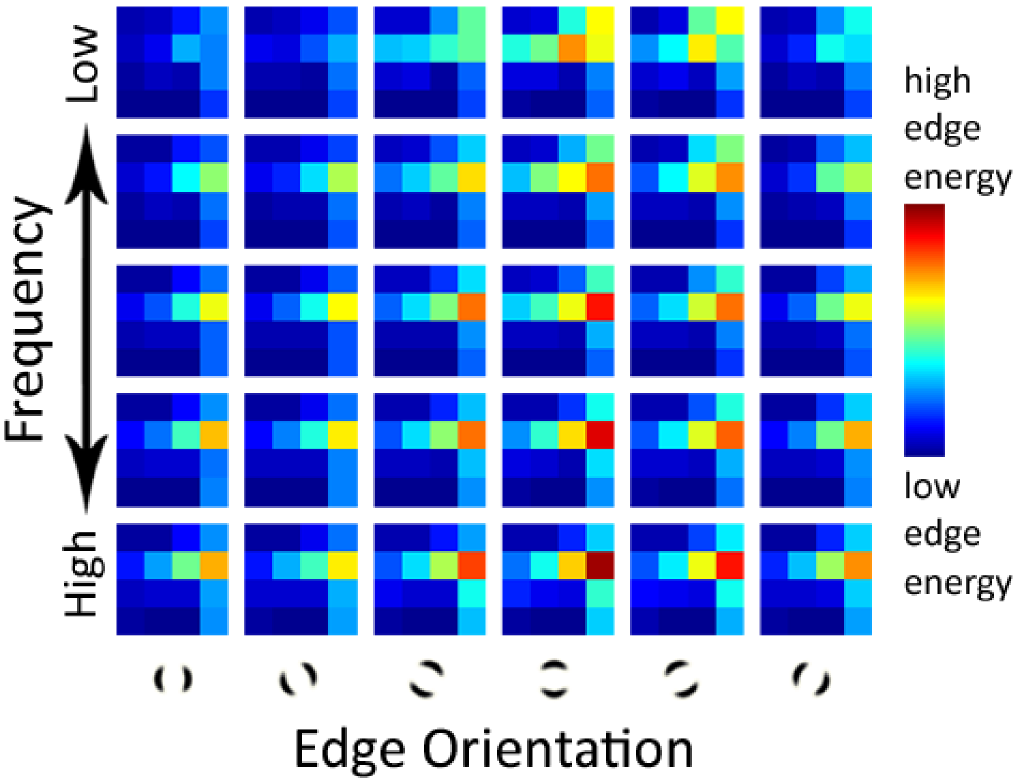
Scene Descriptor





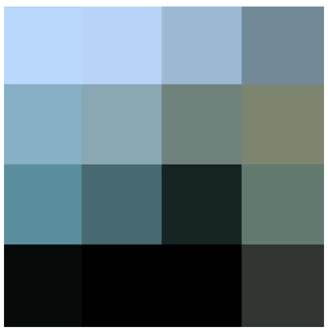
Scene Descriptor



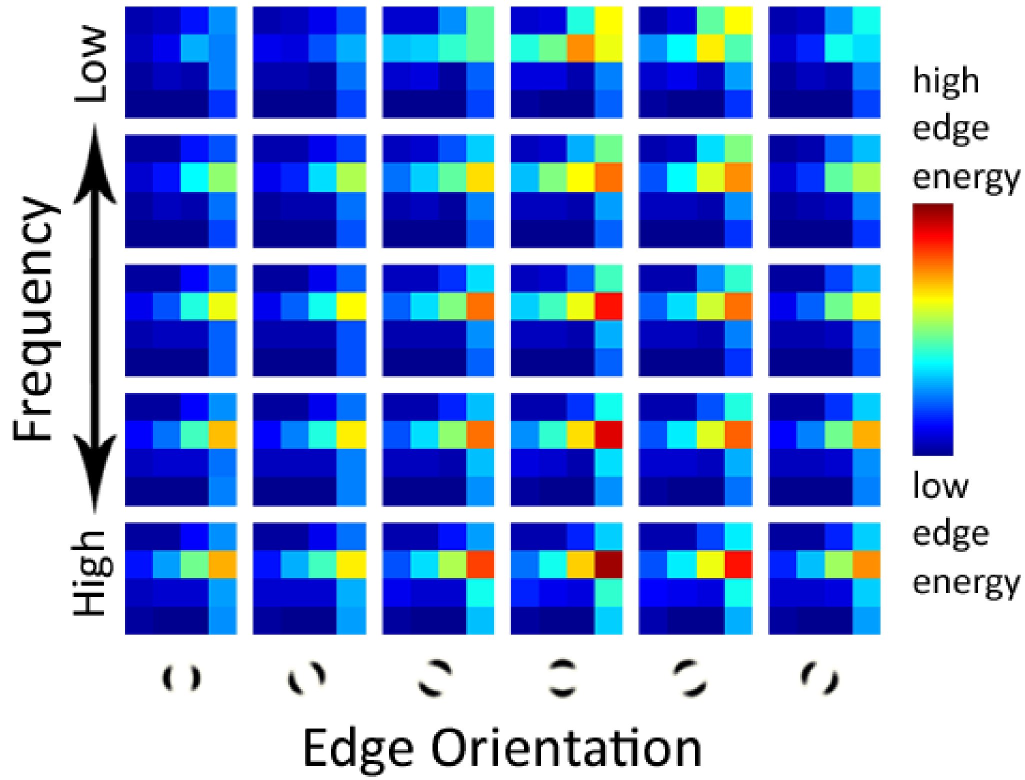


Scene Gist Descriptor
(Oliva and Torralba 2001)

Scene Descriptor

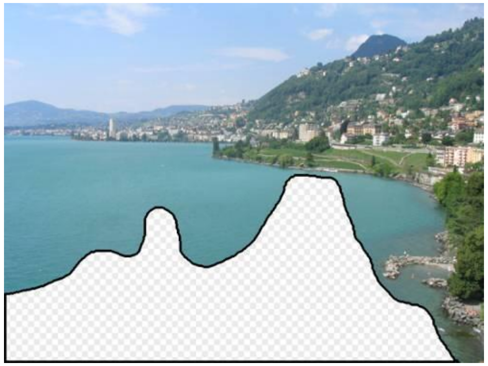


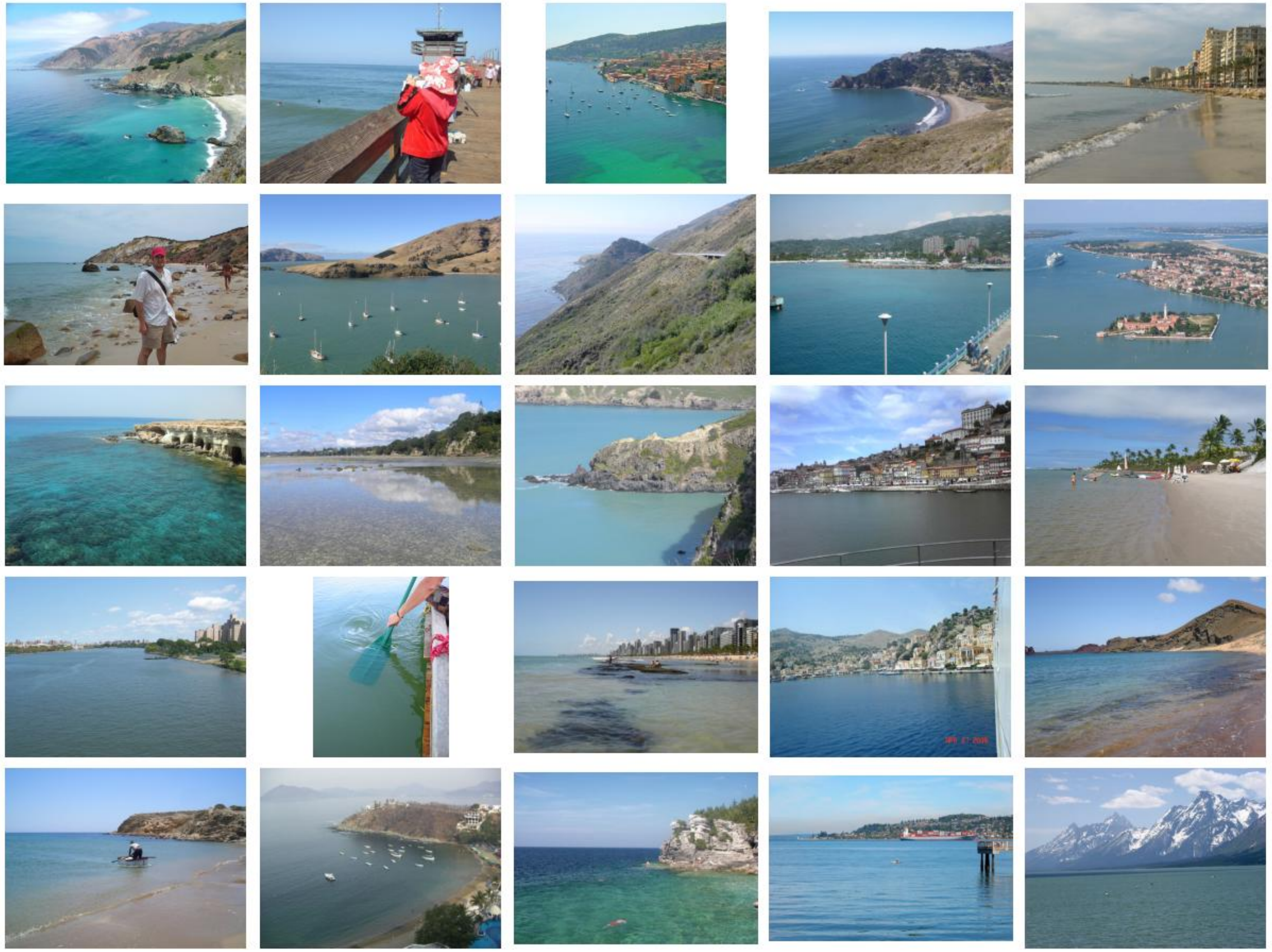
+



Scene Gist Descriptor
(Oliva and Torralba 2001)

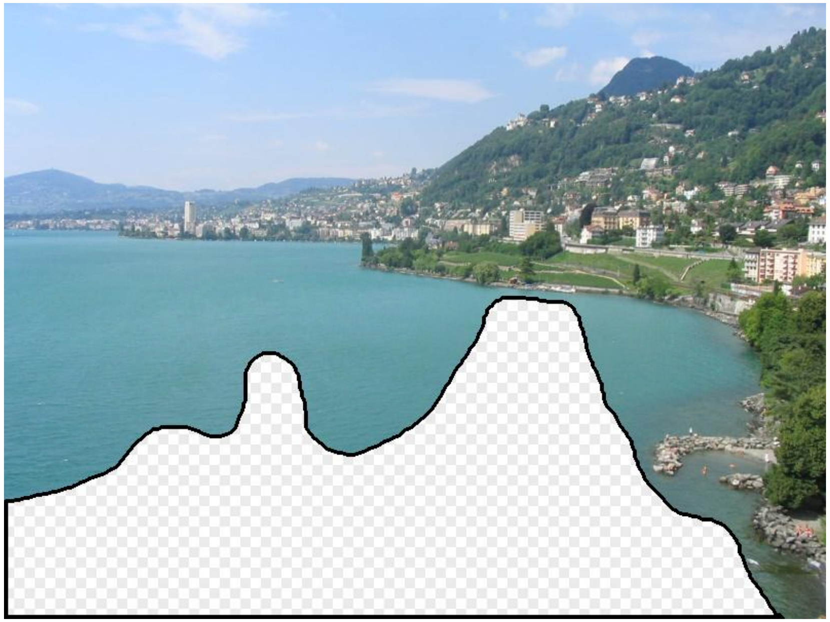
2 Million Flickr Images





... 200 total

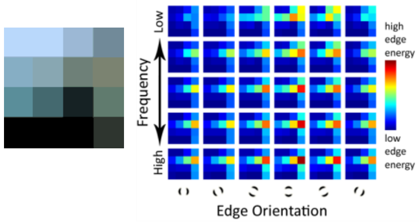
Context Matching



Graph cut + Poisson blending

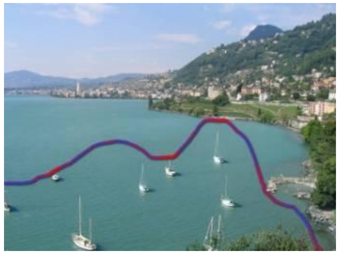
Result Ranking

We assign each of the 200 results a score which is the sum of:

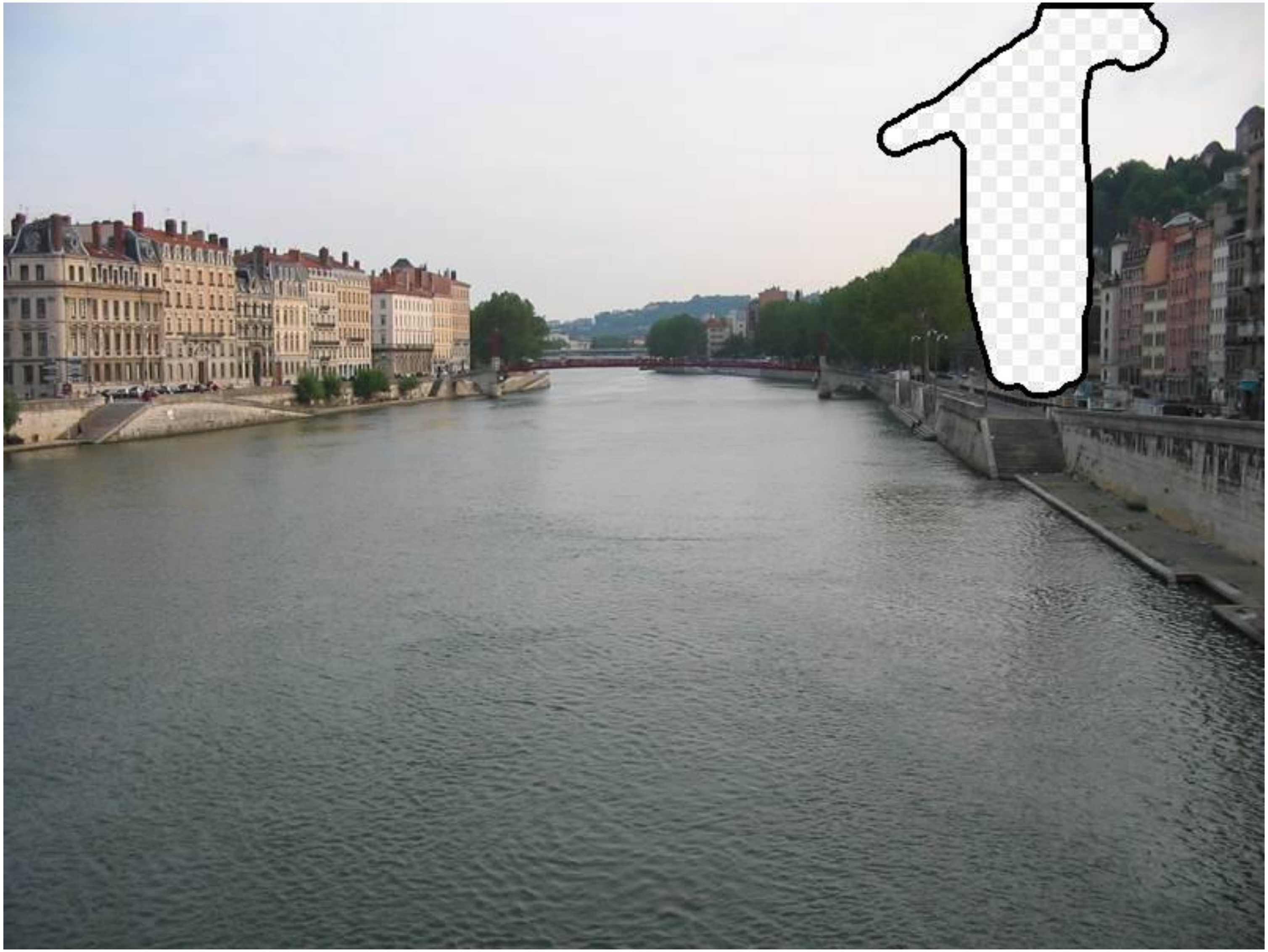


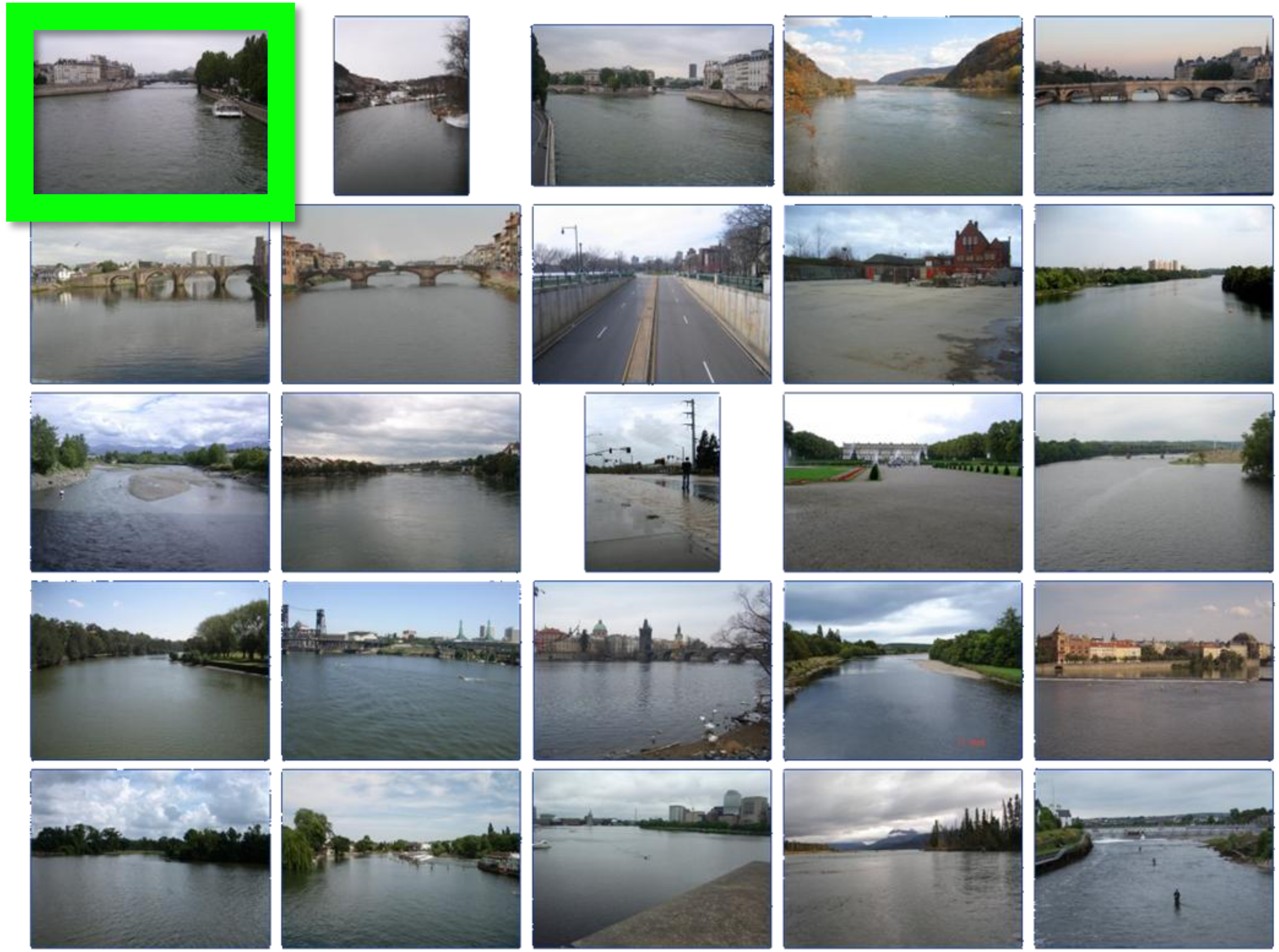
The scene matching distance

The context matching distance
(color + texture)

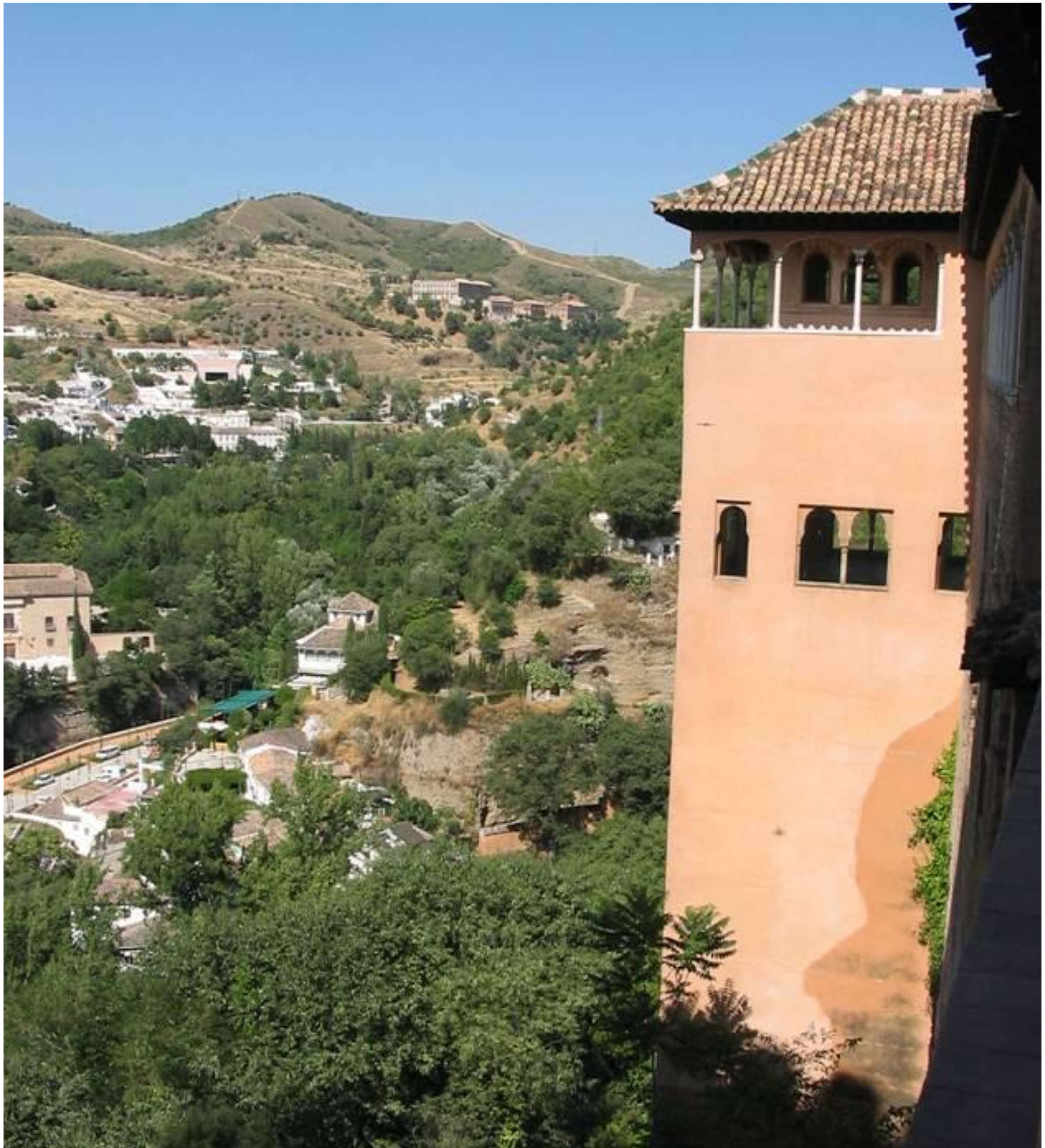


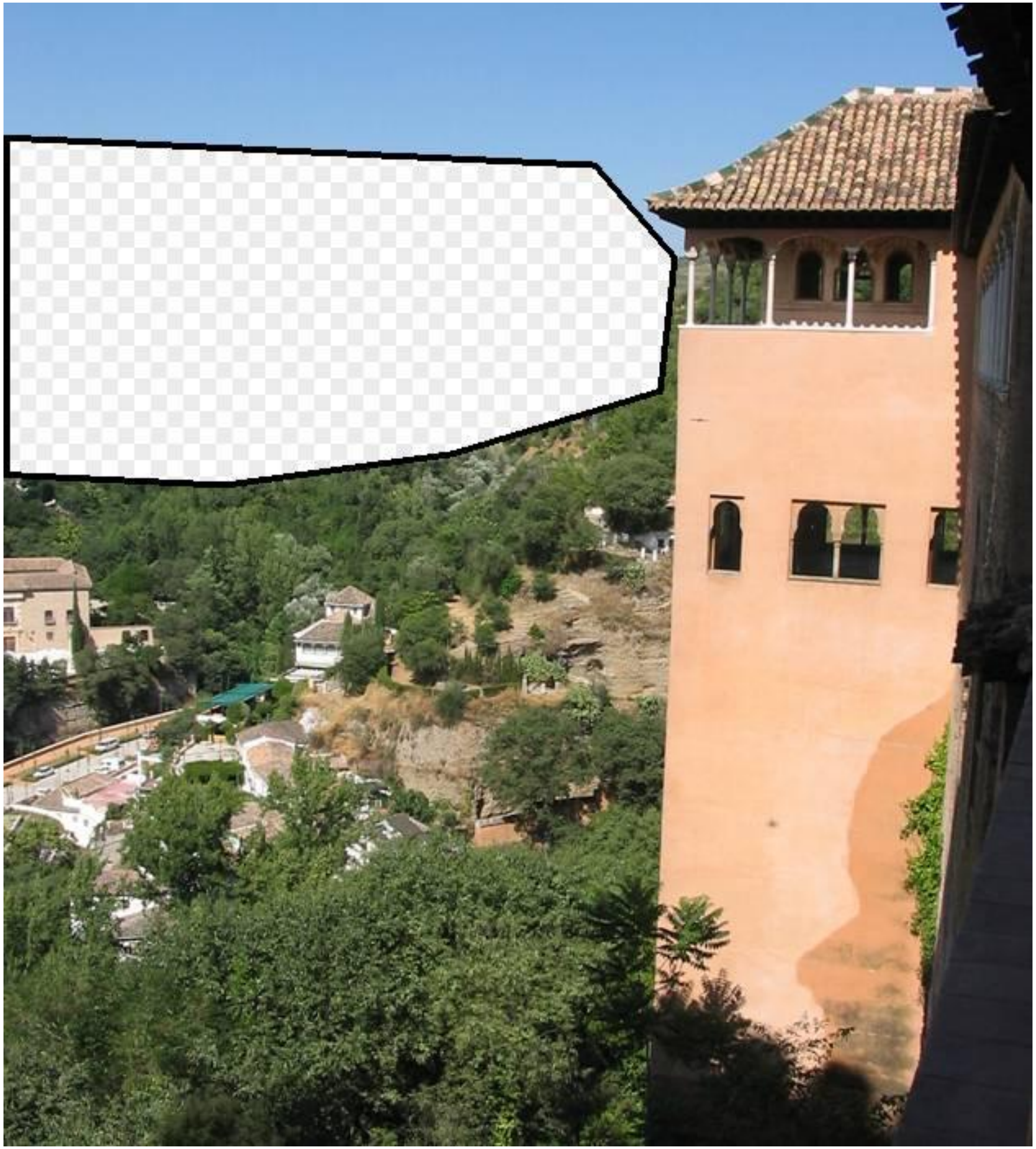
The graph cut cost





… 200 scene matches





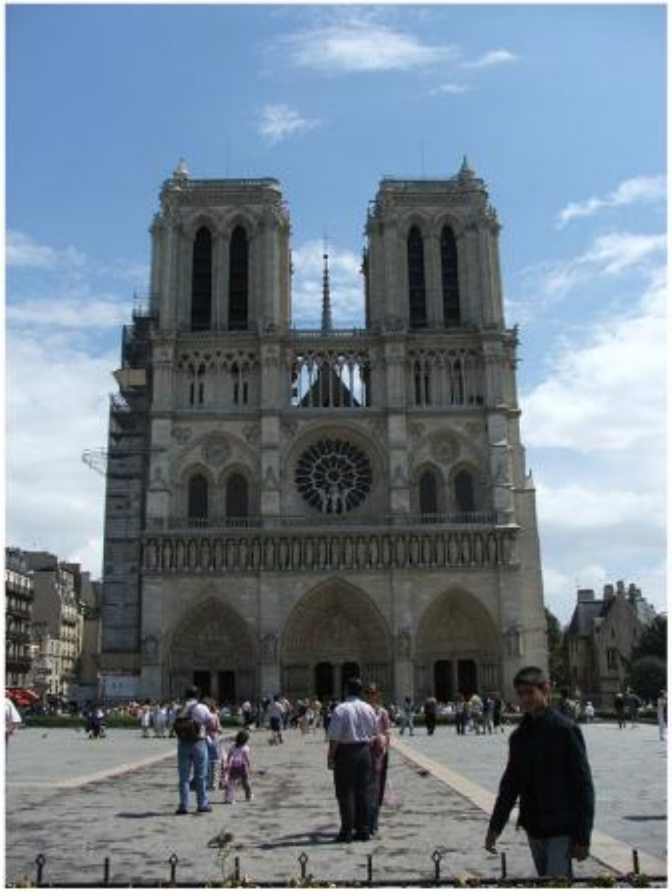
Which is the original?

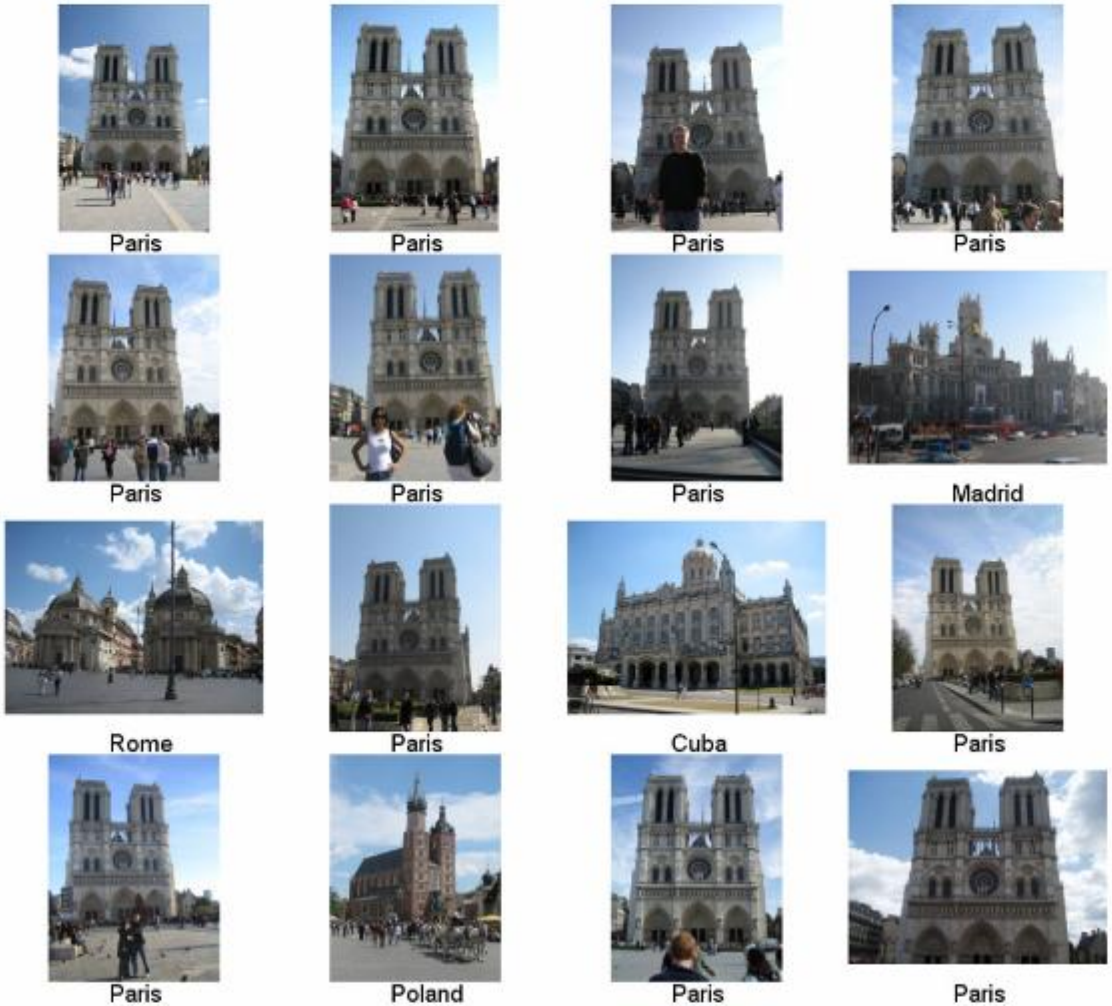
im2gps (Hays & Efros, CVPR 2008)

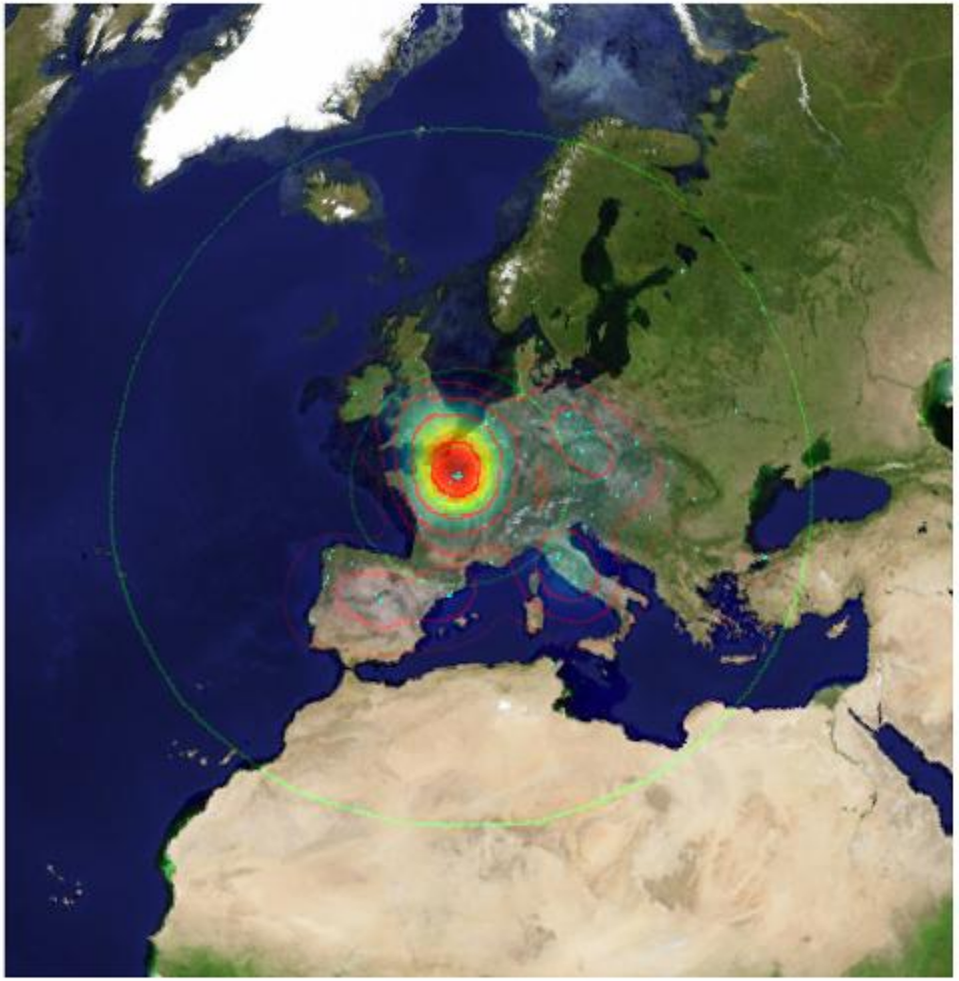
6 million geo-tagged Flickr images

http://graphics.cs.cmu.edu/projects/im2gps/

How much can an image tell about its geographic location?



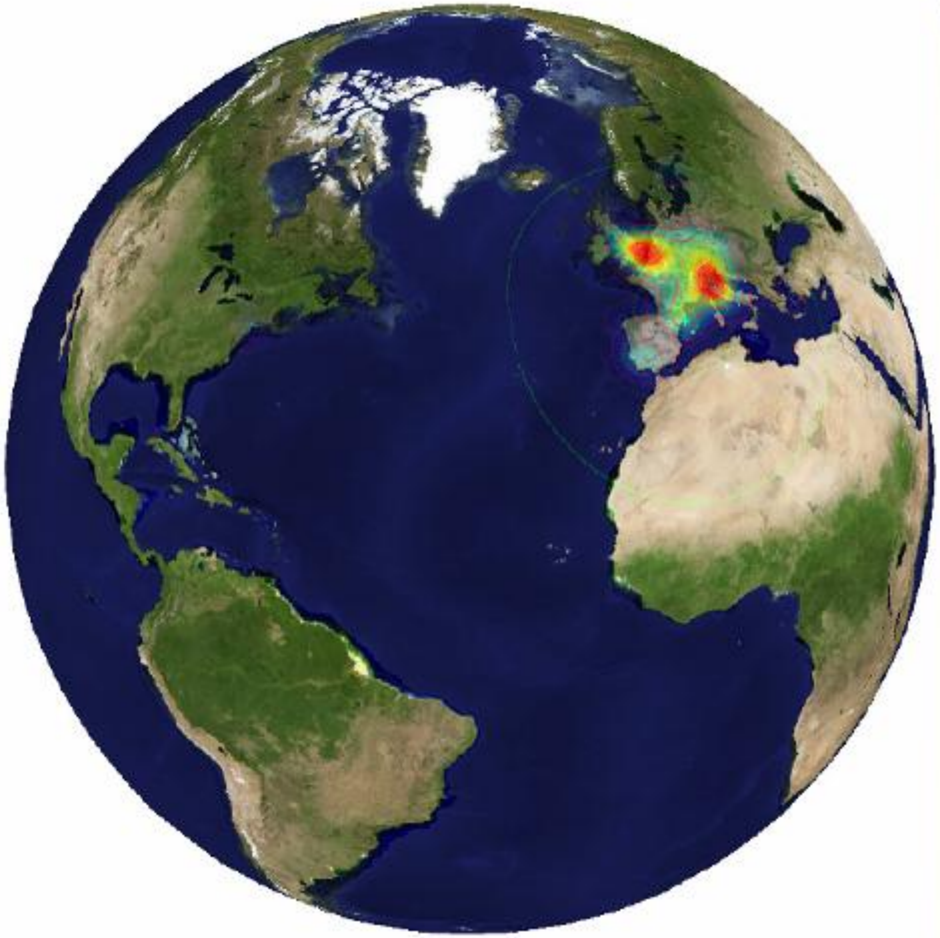

Paris
Paris
Paris
Paris
Paris
Paris
Paris
Madrid
Rome
Paris
Cuba
Paris
Paris
Poland
Paris
Paris

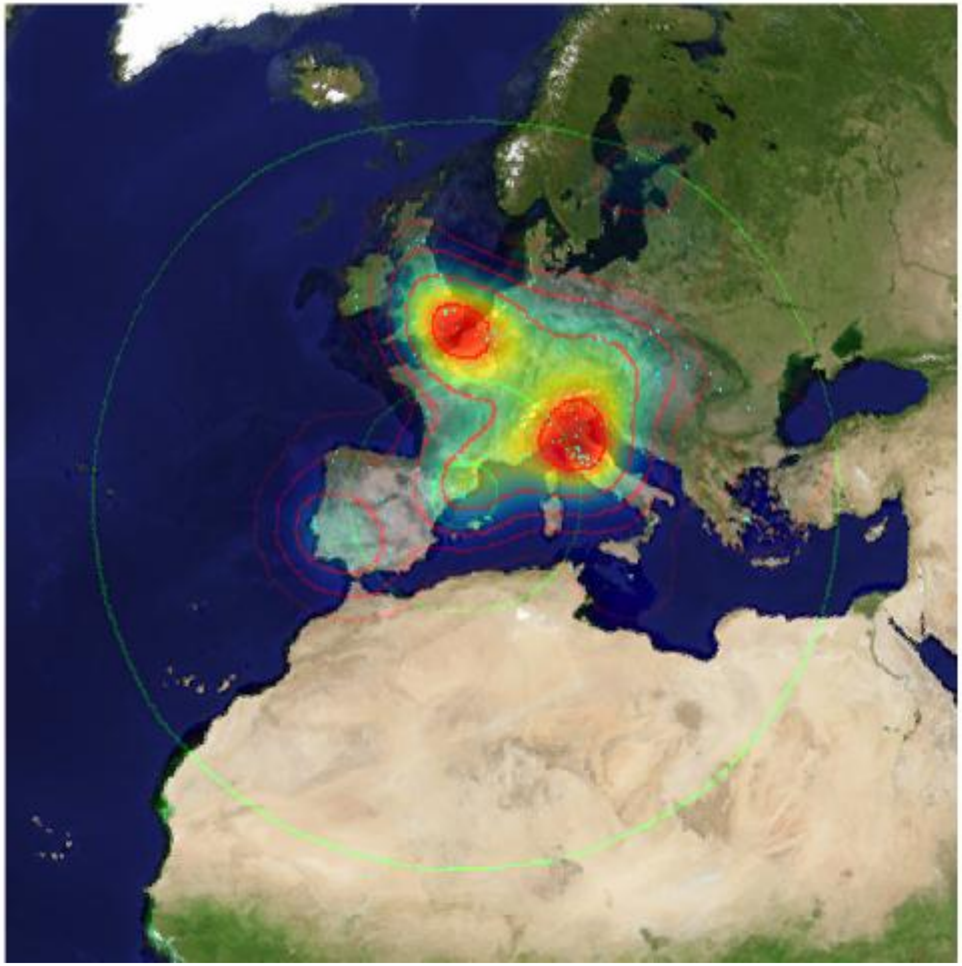


Im2gps

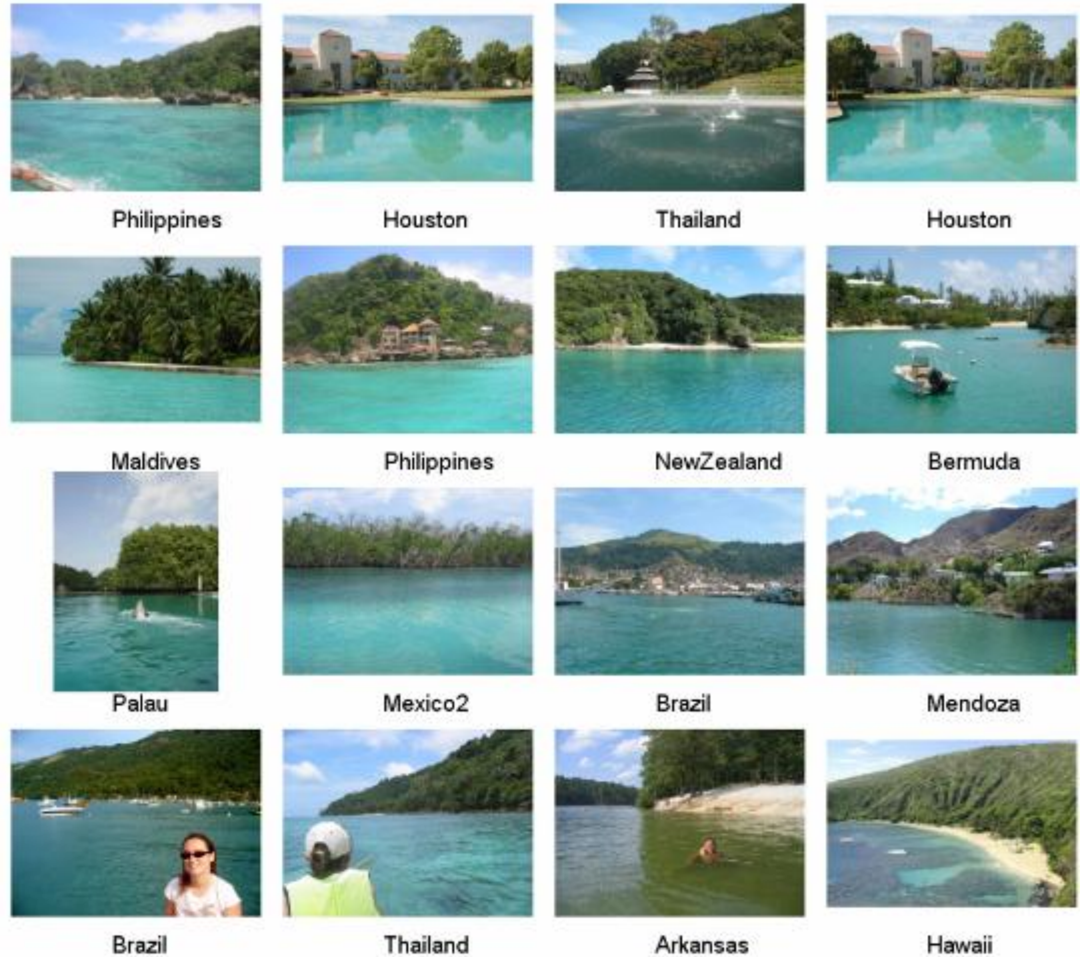
Example Scene Matches

Voting Scheme

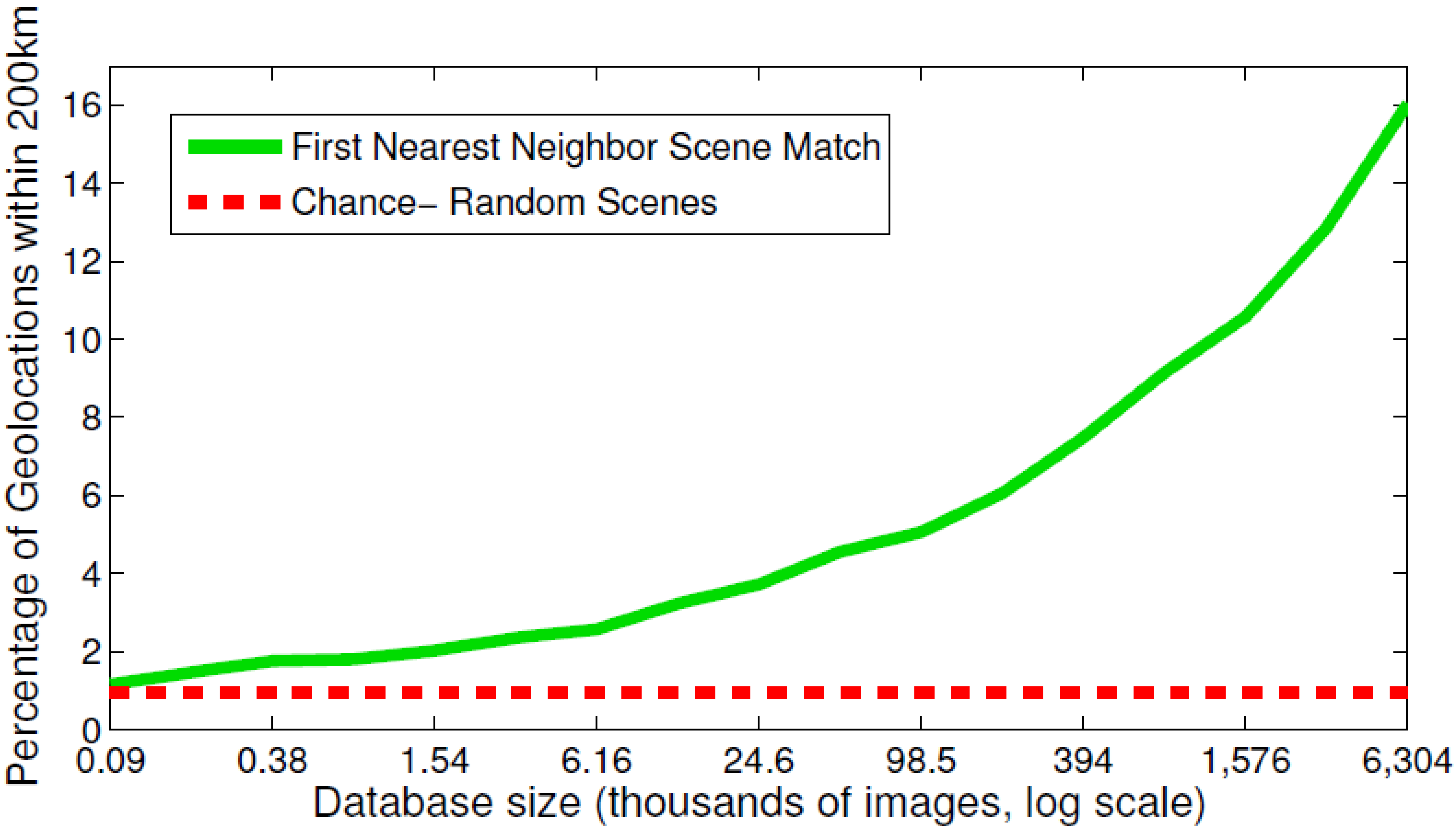




im2gps

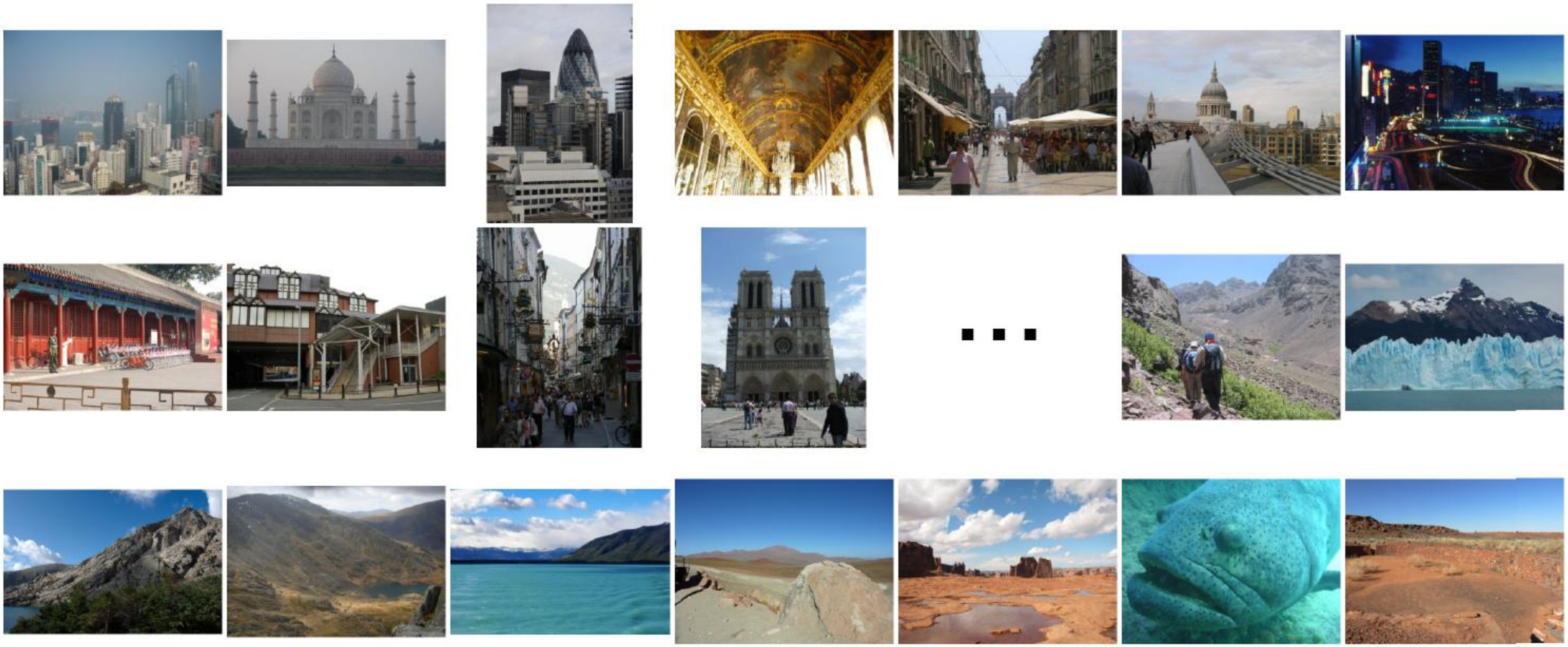

Philippines
Houston
Thailand
Houston
Maldives
Philippines
NewZealand
Bermuda
Palau
Mexico2
Brazil
Mendoza
Brazil
Thailand
Arkansas
Hawaii


Effect of Dataset Size



Population density ranking

High Predicted Density



Low Predicted Density

Where is This?

[Olga Vesselova, Vangelis Kalogerakis, Aaron Hertzmann, James Hays, Alexei A. Efros. Image Sequence Geolocation. ICCV'09]

Where is This?

Where are These?

15:14,
June 18th, 2006

16:31,
June 18th, 2006

Where are These?

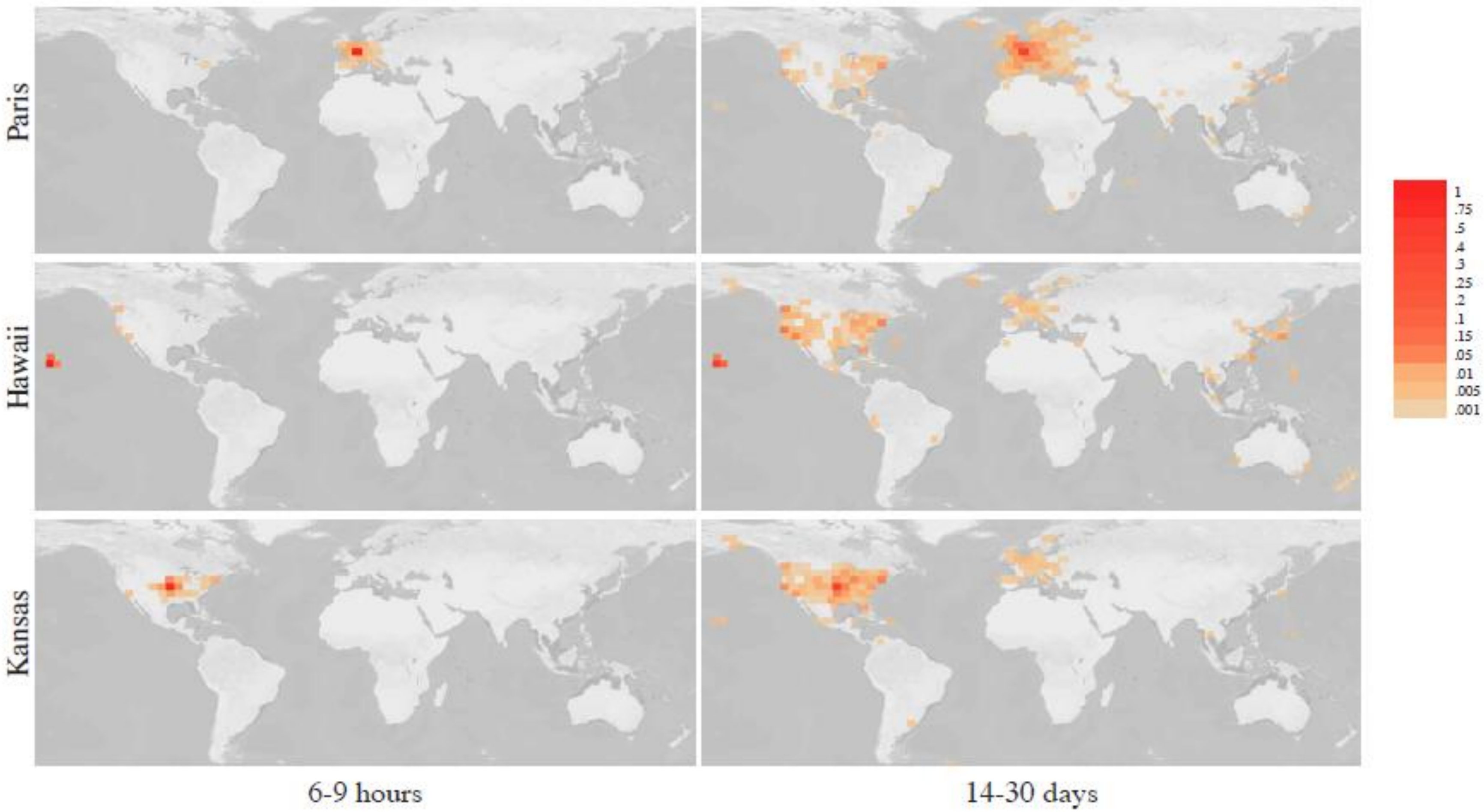
15:14,
June 18th, 2006

16:31,
June 18th, 2006

17:24,
June 19th, 2006

Results

- im2gps – 10% (geo-loc within 400 km)
- temporal im2gps – 56%



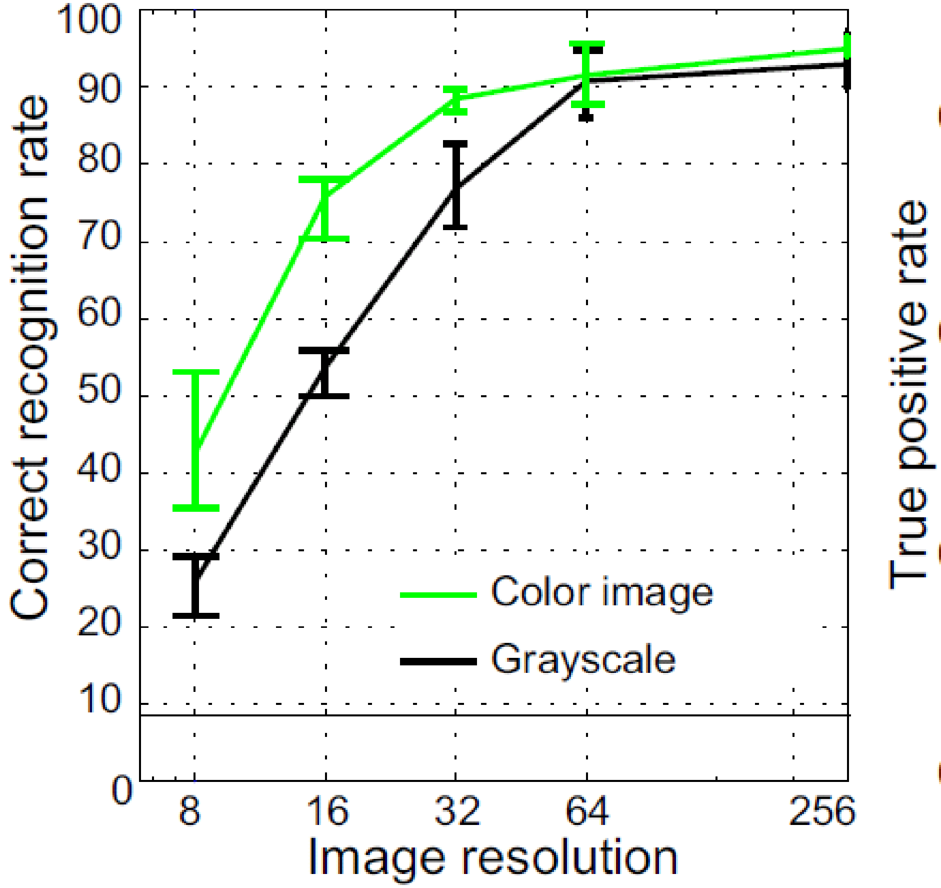
Tiny Images

80 million tiny images: a large dataset for non-parametric object and scene recognition
Antonio Torralba, Rob Fergus and William T. Freeman. PAMI 2008.

http://groups.csail.mit.edu/vision/TinyImages/

c) Segmentation of 32x32 images

Human Scene Recognition

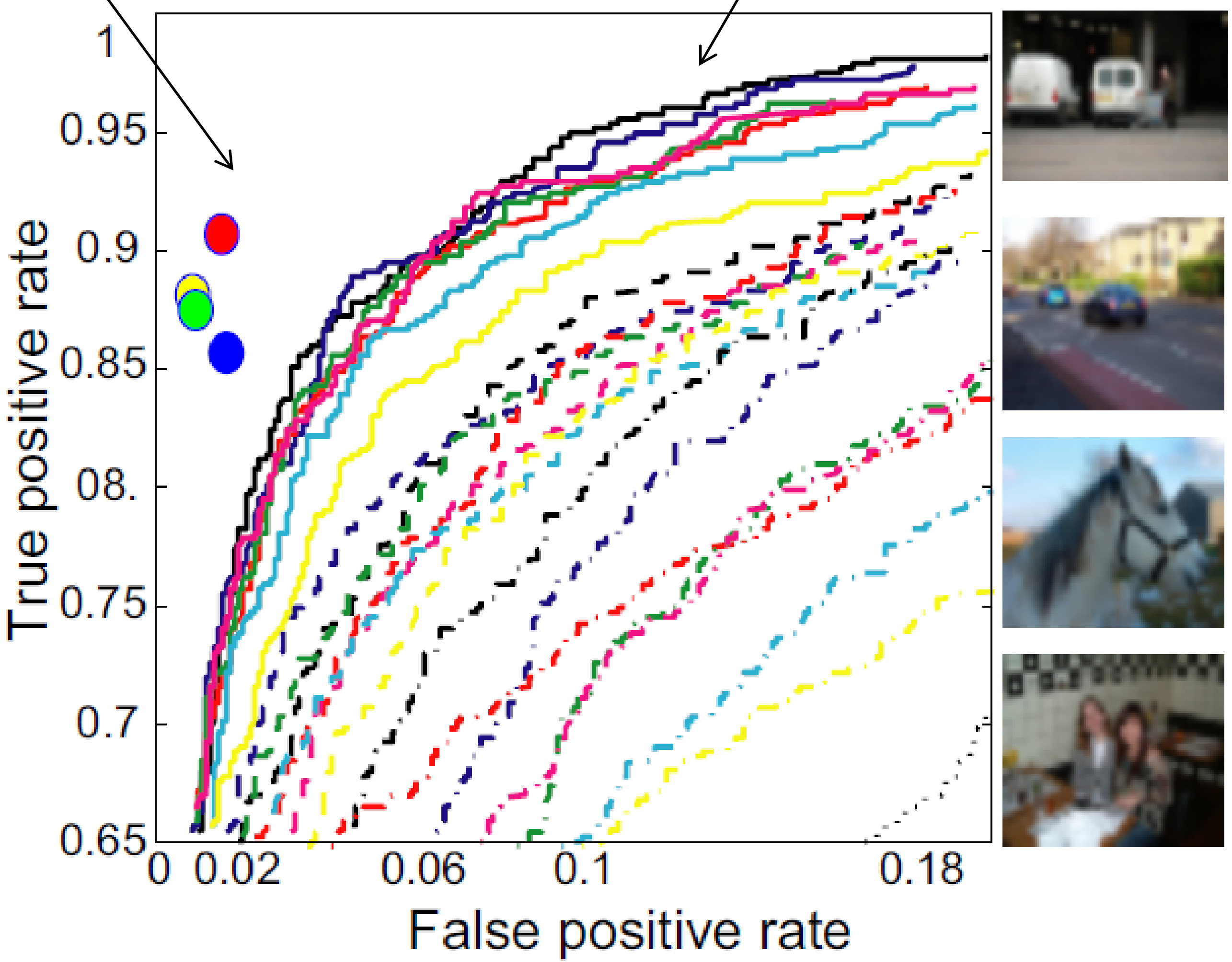


a) Scene recognition

Humans vs. Computers: Car-Image Classification

Humans for 32 pixel tall images

Various computer vision algorithms for full resolution images



Powers of 10

Number of images on my hard drive: 10^4

Number of images seen during my first 10 years: 10^8

(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: 10^{20}

106,456,367,669 humans[1] * 60 years * 3 images/second * 60 * 60 * 16 * 365 =

1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: 10^{88}

Number of all 32x32 images: 10^{7373}

$256^{32*32*3} \sim 10^{7373}$

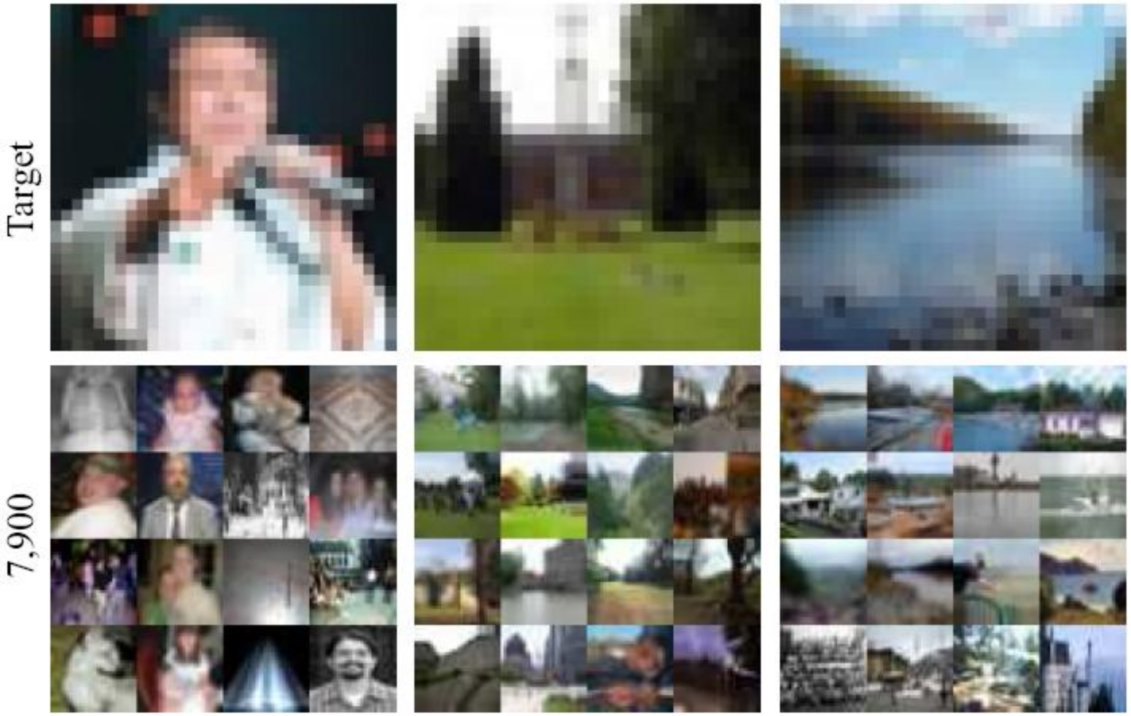
Scenes are unique

But not all scenes are so original

Lots

Of

Images

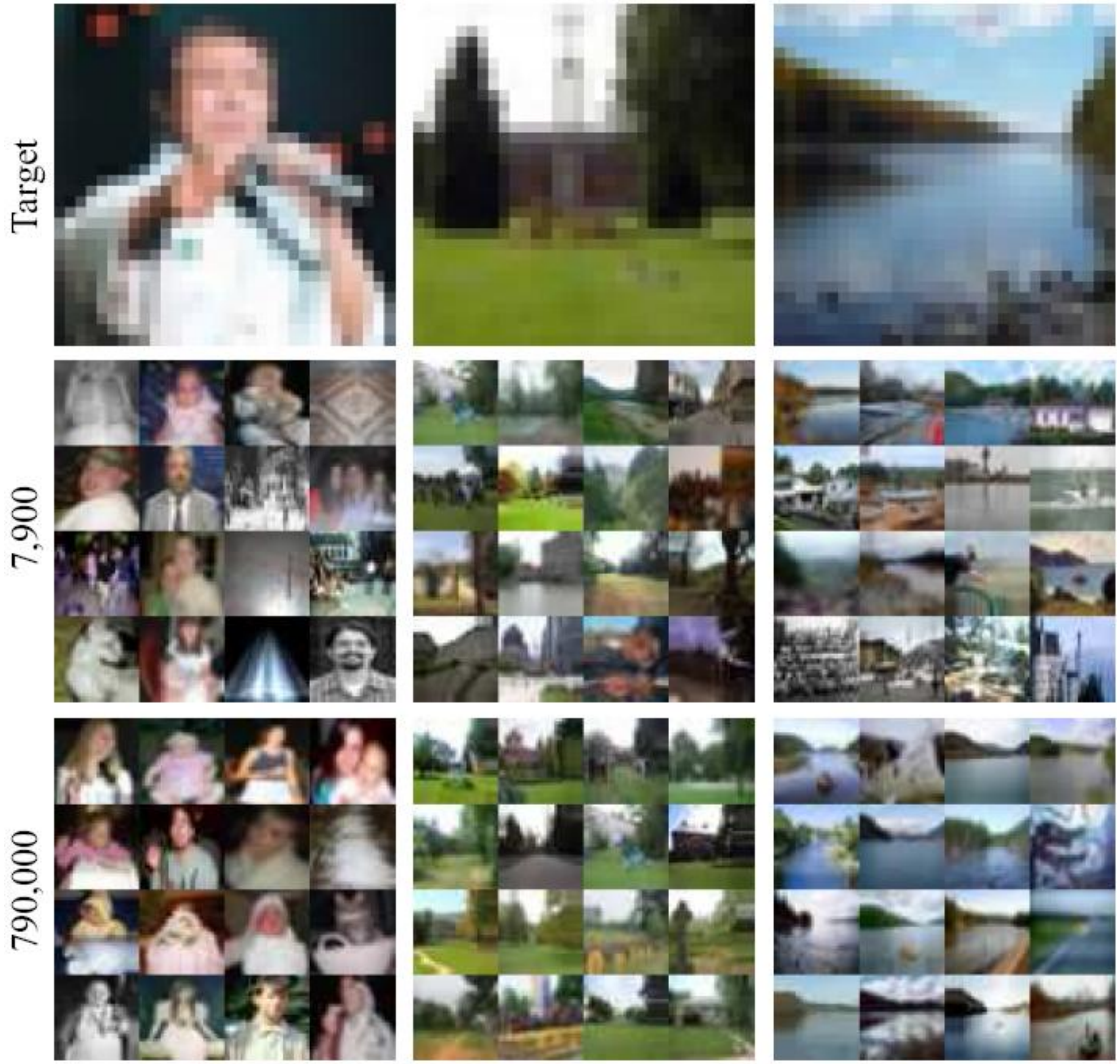


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

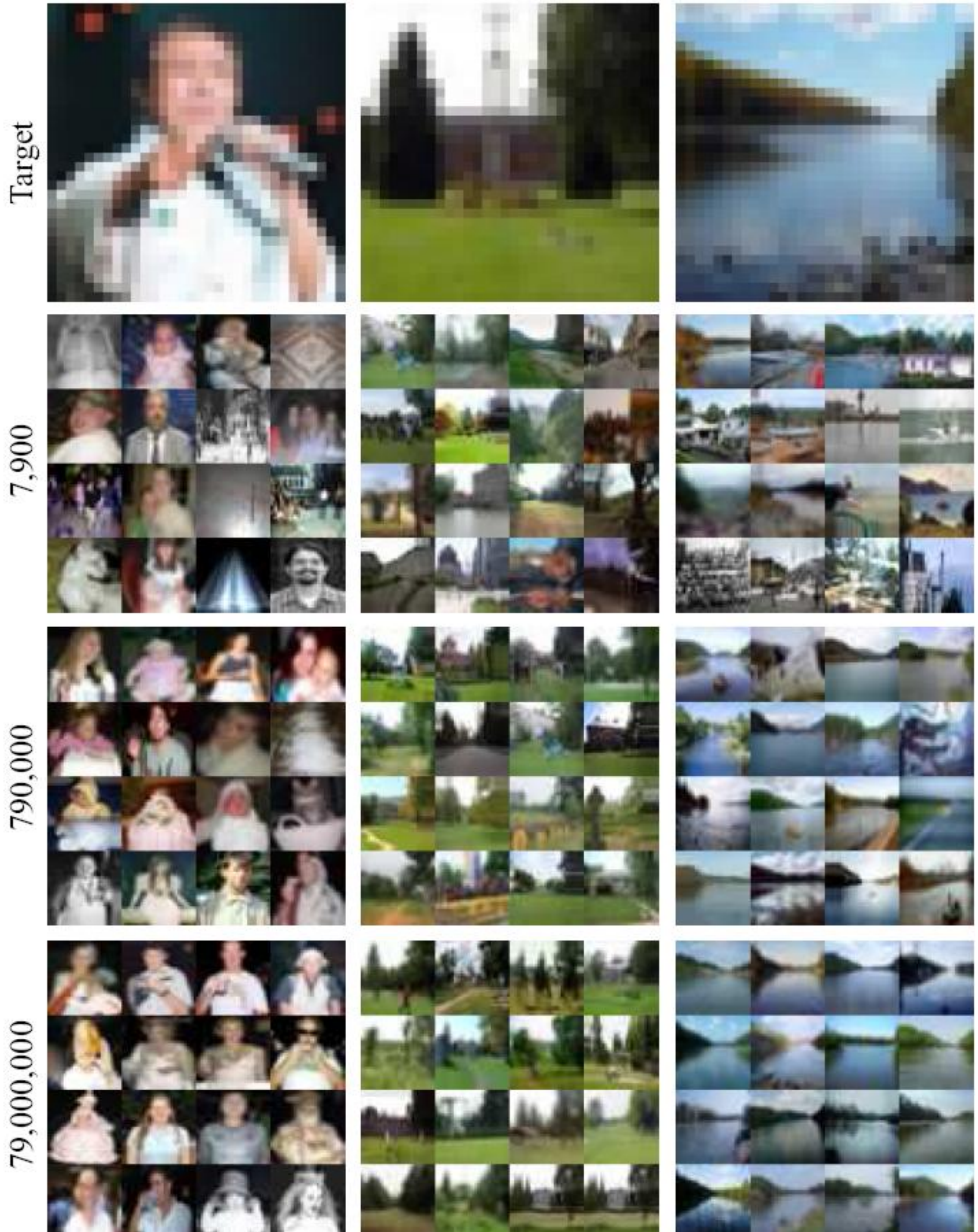


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images



Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match

Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

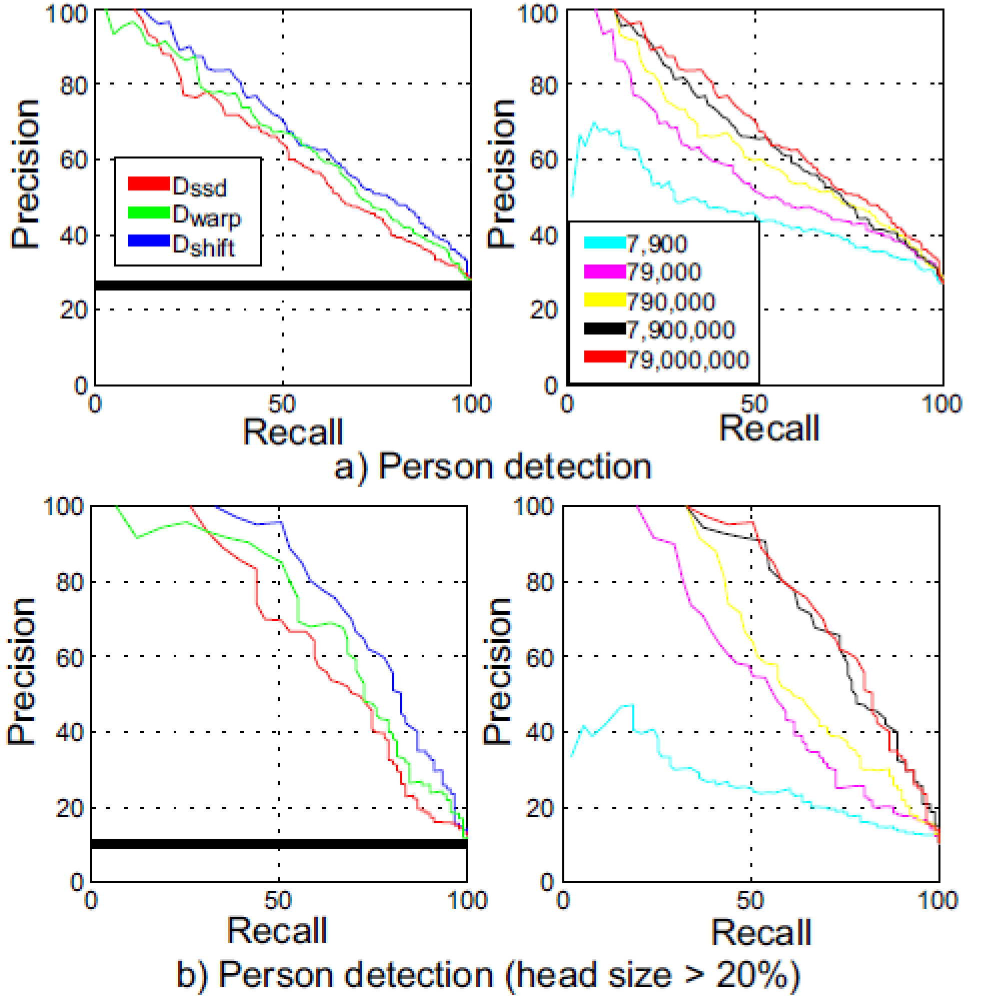
Matches (w/ color)

Avg Color of Match

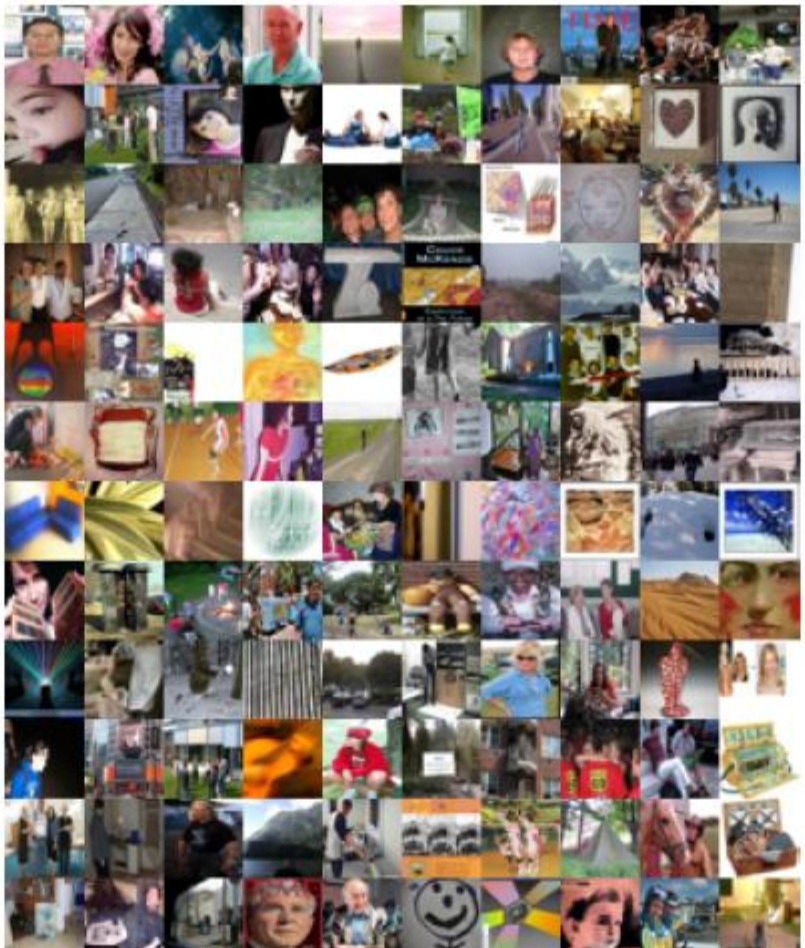
Application: Person Detection

80 million "tiny images" downloaded by keyword search.

80 nearest neighbors vote for image category.



Re-ranking Altavista search for “person”



a) Altavista ranking

b) Sorted by the tiny images

Recognition by Association

Rather than categorizing objects, associate them with stored examples of objects and transfer the associated labels.

Malisiewicz and Efros (CVPR 2008)

Training procedure

- Learn a region similarity measure from hand-segmented objects in LabelMe

- Similarity features
  - Shape: region mask, pixel area, bounding box size
  - Texture: normalized texton histogram
  - Color: mean RGB, std RGB, color histogram
  - Position: coarse 8x8 image mask, coords of top/bottom pixels

Training procedure

- Learn a distance/similarity measure *for each region*
  - Minimize distance to K most similar examples from same category
  - Maximize distance to examples from other categories

distance weights

distance measures

$$\{\mathbf{w}^*, \boldsymbol{\alpha}^*\} = \operatorname*{argmin}_{\mathbf{w},\boldsymbol{\alpha}} f(\mathbf{w}, \boldsymbol{\alpha})$$

$$f(\mathbf{w}, \boldsymbol{\alpha}) = \sum_{i \in C} \alpha_i L(-\mathbf{w} \cdot \mathbf{d}_i) + \sum_{i \notin C} L(\mathbf{w} \cdot \mathbf{d}_i)$$

$$\mathbf{w} \geq 0, \ \alpha_j \in \{0, 1\}$$

$$\textstyle\sum_j \alpha_j = K$$

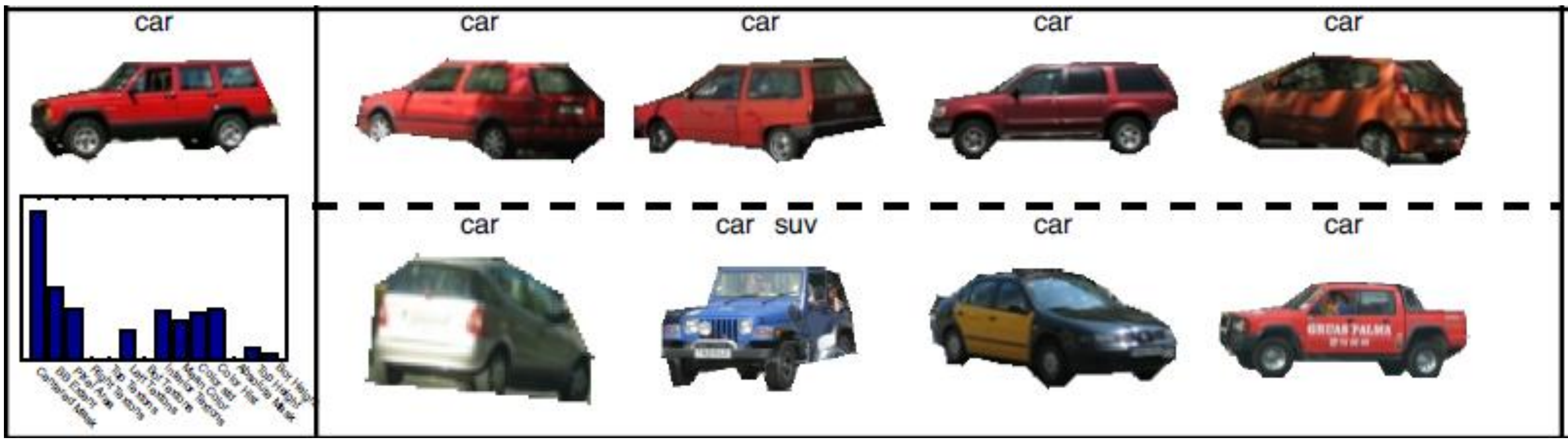
Set to 1 for K nearest examples

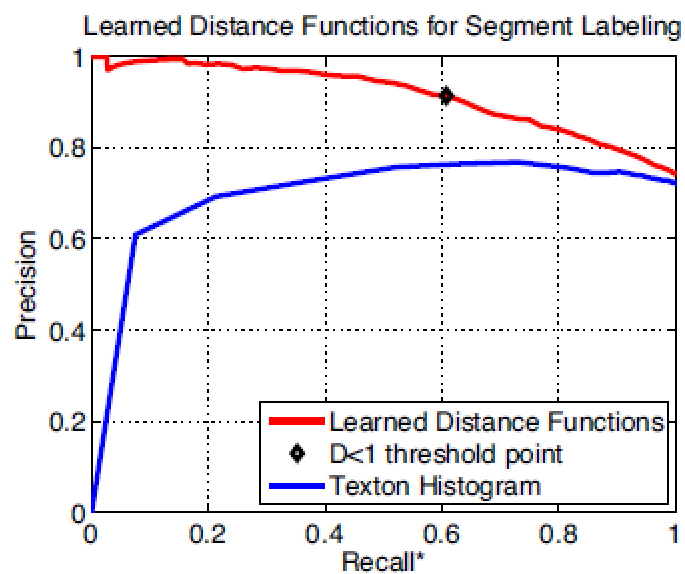
Hinge Loss

Learned Similarity Measure

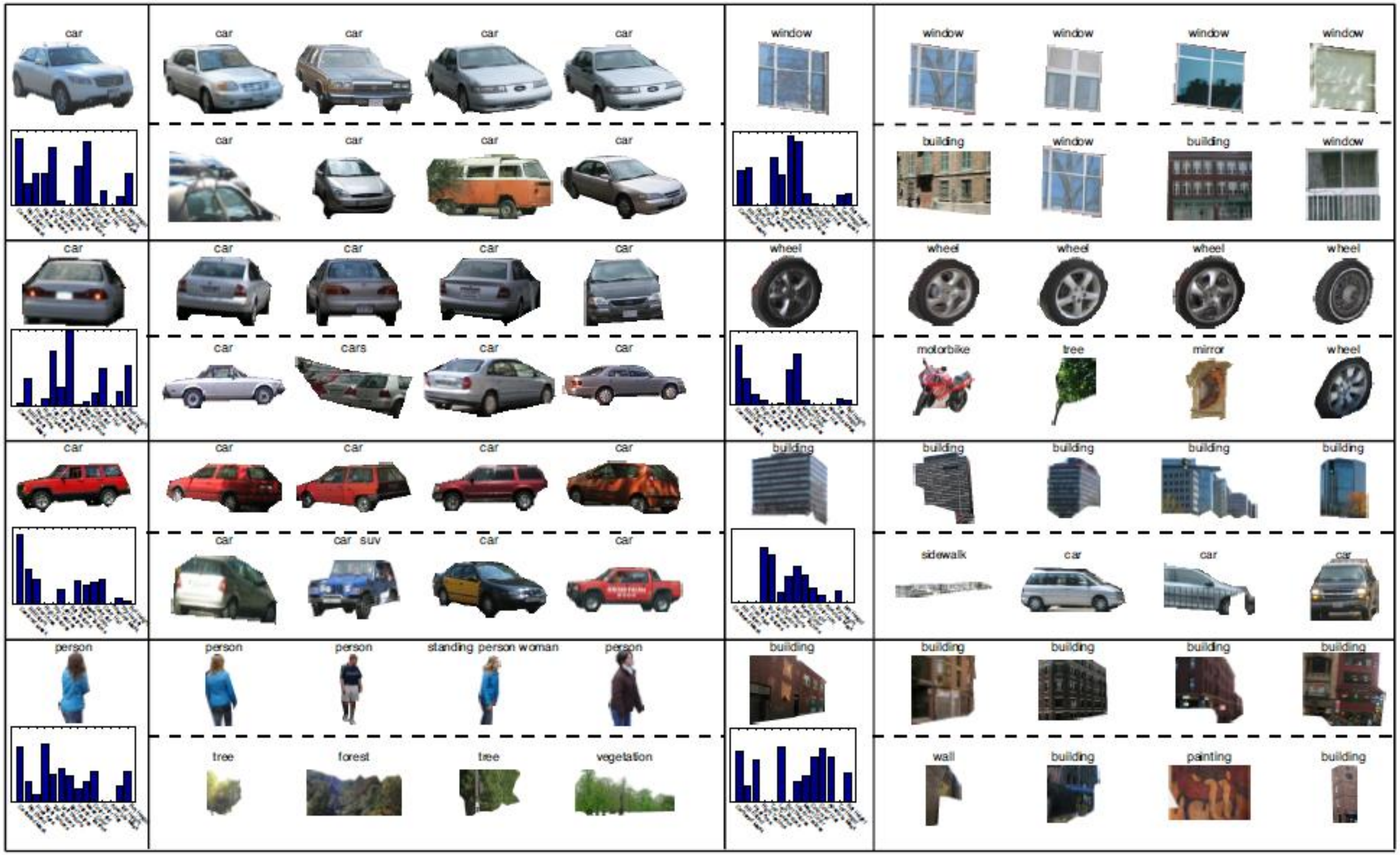
Learned Distance

Texton Distance



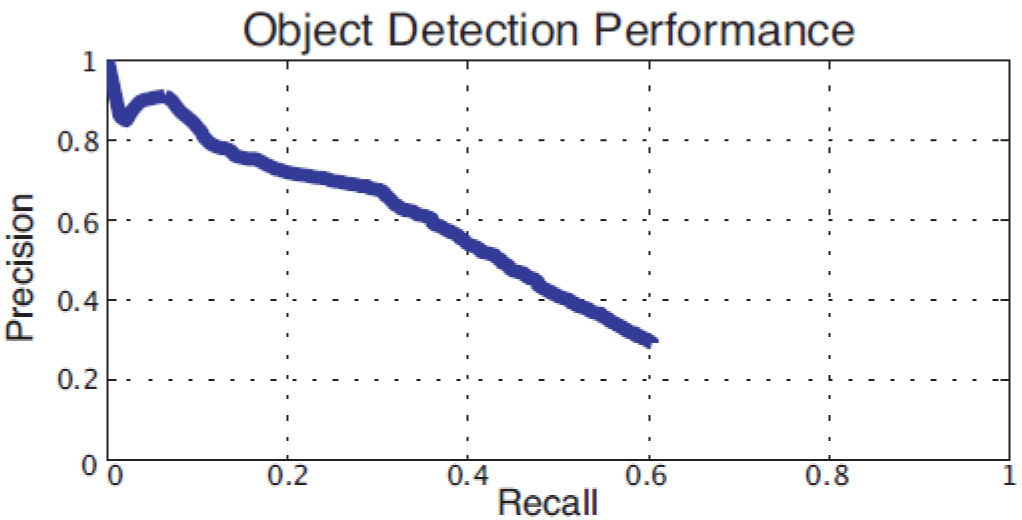


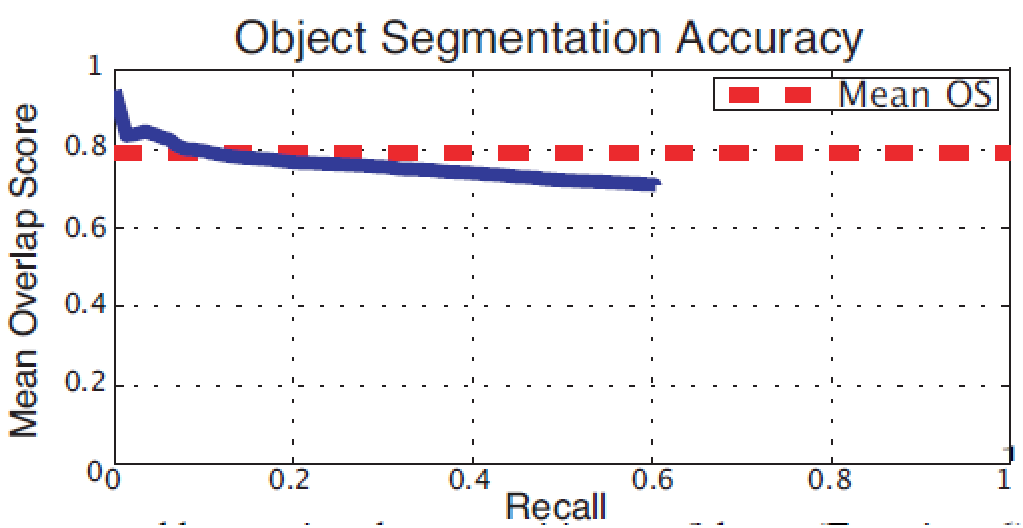
Learned Similarity Measure



Testing procedure

- Create multiple segmentations (MeanShift + Ncuts)
- Find similar object regions in training set; each votes for the object label
- What about bad segments?
  - Most of the time, they don't match any objects in the training set
  - Consider only associations with distance < 1


Object Detection Performance
Precision
Recall
0
0.2
0.4
0.6
0.8
1


Object Segmentation Accuracy
Mean OS
Mean Overlap Score
Recall
0
0.2
0.4
0.6
0.8
1

Automatic Parses

Summary

- With billions of images on the web, it’s often possible to find a close nearest neighbor
- In such cases, we can shortcut hard problems by “looking up” the answer, stealing the labels from our nearest neighbor
- For example, simple (or learned) associations can be used to synthesize background regions, colorize, or recognize objects