Visual Data on the Internet

With slides from James Hays, Antonio Torralba, and Frederic Heger

Cassandra Jones:
https://youtu.be/5H7WrIBrDRg?t=161

CS194: Image Manipulation & Computational Photography
Alexei Efros, UC Berkeley, Fall 2018

facebook

140 billion images
6 billion added monthly

flickr

6 billion images

the simple image sharer
imgur

1 billion images
served daily

YouTube

72 hours uploaded
every minute

3.5 trillion photographs

90% of net traffic will be visual!

Too Big for Humans

[Perona 2010]

Big issues

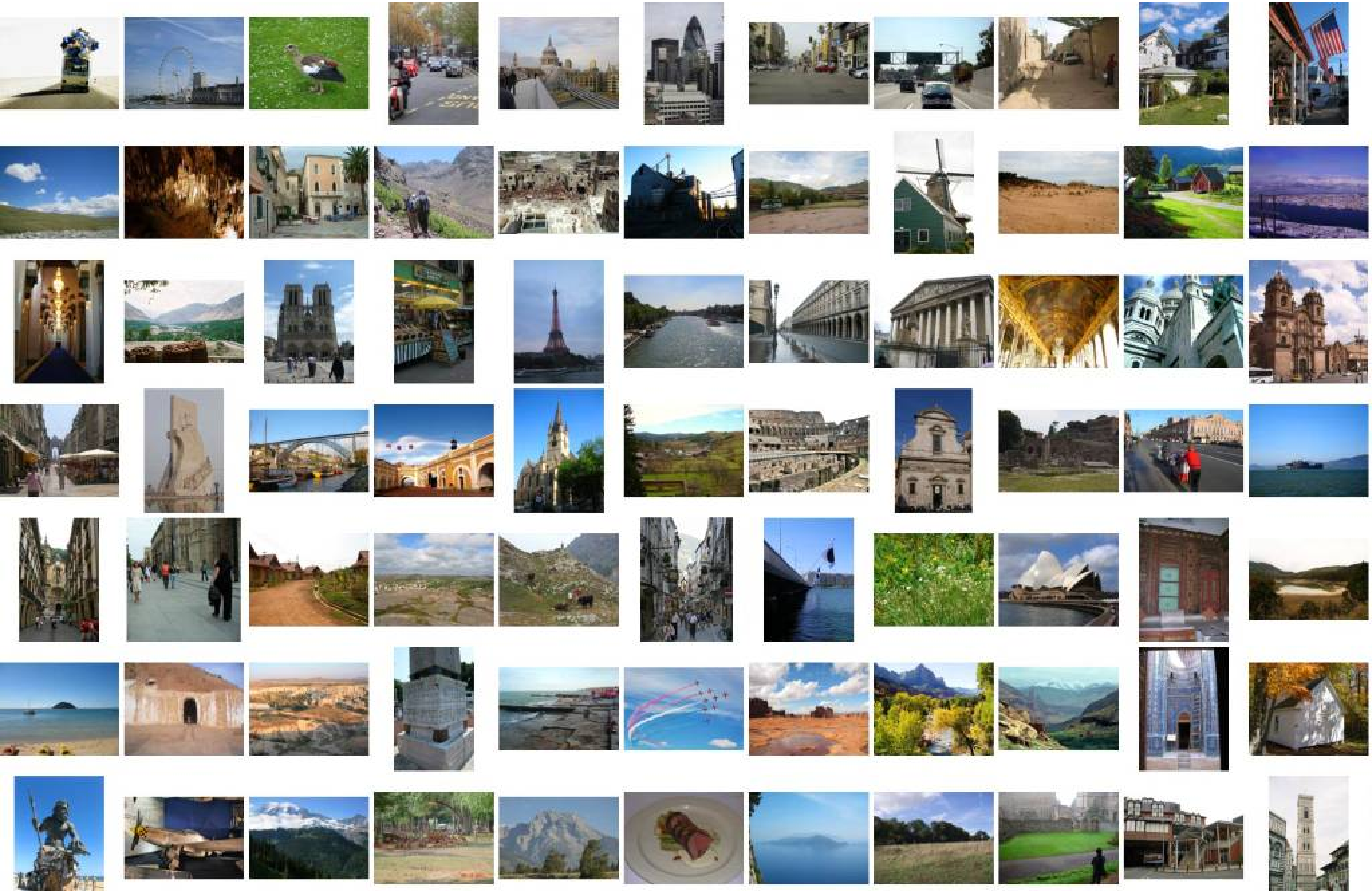
- What is out there on the Internet? How do we get it? What can we do with it?
- How do we compute distances between images?

Subject-specific Data

Photos of Coliseum

Portraits of Bill Clinton

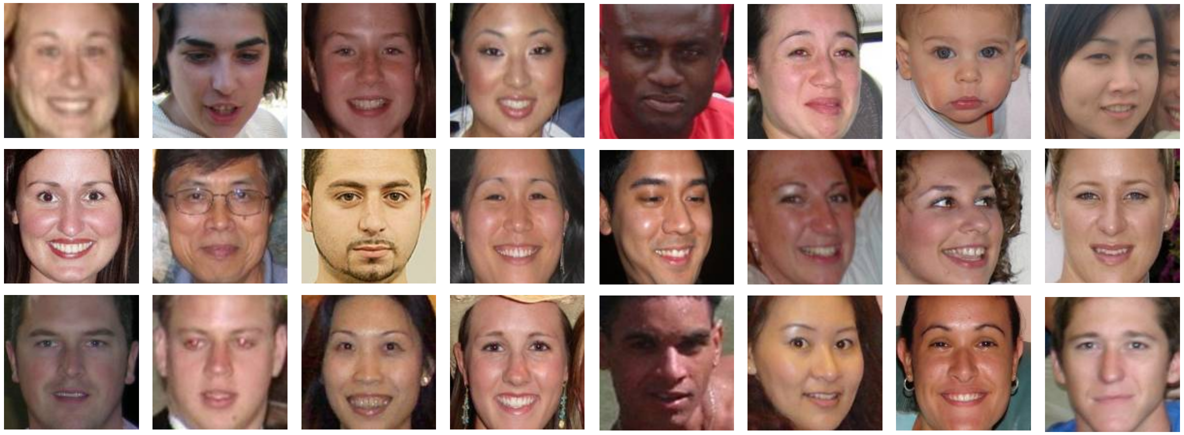
Much of Captured World is "generic"



Generic Data

street scenes

Food plates



faces

pedestrians

https://www.youtube.com/watch?v=ZTpqd3Fvaq8

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)

Is Generic Data useful?

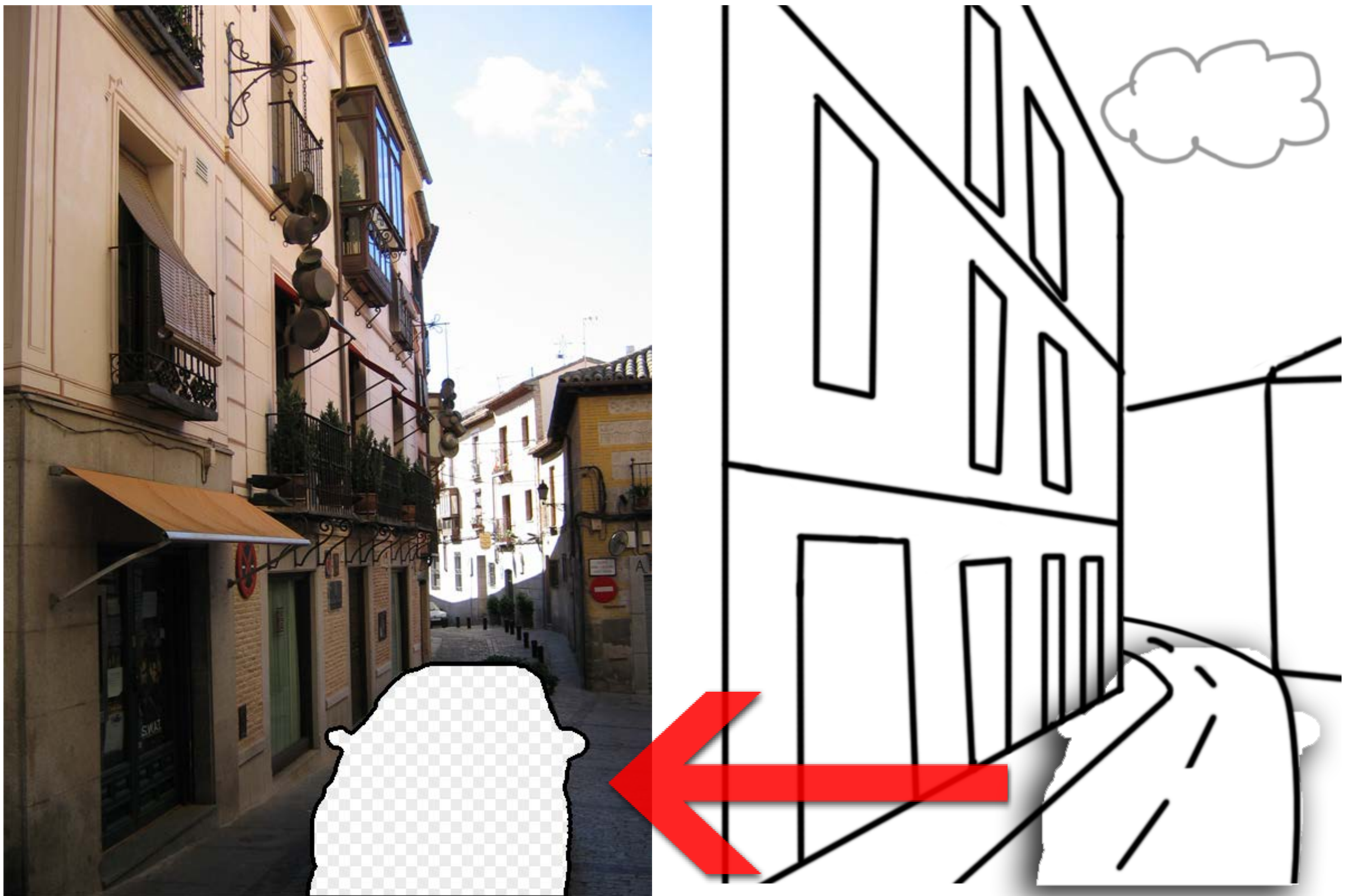
A motivating example…

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

Diffusion Result

Efros and Leung result

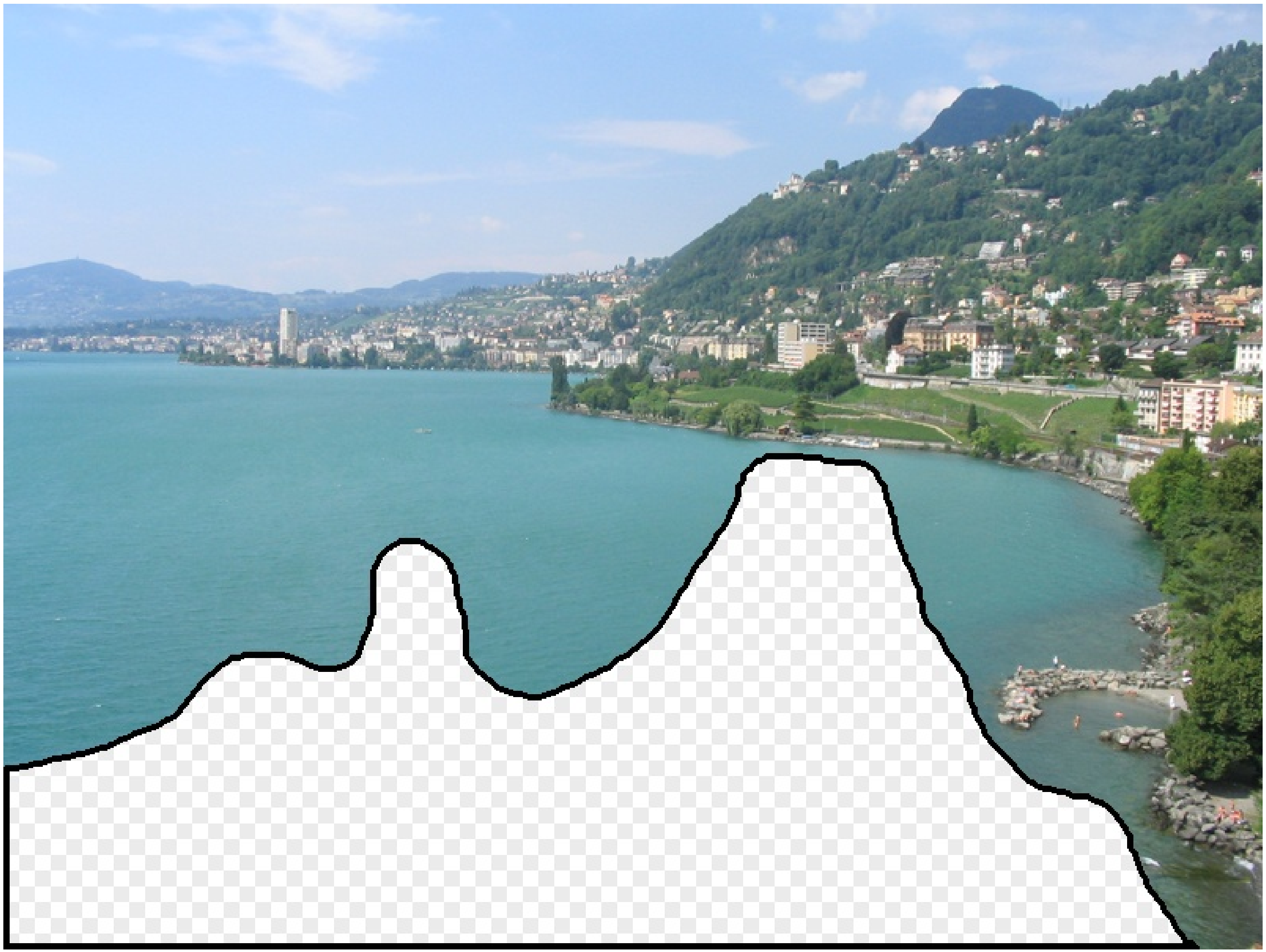
Scene Matching for Image Completion



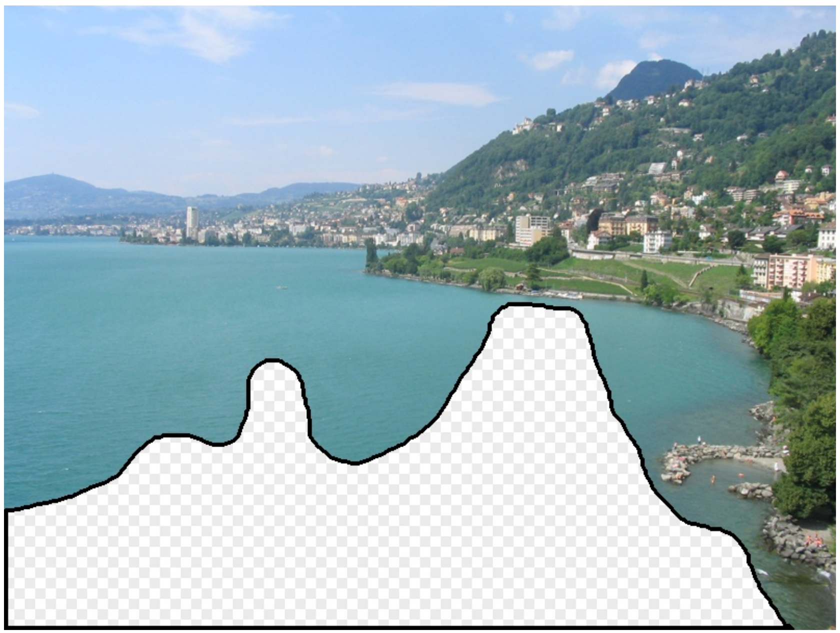
Scene Completion Result

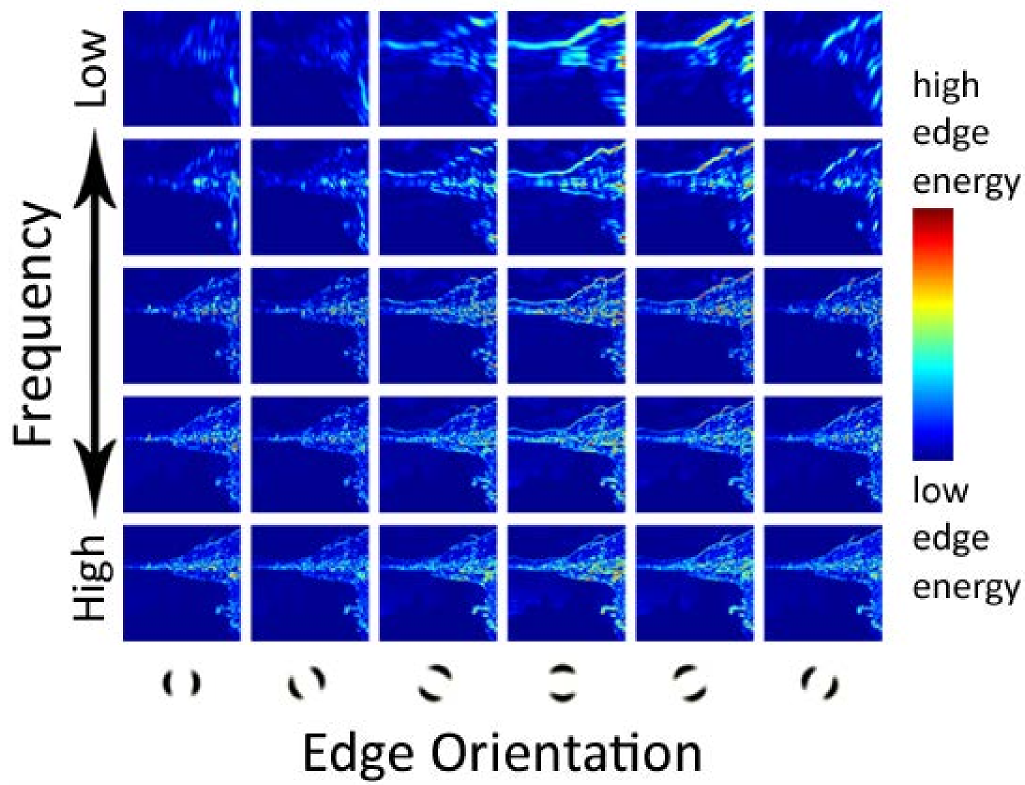
The Algorithm

Scene Matching

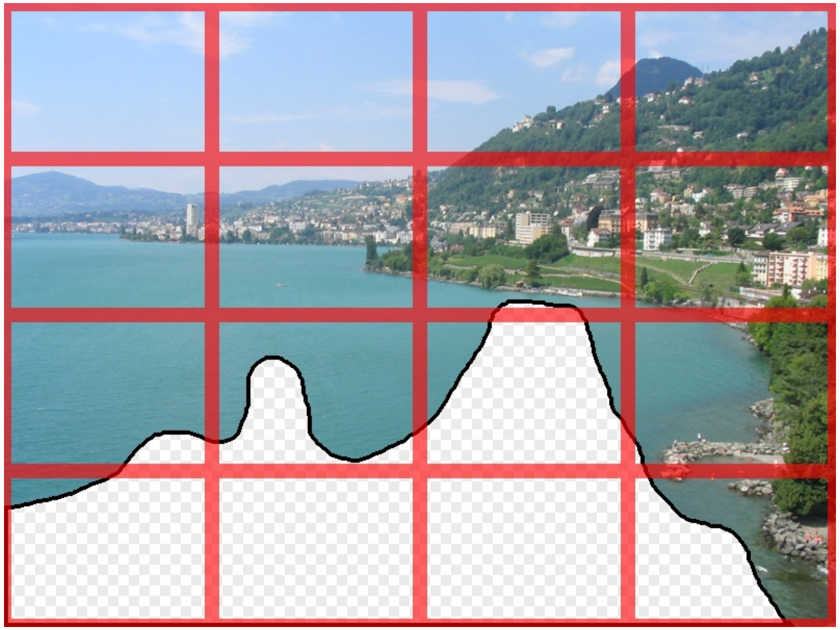


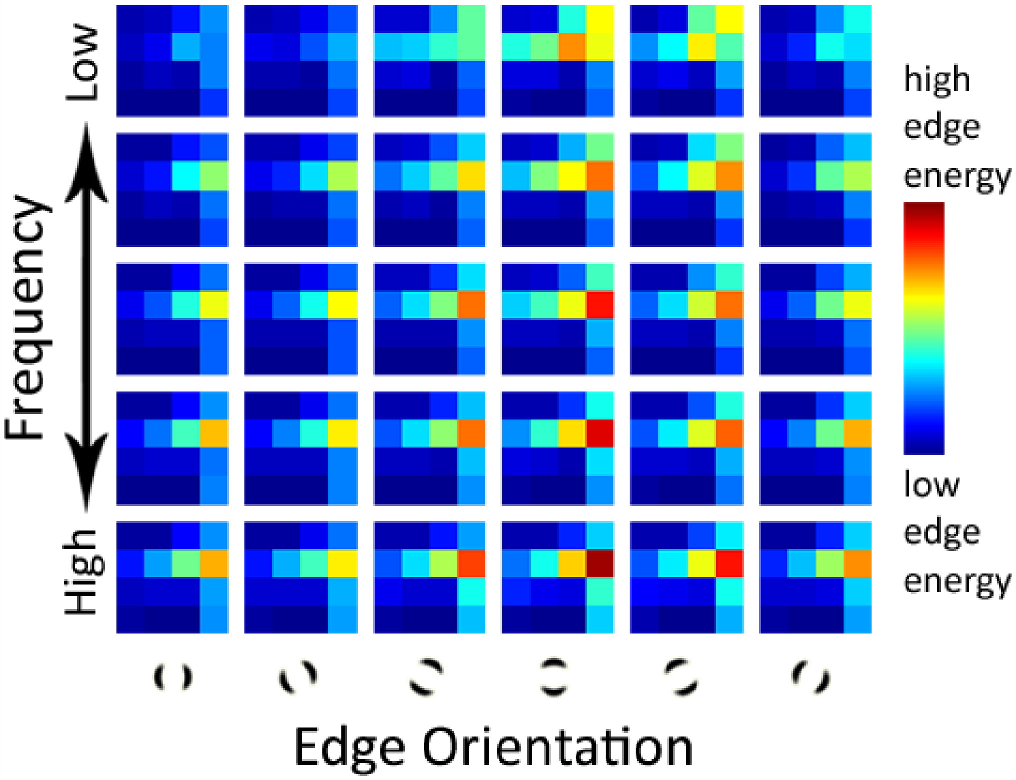
Scene Descriptor





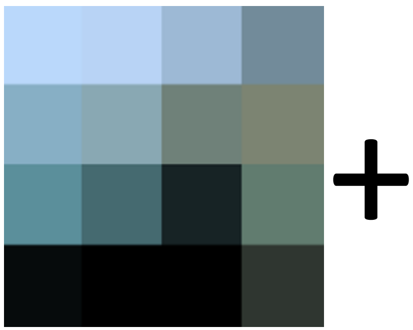
Scene Descriptor

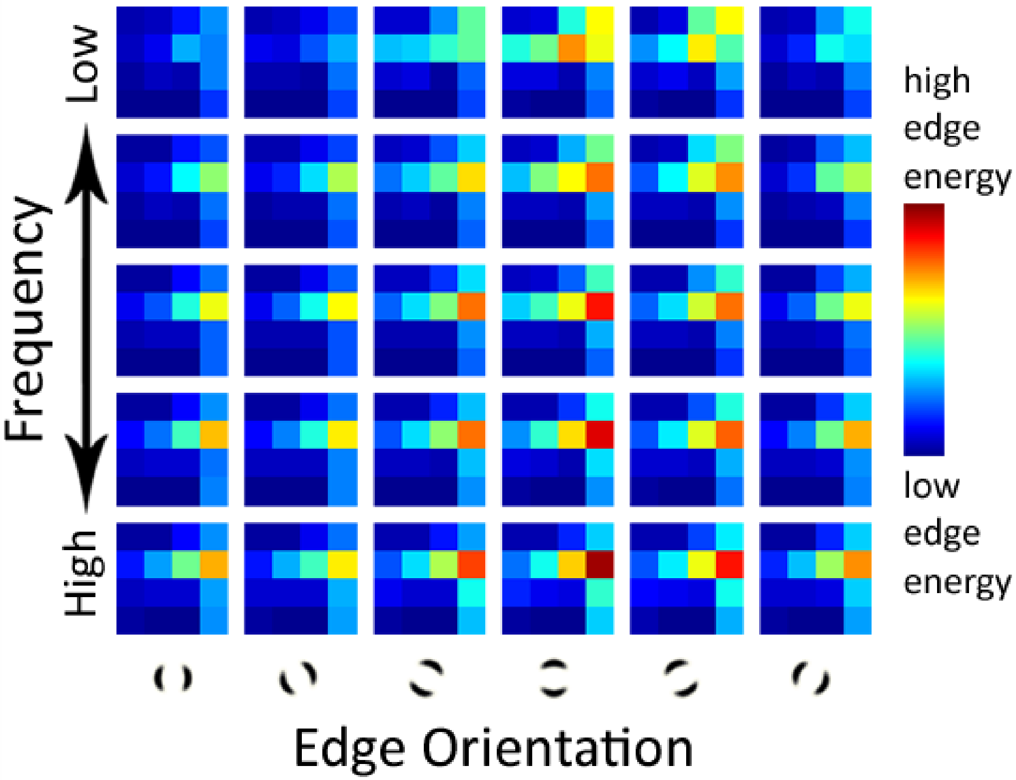




Scene Gist Descriptor
(Oliva and Torralba 2001)

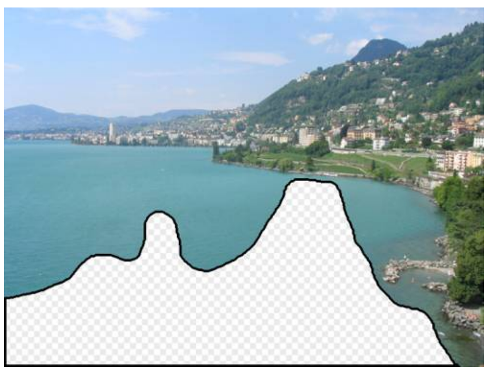
Scene Descriptor

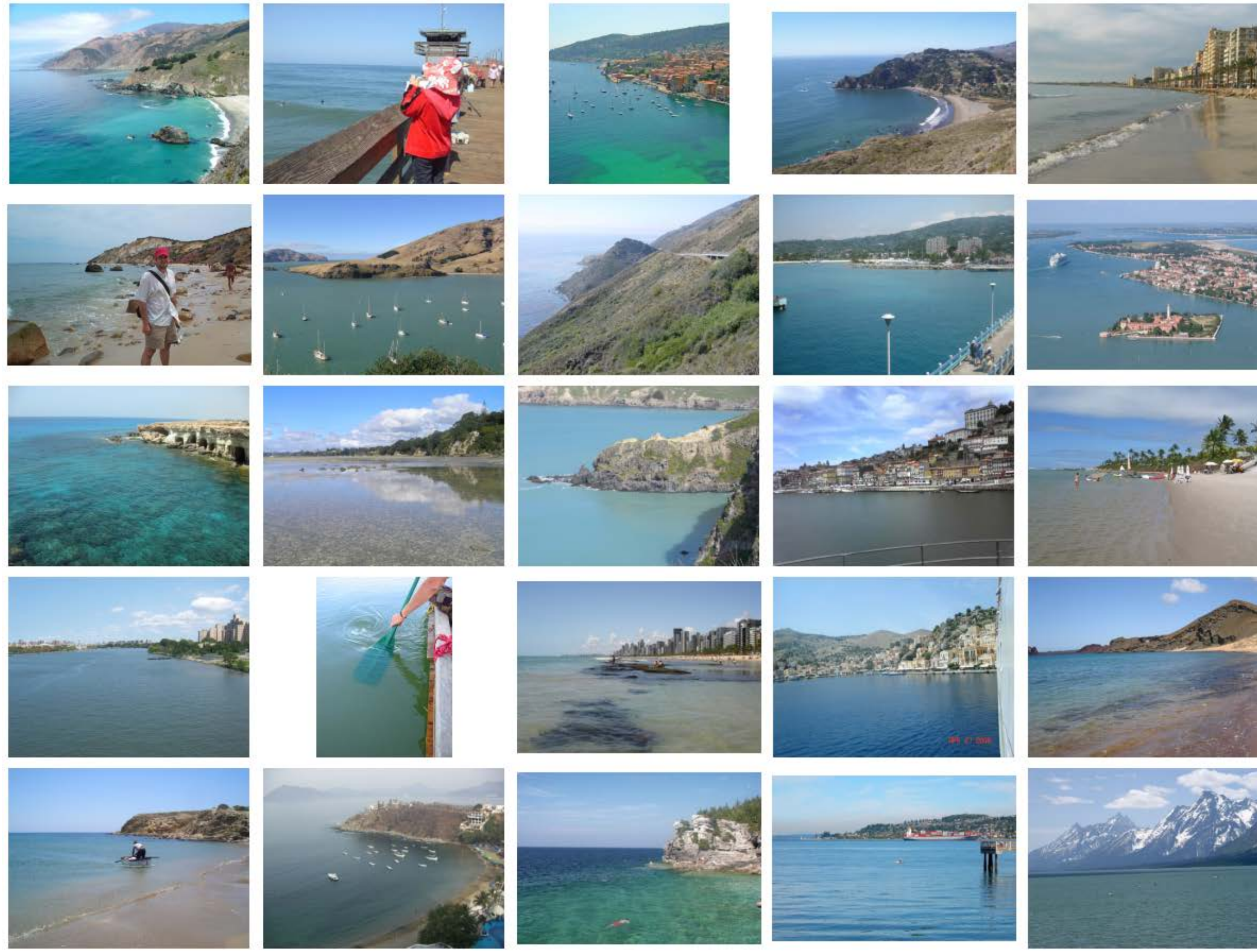




Scene Gist Descriptor
(Oliva and Torralba 2001)

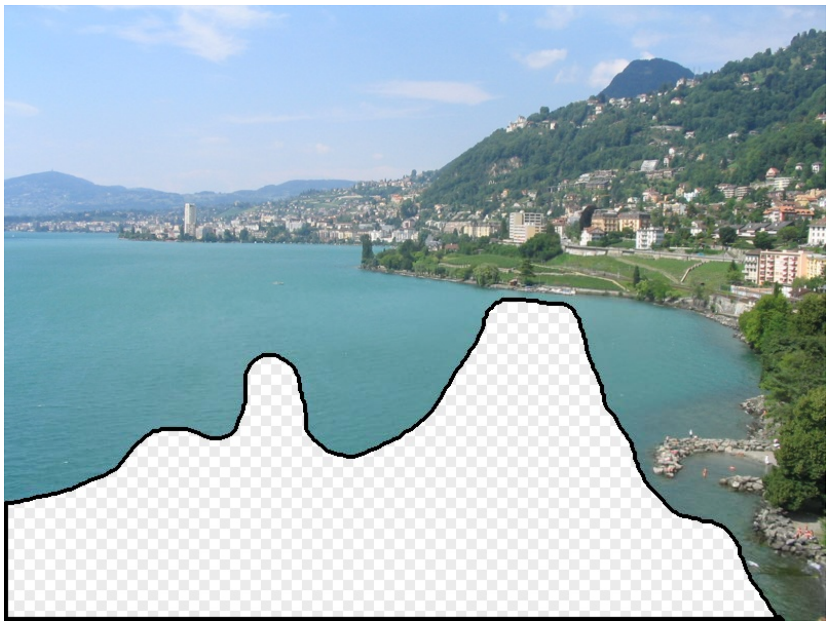
2 Million Flickr Images





… 200 total

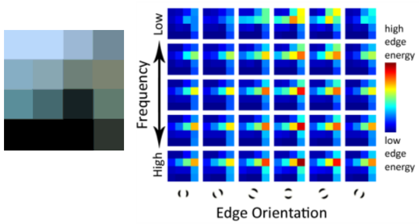
Context Matching



Graph cut + Poisson blending

Result Ranking

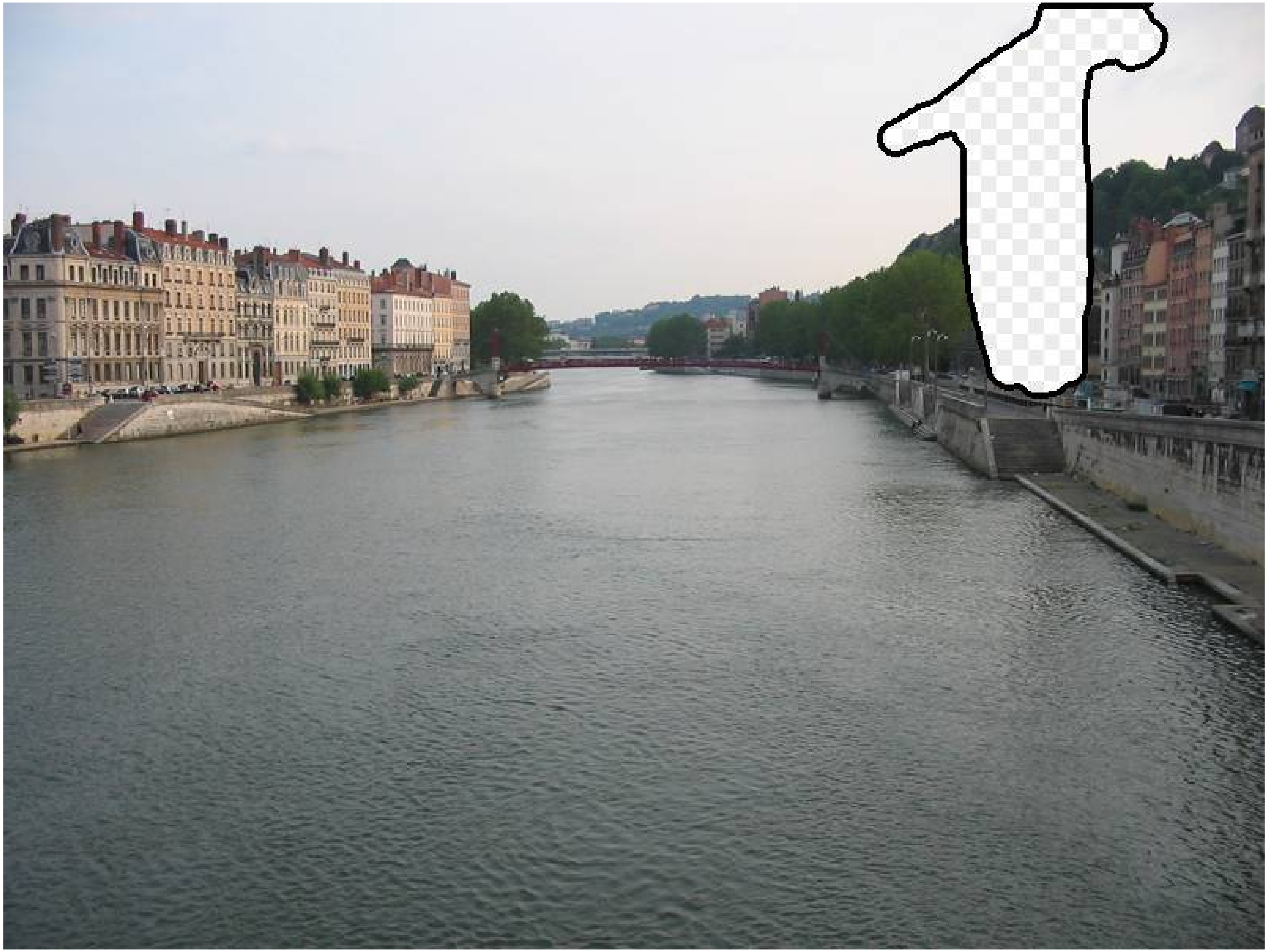
We assign each of the 200 results a score which is the sum of:

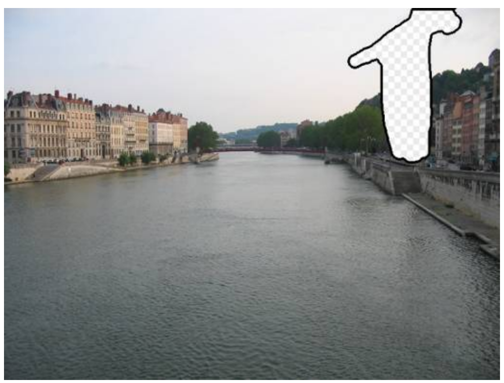


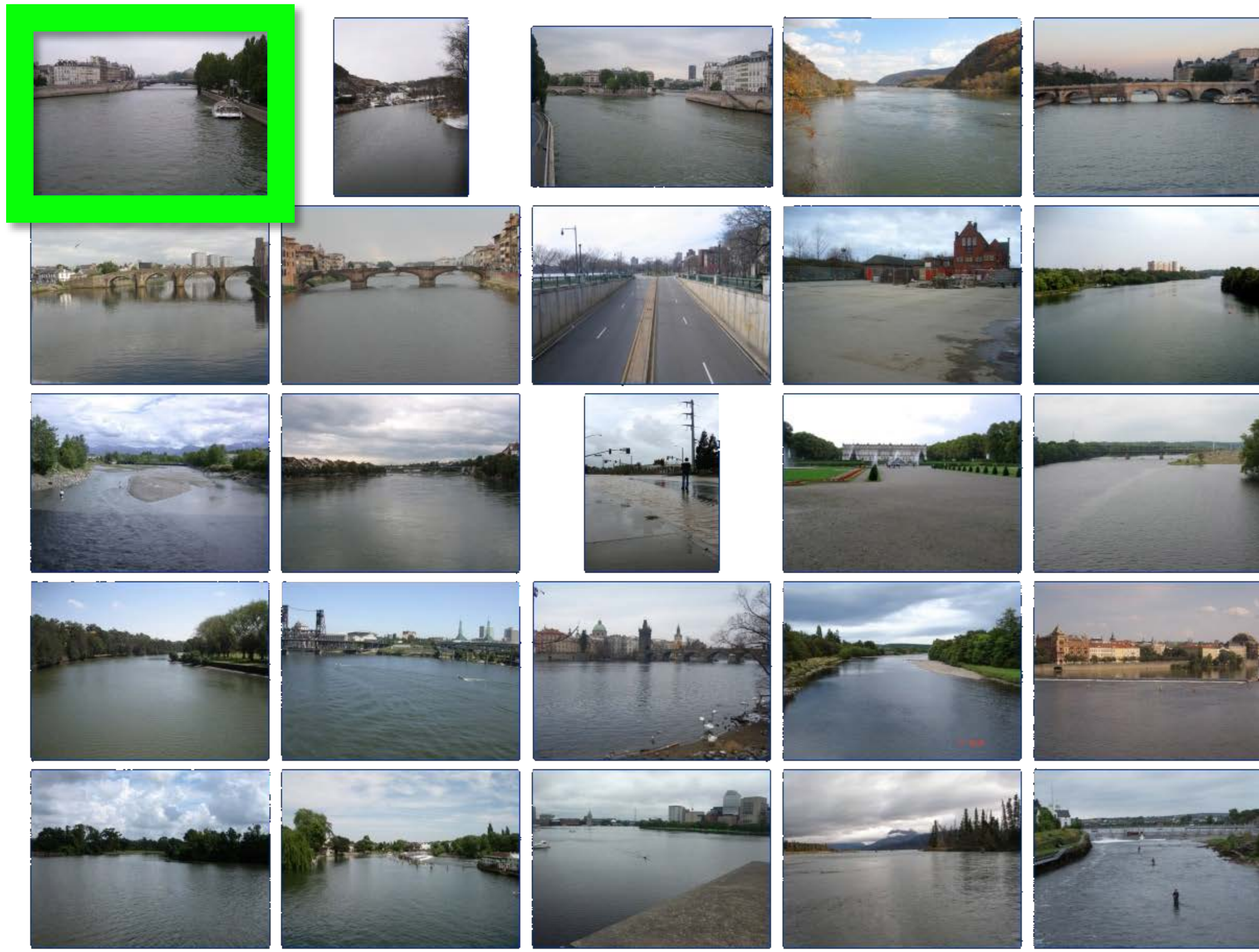
The scene matching distance

The context matching distance
(color + texture)

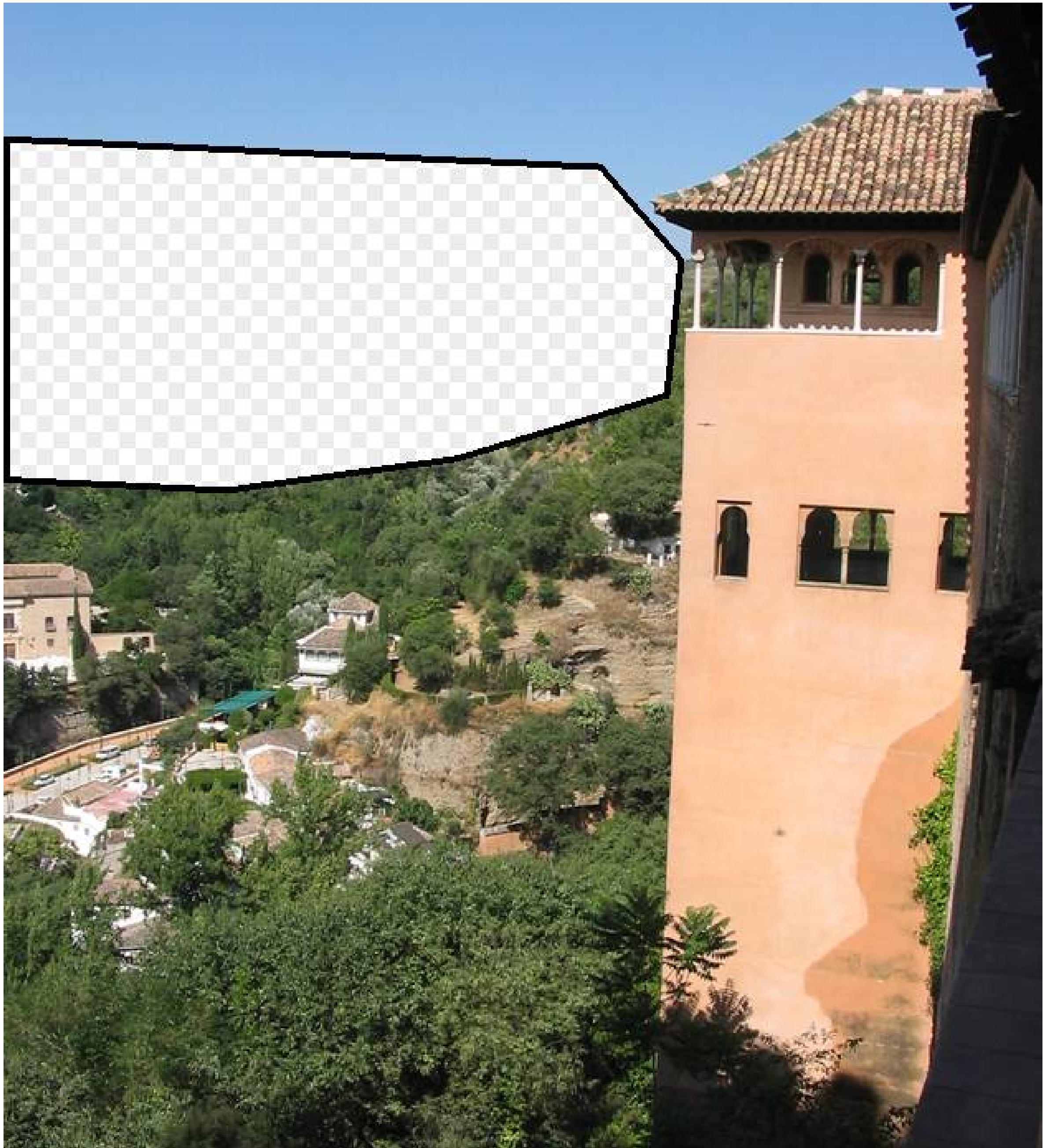
The graph cut cost

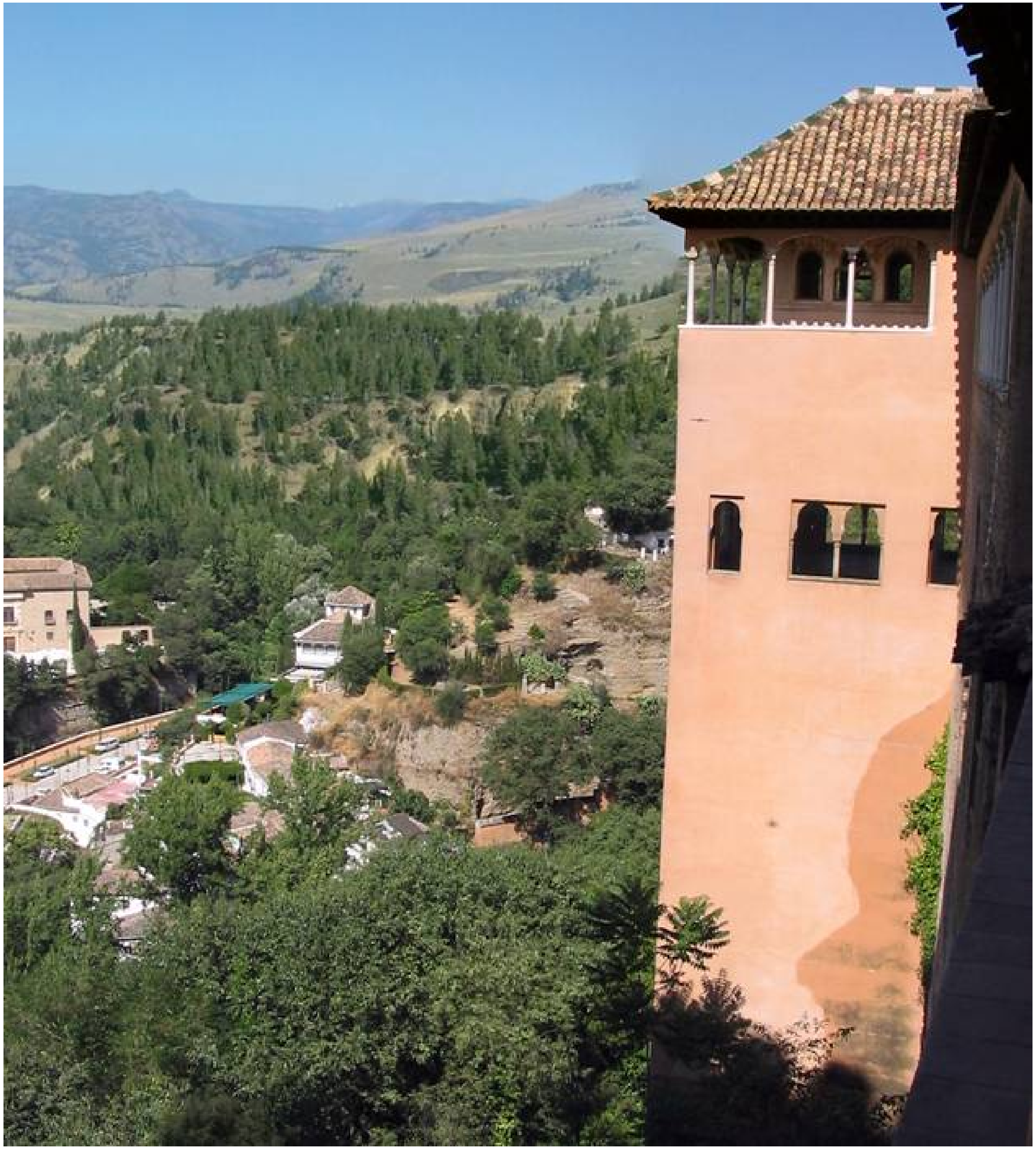




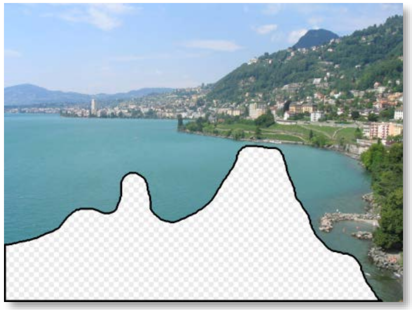


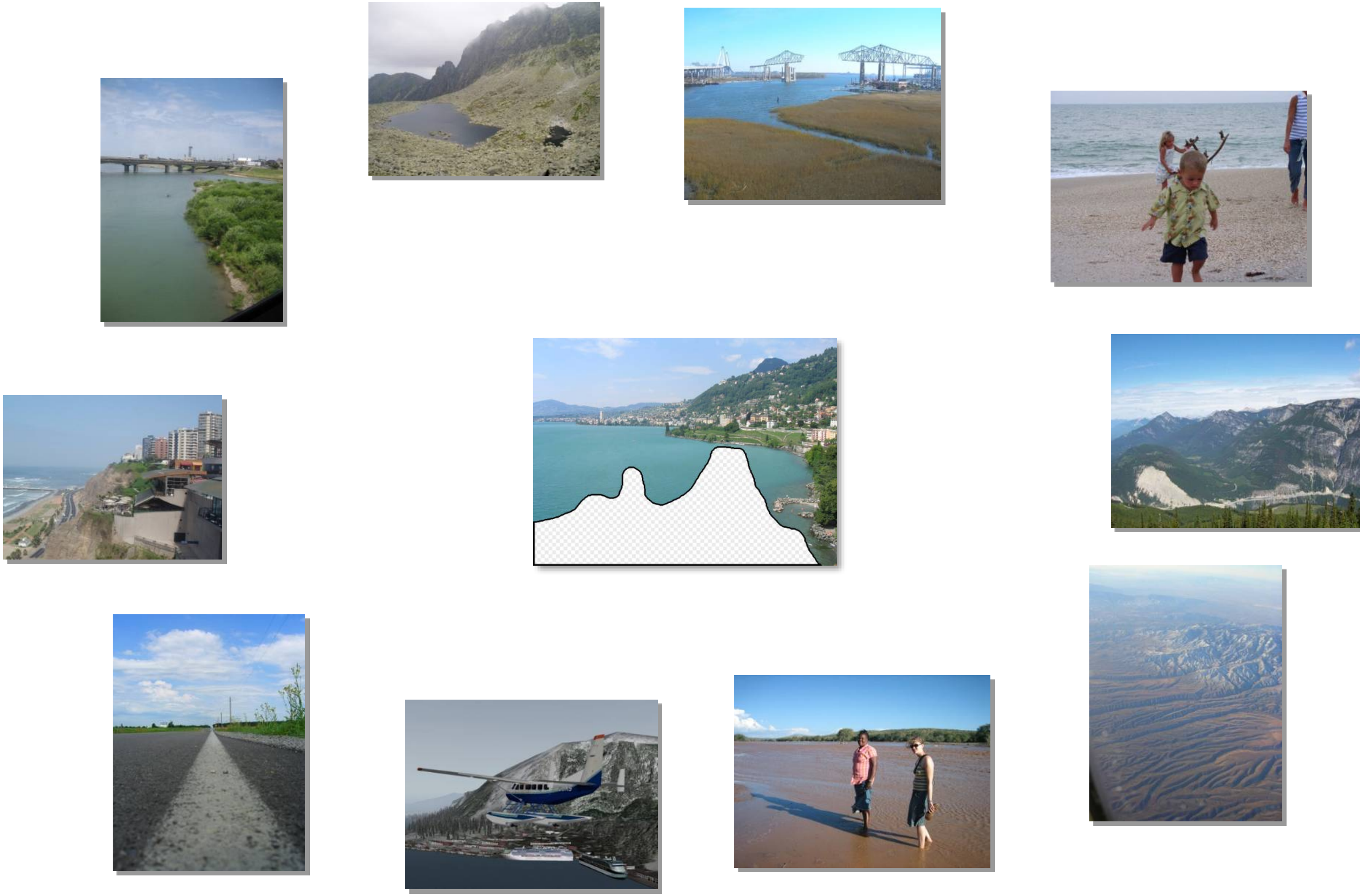
… 200 scene matches

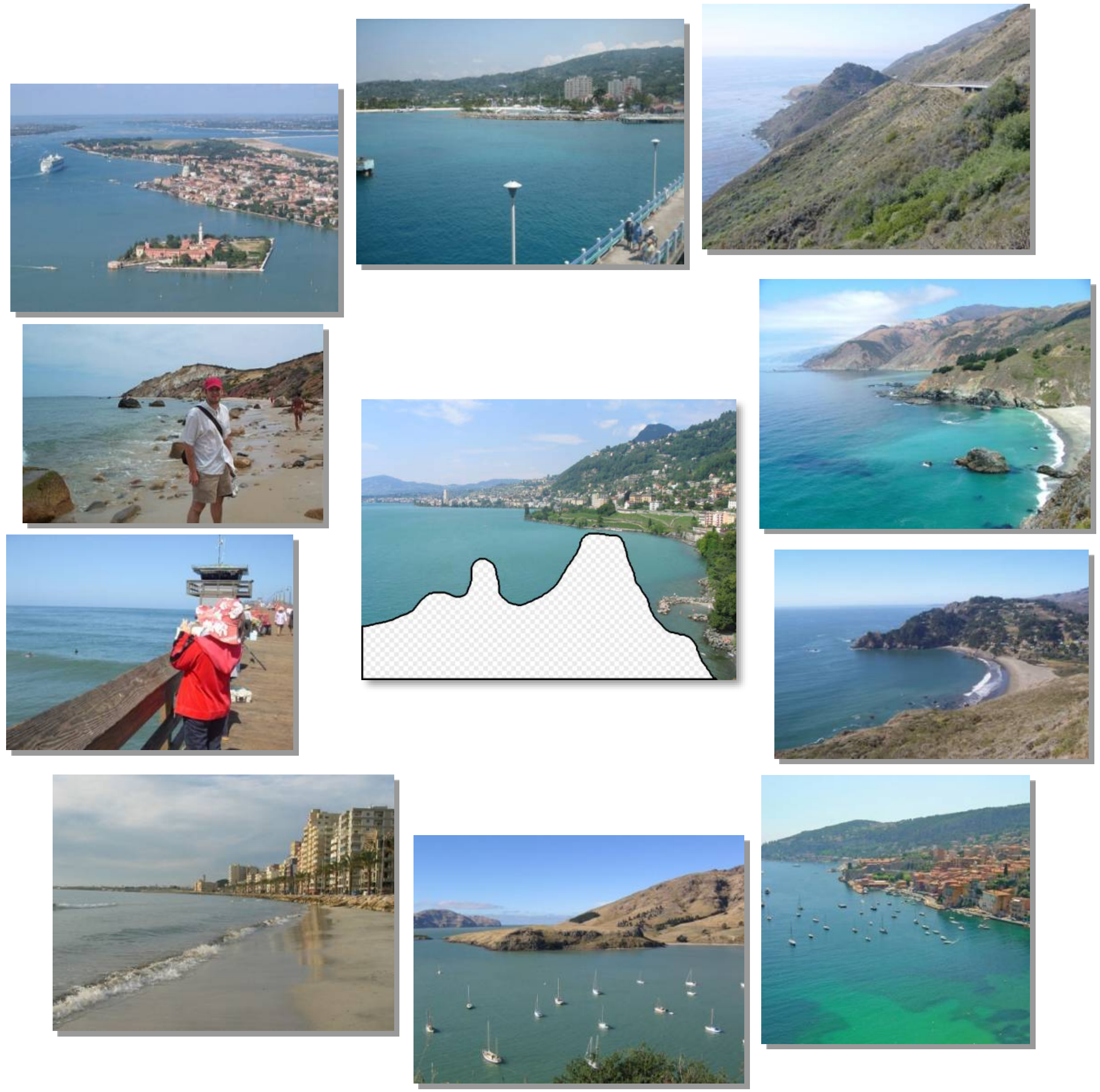


Why does it work?





Nearest neighbors from a
collection of 20 thousand images



Nearest neighbors from a collection of 2 million images

“Unreasonable Effectiveness of Data”

[Halevy, Norvig, Pereira 2009]

- Parts of our world can be explained by elegant mathematics
  - physics, chemistry, astronomy, etc.
- But much cannot
  - psychology, economics, genetics, etc.
- Enter <u>The Data!</u>
  - Great advances in several fields:
    - e.g. speech recognition, machine translation
    - Case study: Google

- 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>

How about visual data?

- text is simple:
  - – clean, segmented, compact, 1D
- Visual data is much harder:
  - – Noisy, unsegmented, high entropy, 2D/3D

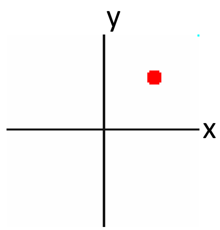
Quick Overview

Comparing Images

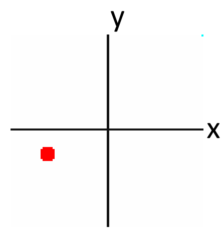
Uses of Visual Data

The Dangers of Data

Distance Metrics



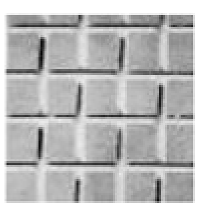
-



= Euclidian distance of 5 units

-

= Grayvalue distance of 50 values



-

= ?

SSD says these are not similar

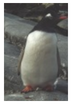
?

Gist of a scene

- Need a full image descriptor, to capture the context

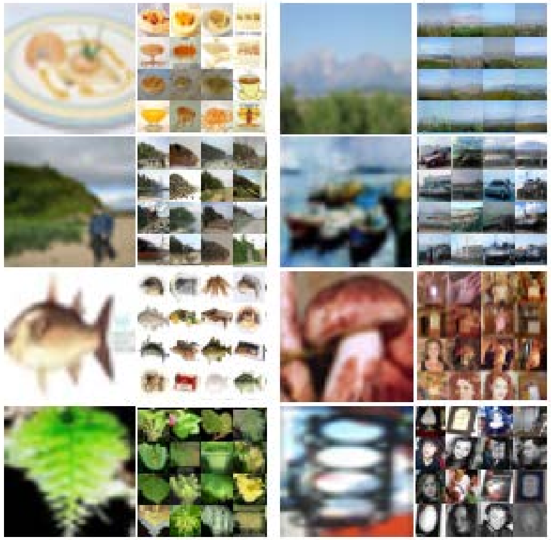
- But still want it to be not too high-dimentional (else nothing will look similar)

Make them tiny!



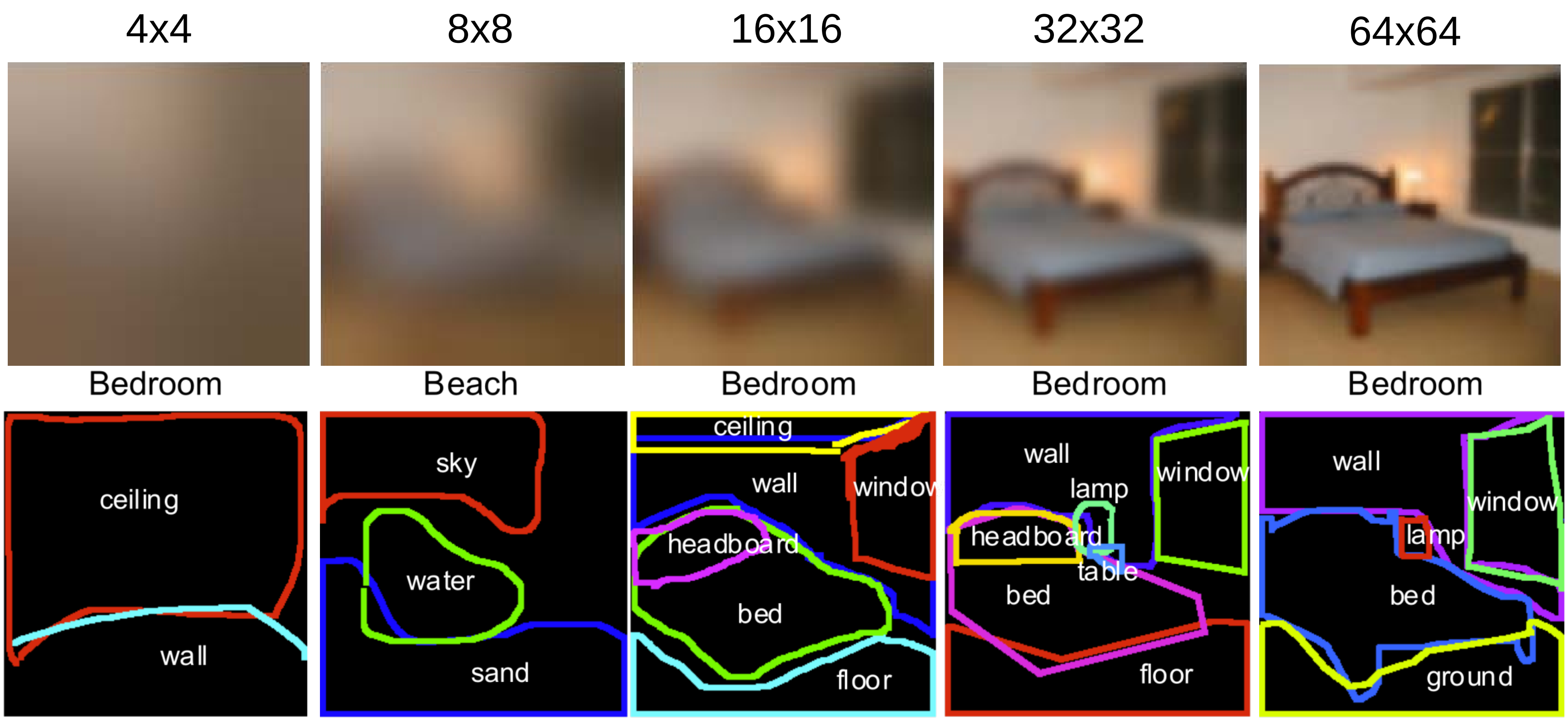
?

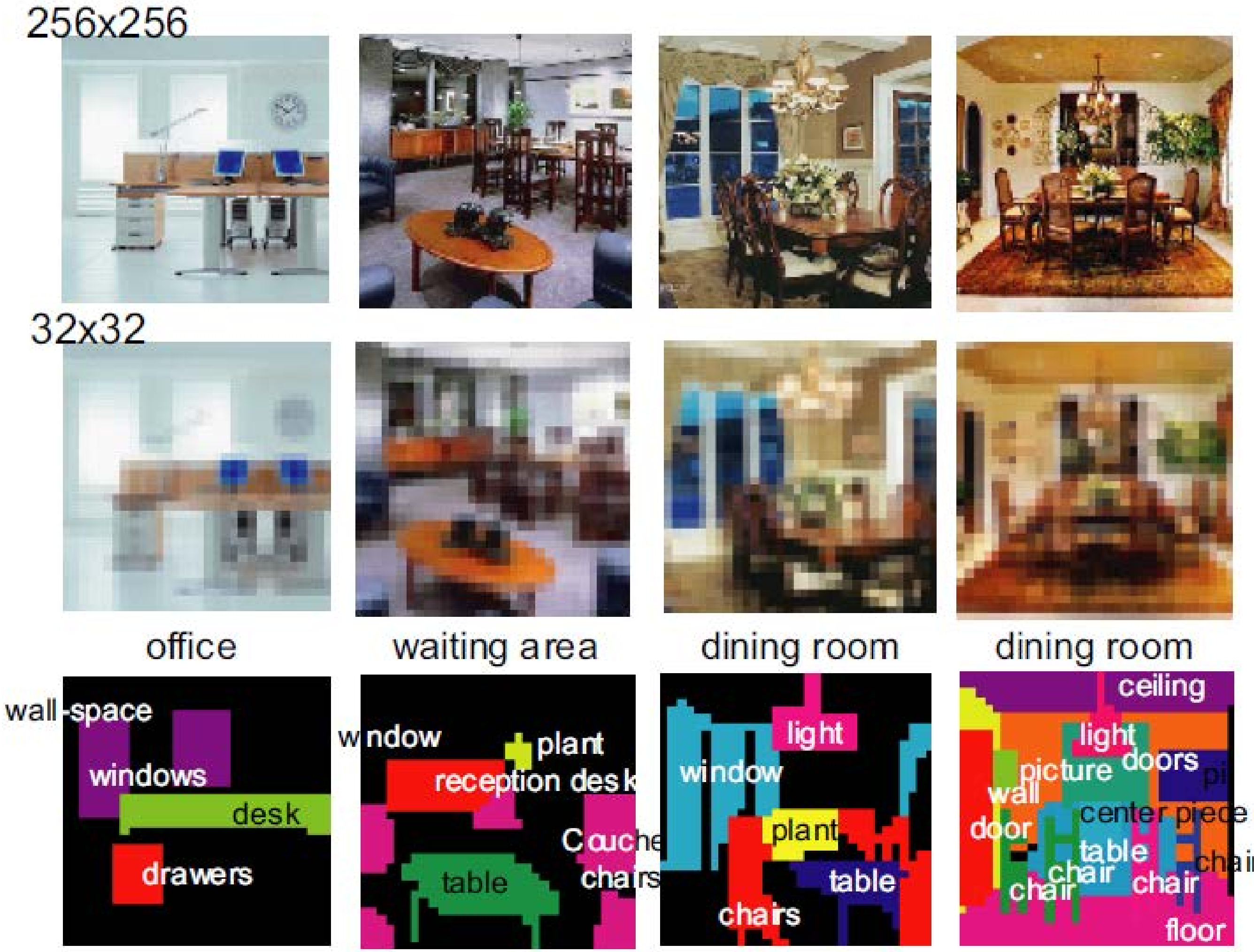
Tiny Images



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

Tiny Images pack a punch!

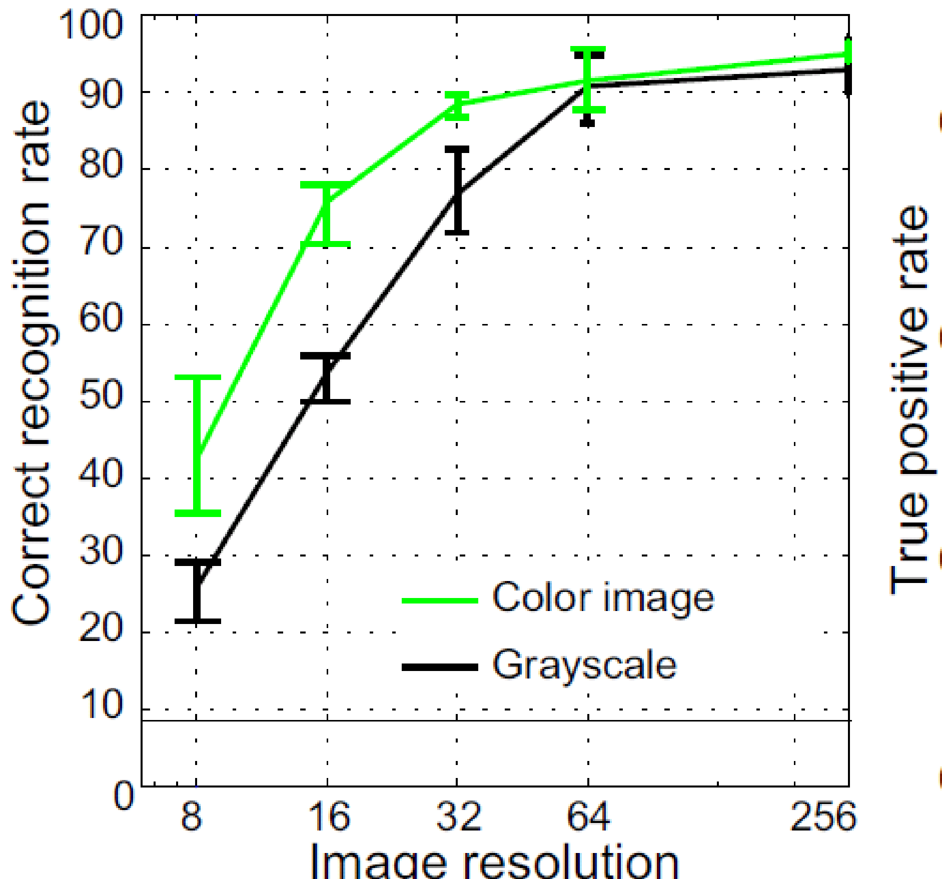




c) Segmentation of 32x32 images

Image Segmentation (by humans)

Human Scene Recognition



a) Scene recognition

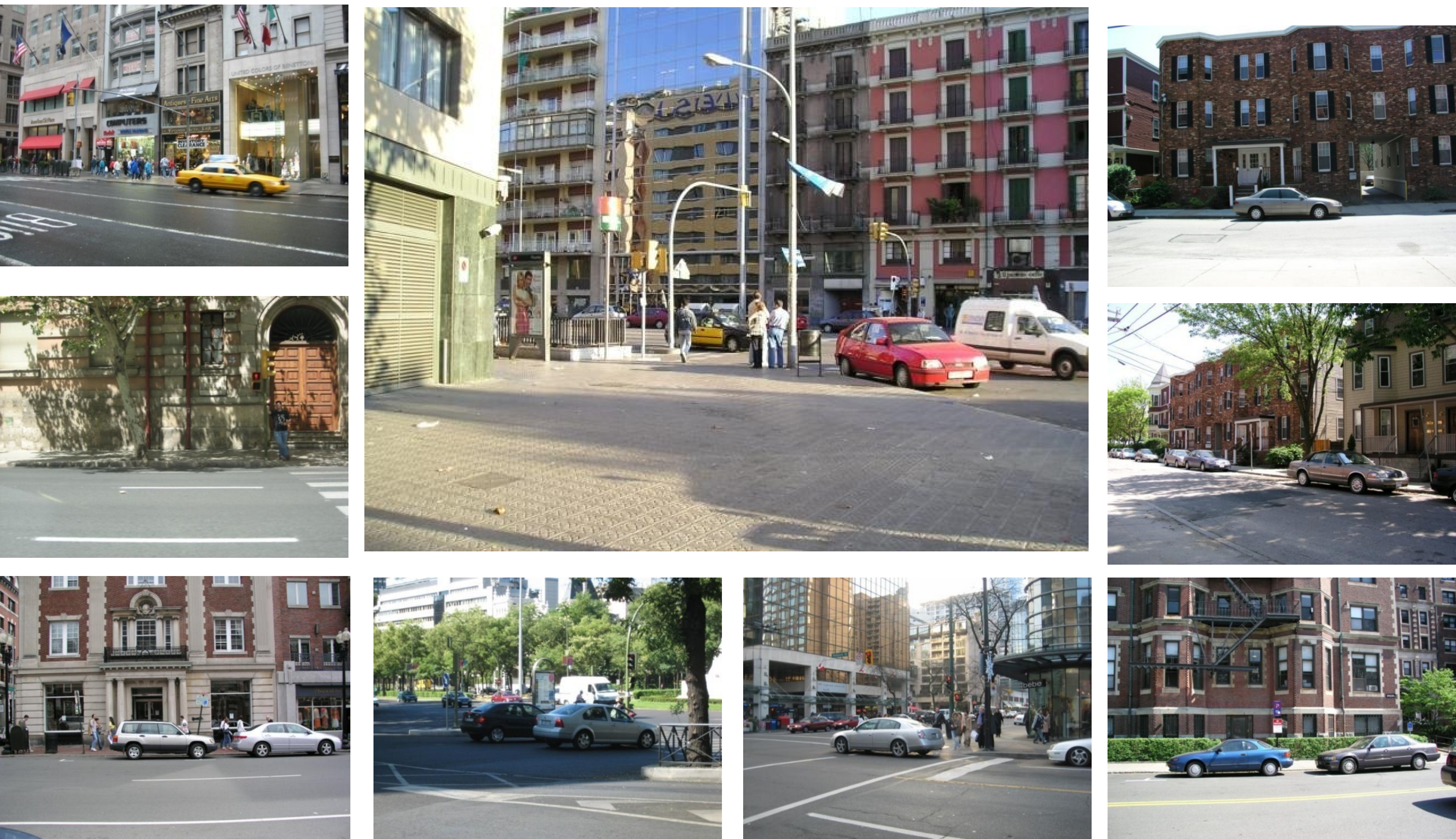
Tiny Images Project Page

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

Scenes are unique

But not all scenes are so original

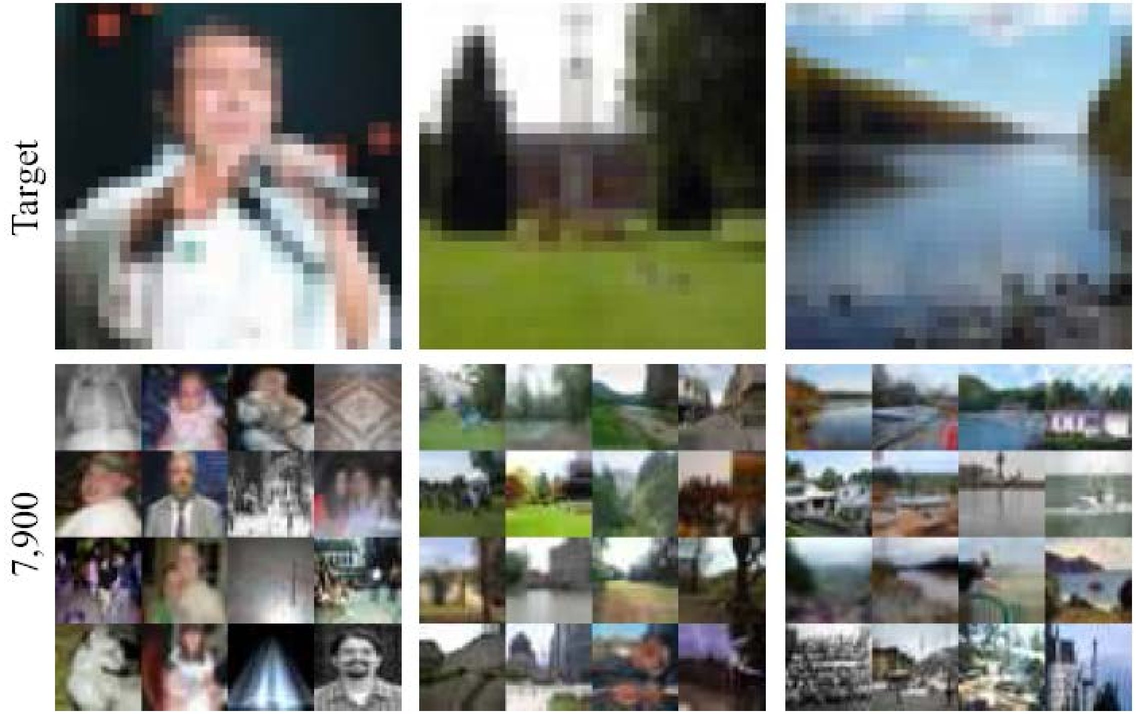
But not all scenes are so original



Lots

Of

Images

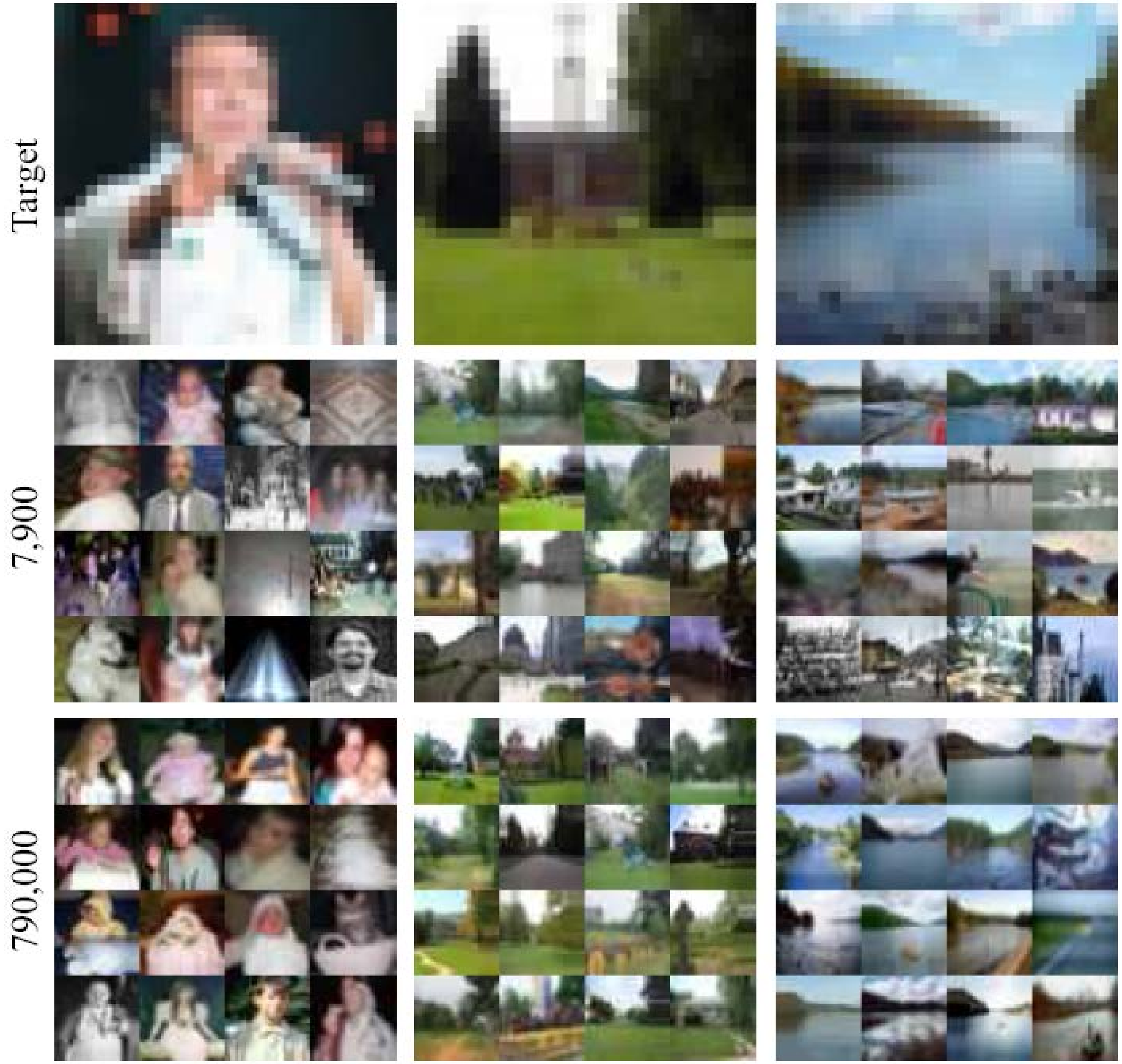


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

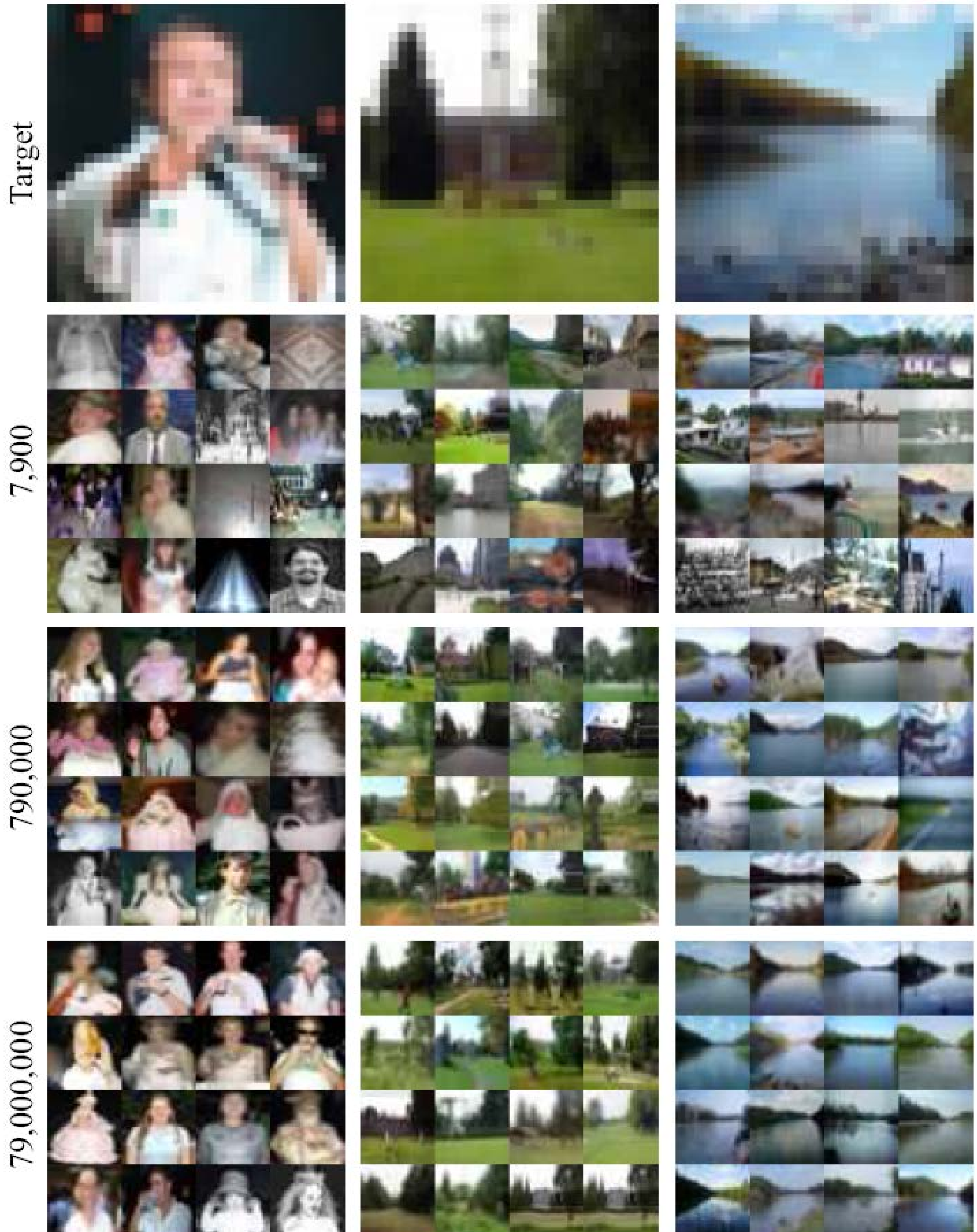


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

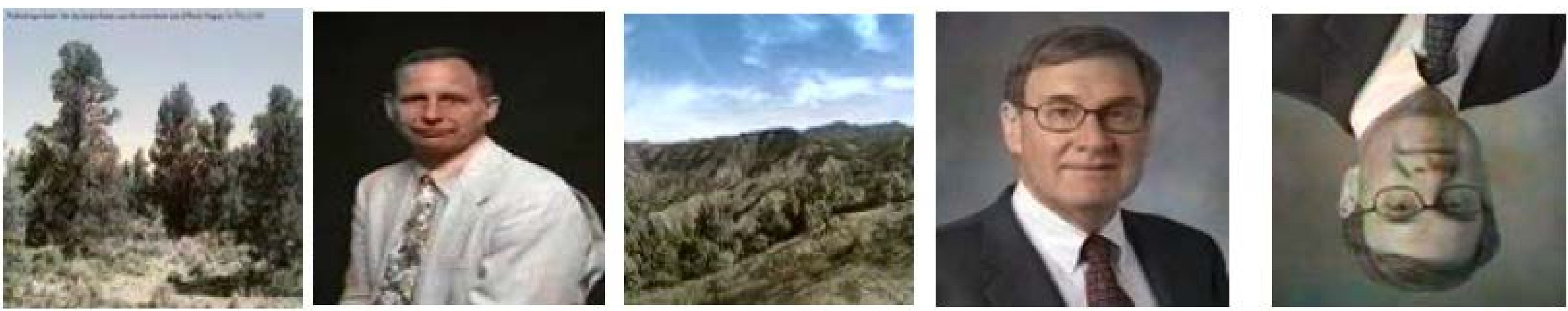
Images



Automatic Colorization Result

Grayscale input High resolution

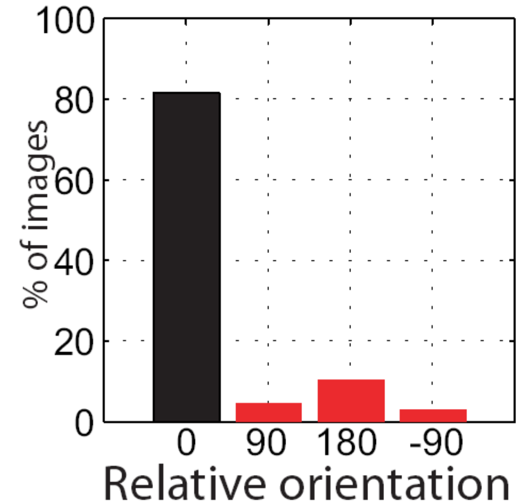
Colorization of input using average



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

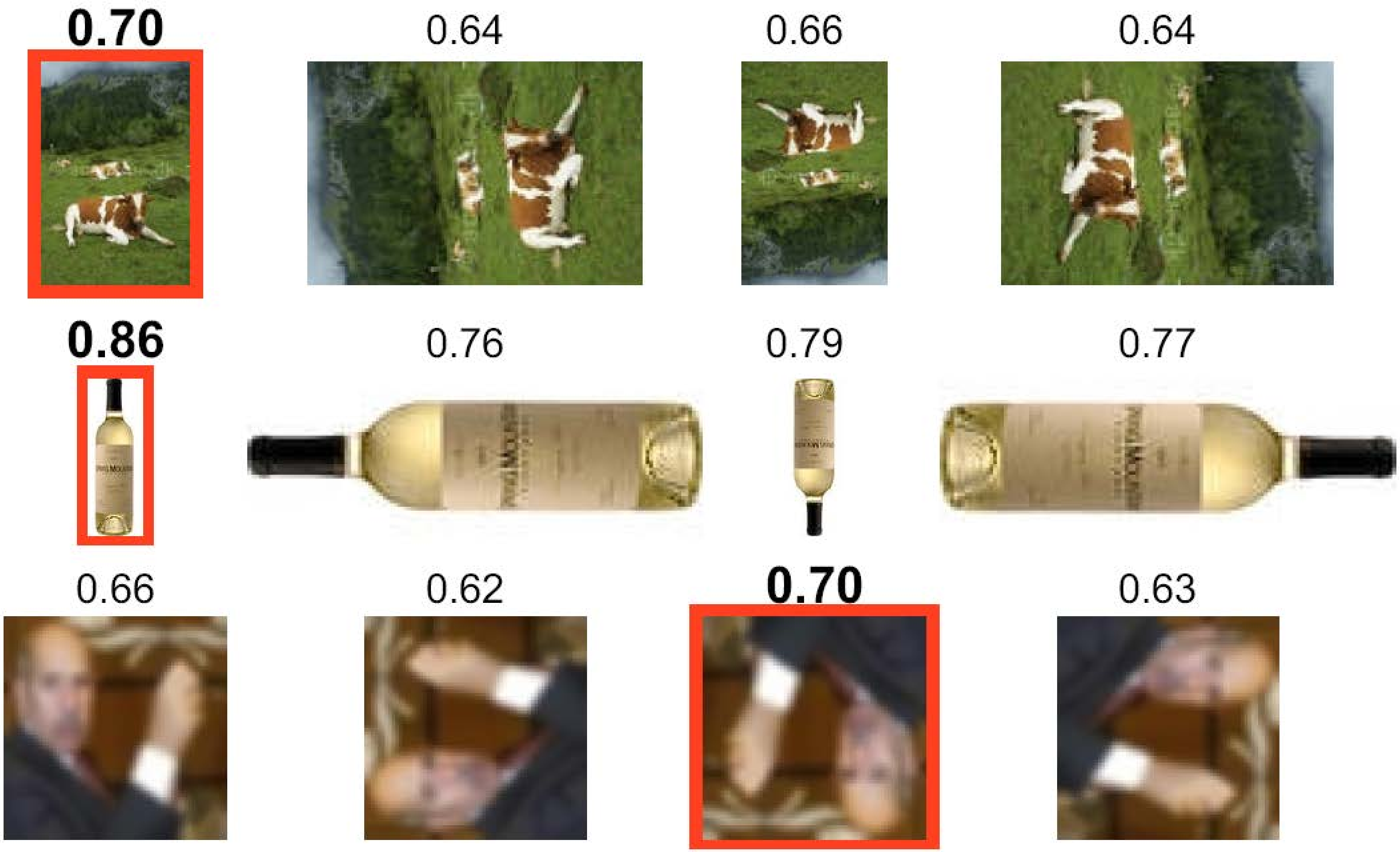
Automatic Orientation

- Many images have ambiguous orientation
- Look at top 25% by confidence:

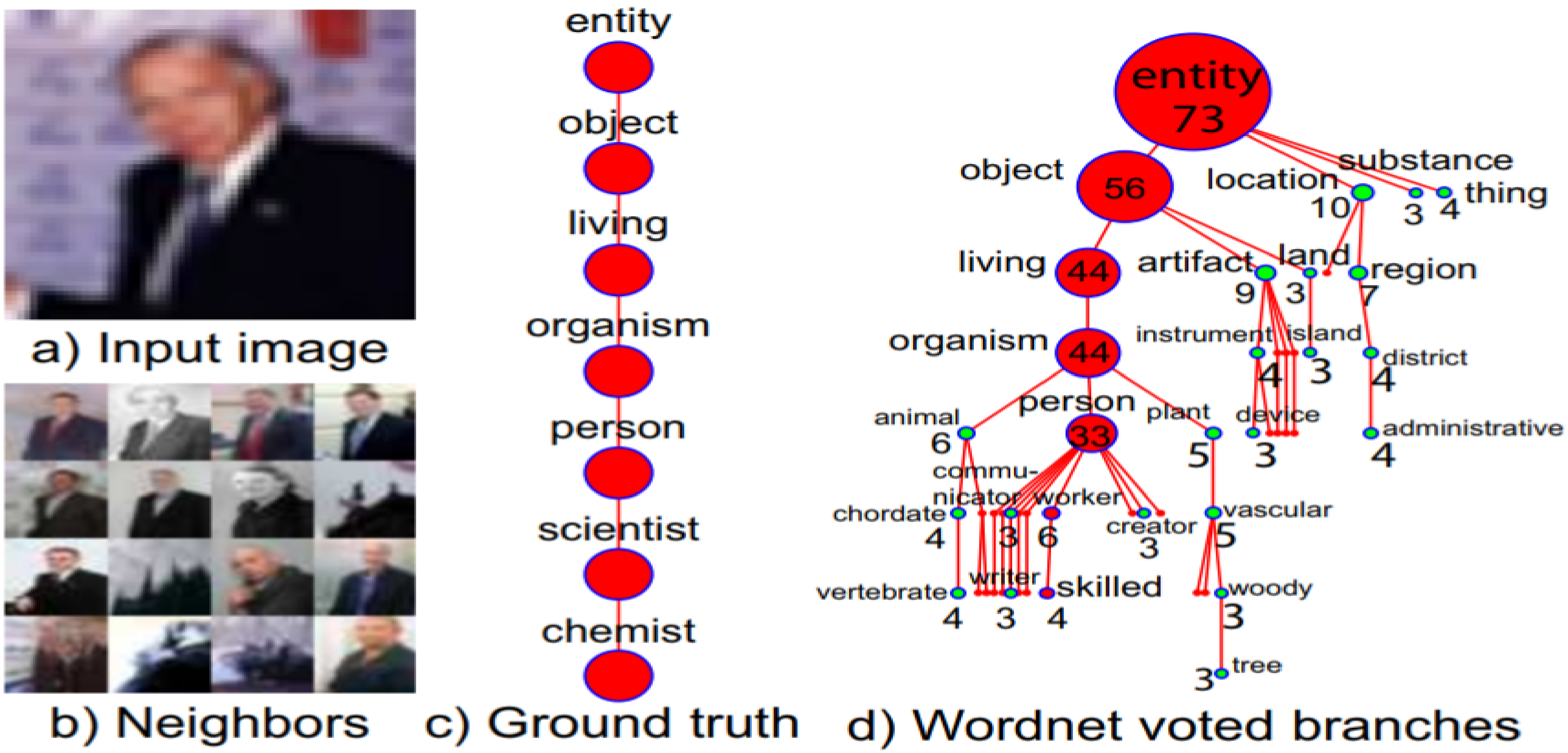


- Examples of high and low confidence images:

Automatic Orientation Examples



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

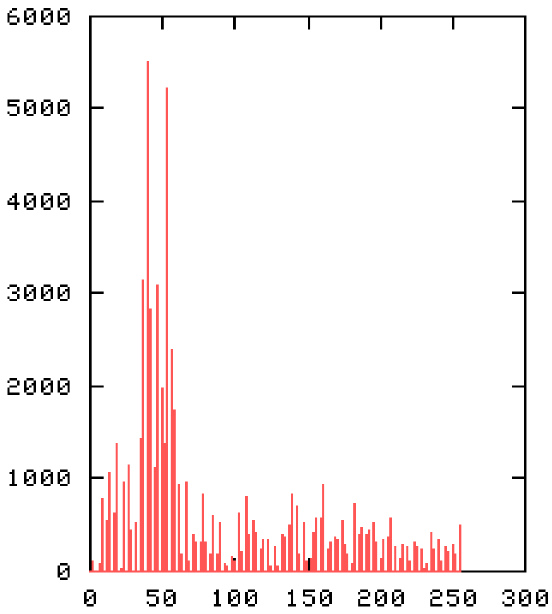

a) Input image
b) Neighbors
entity
object
living
organism
person
scientist
chemist
c) Ground truth
entity 73
object 56
location 10
substance 3
thing 4
living 44
artifact 9
land 3
region 7
organism 44
instrument 4
island 3
district 4
animal 6
person 33
plant 5
device 3
administrative 4
commu-nicator 3
worker 6
creator 3
chordate 4
vascular 5
vertebrate 4
writer 3
skilled 4
woody 3
tree 3
d) Wordnet voted branches

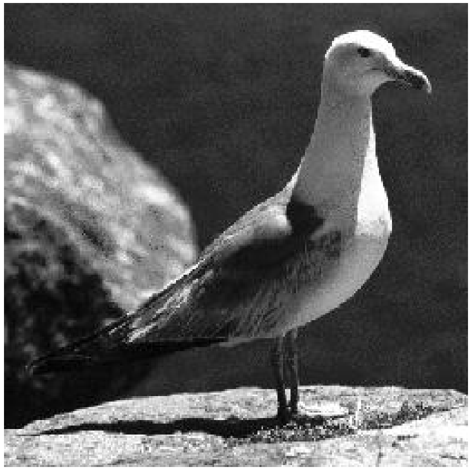
Tiny Images Discussion

- Why SSD?
- Can we build a better image descriptor?

Image Representations: Histograms

Images from Dave Kauchak



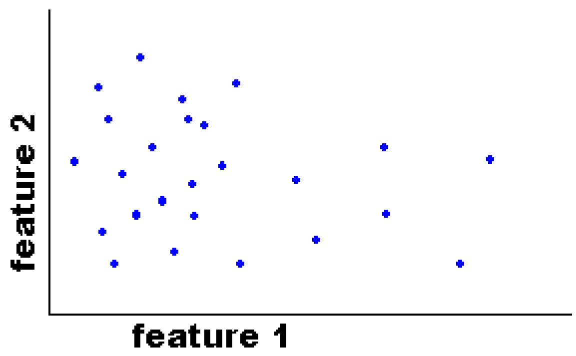


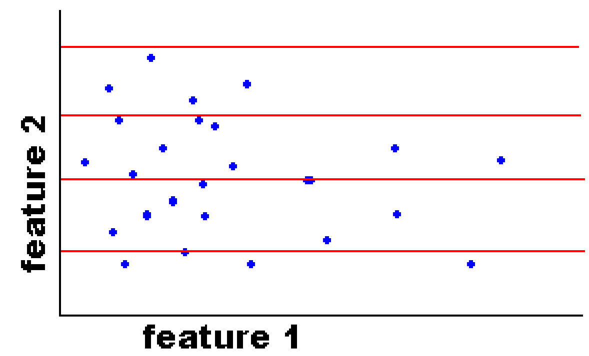
global histogram

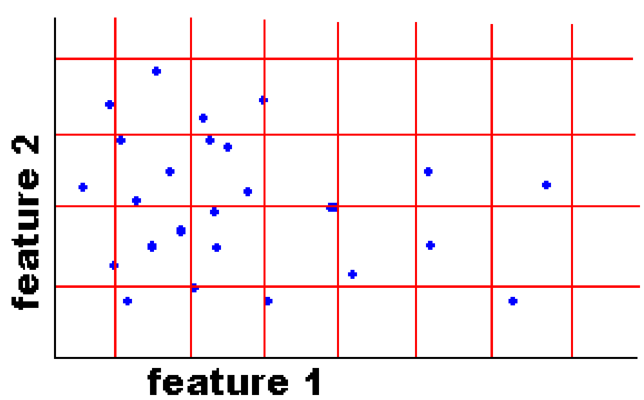
- Represent distribution of features
  - Color, texture, depth, …

Image Representations: Histograms

Images from Dave Kauchak

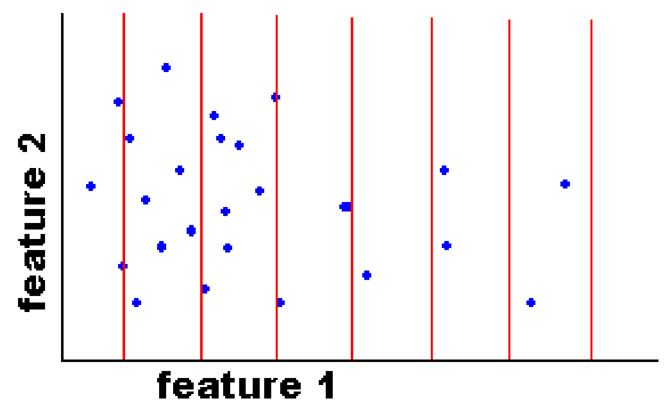






Joint histogram

- Requires lots of data
- Loss of resolution to avoid empty bins

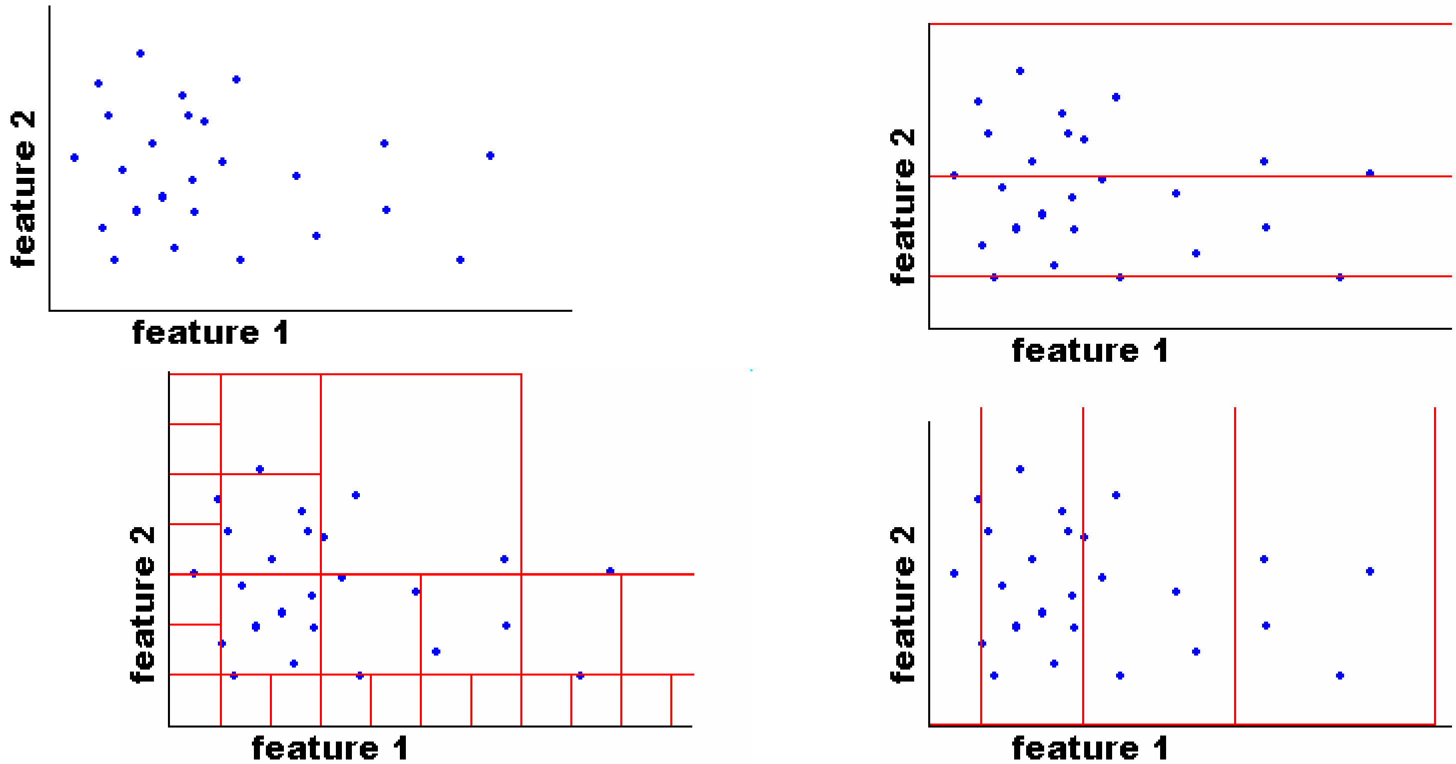


Marginal histogram

- Requires independent features
- More data/bin than joint histogram

Image Representations: Histograms

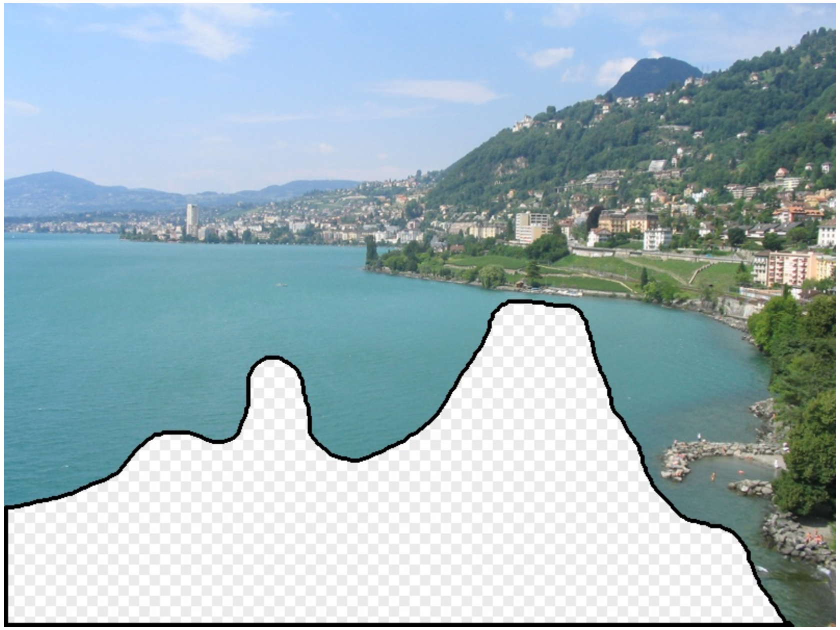
Images from Dave Kauchak

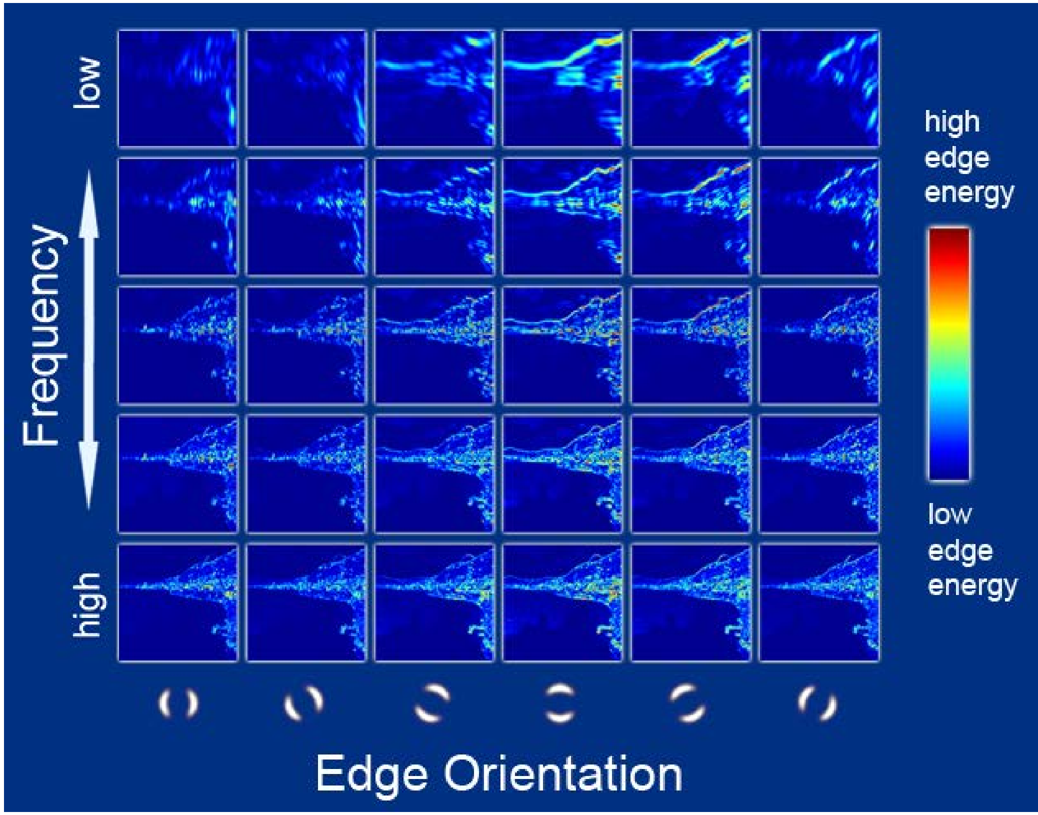


Adaptive binning

- Better data/bin distribution, fewer empty bins
- Can adapt available resolution to relative feature importance

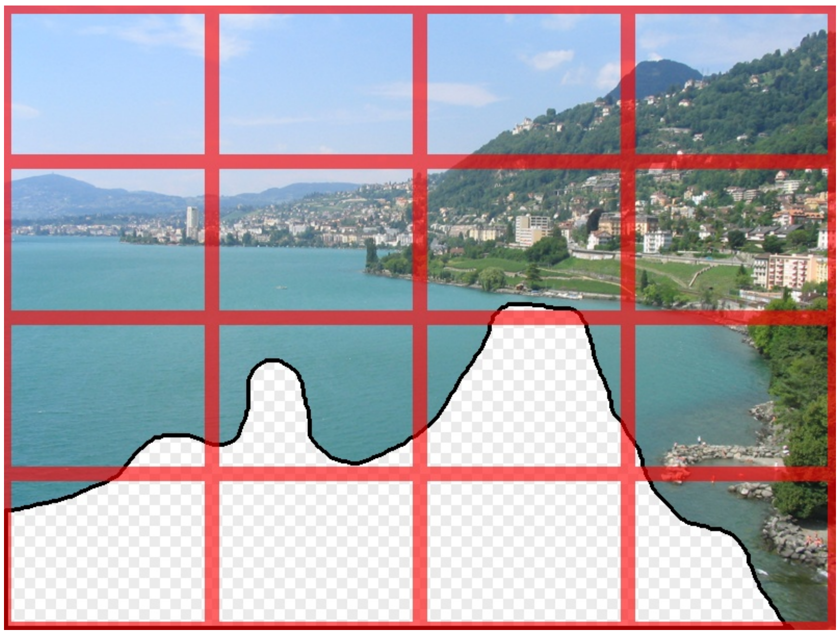
Gist Scene Descriptor

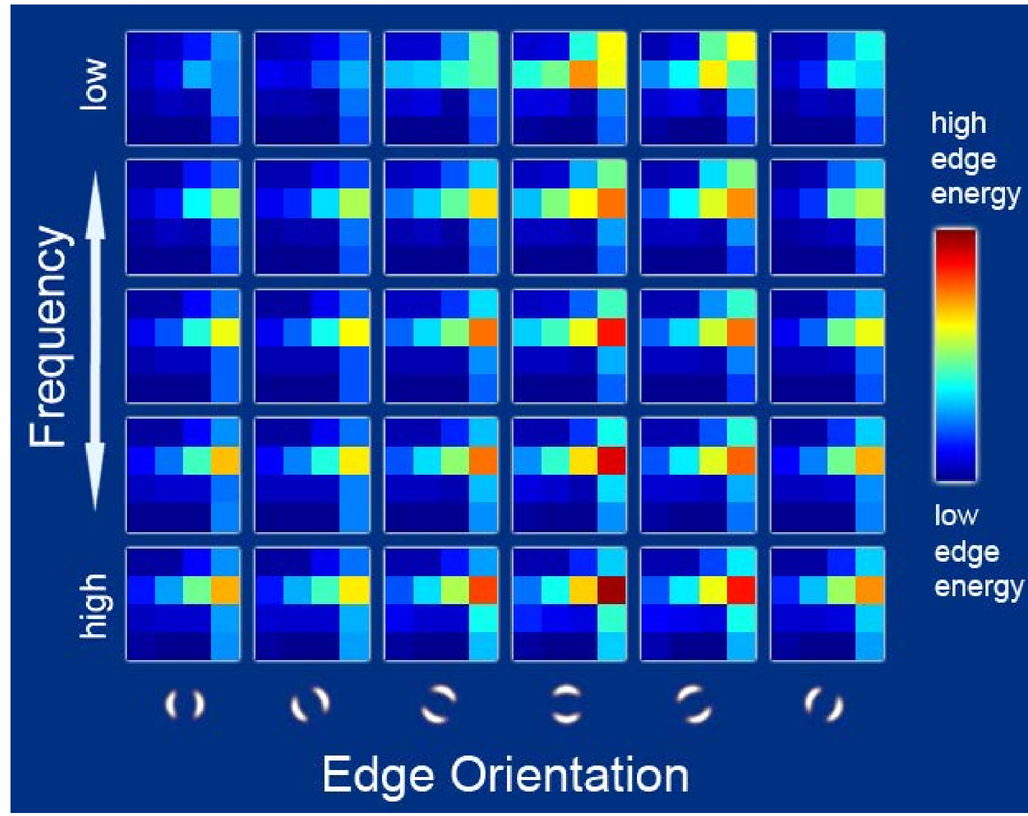




Hays and Efros, SIGGRAPH 2007

Gist Scene Descriptor

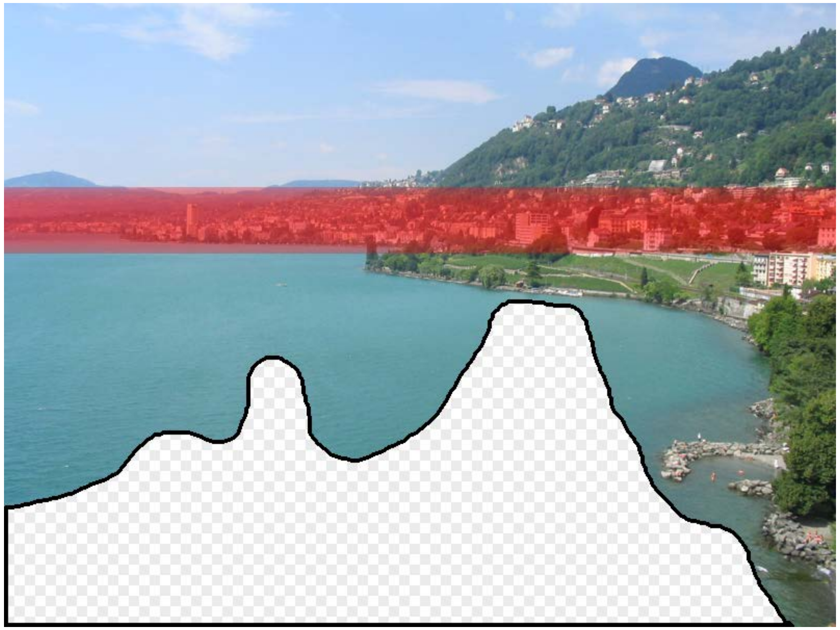


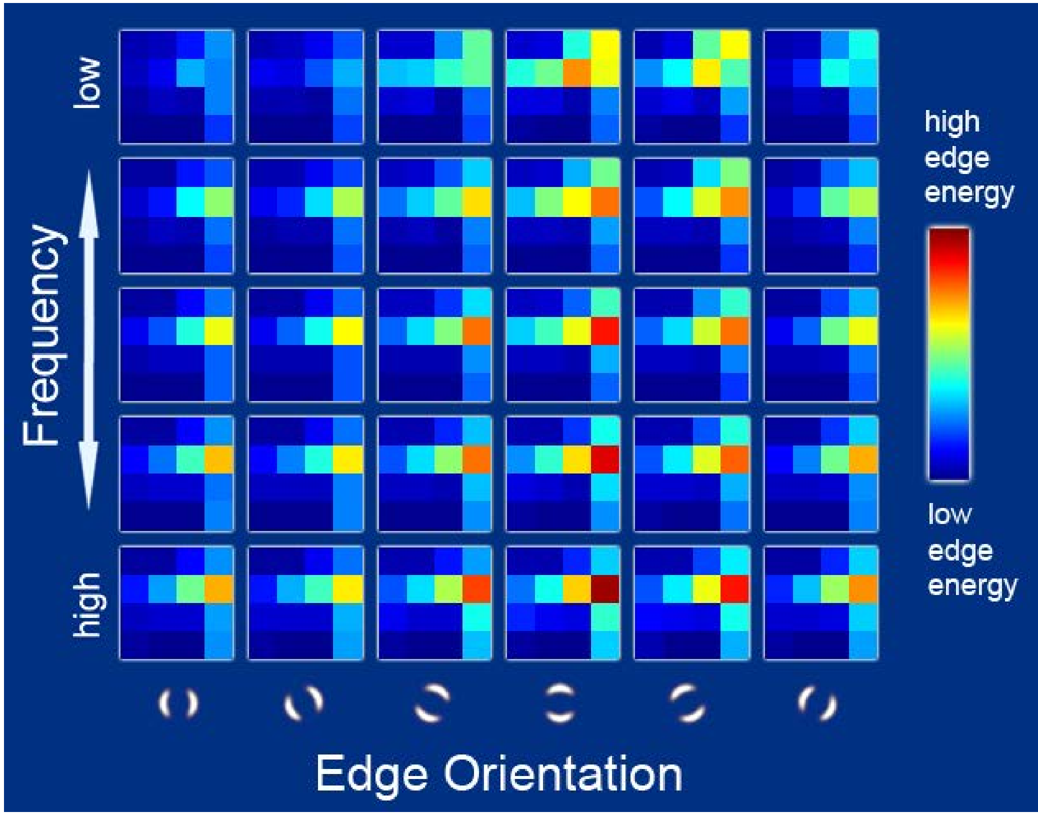


Gist scene descriptor
(Oliva and Torralba 2001)

Hays and Efros, SIGGRAPH 2007

Gist Scene Descriptor

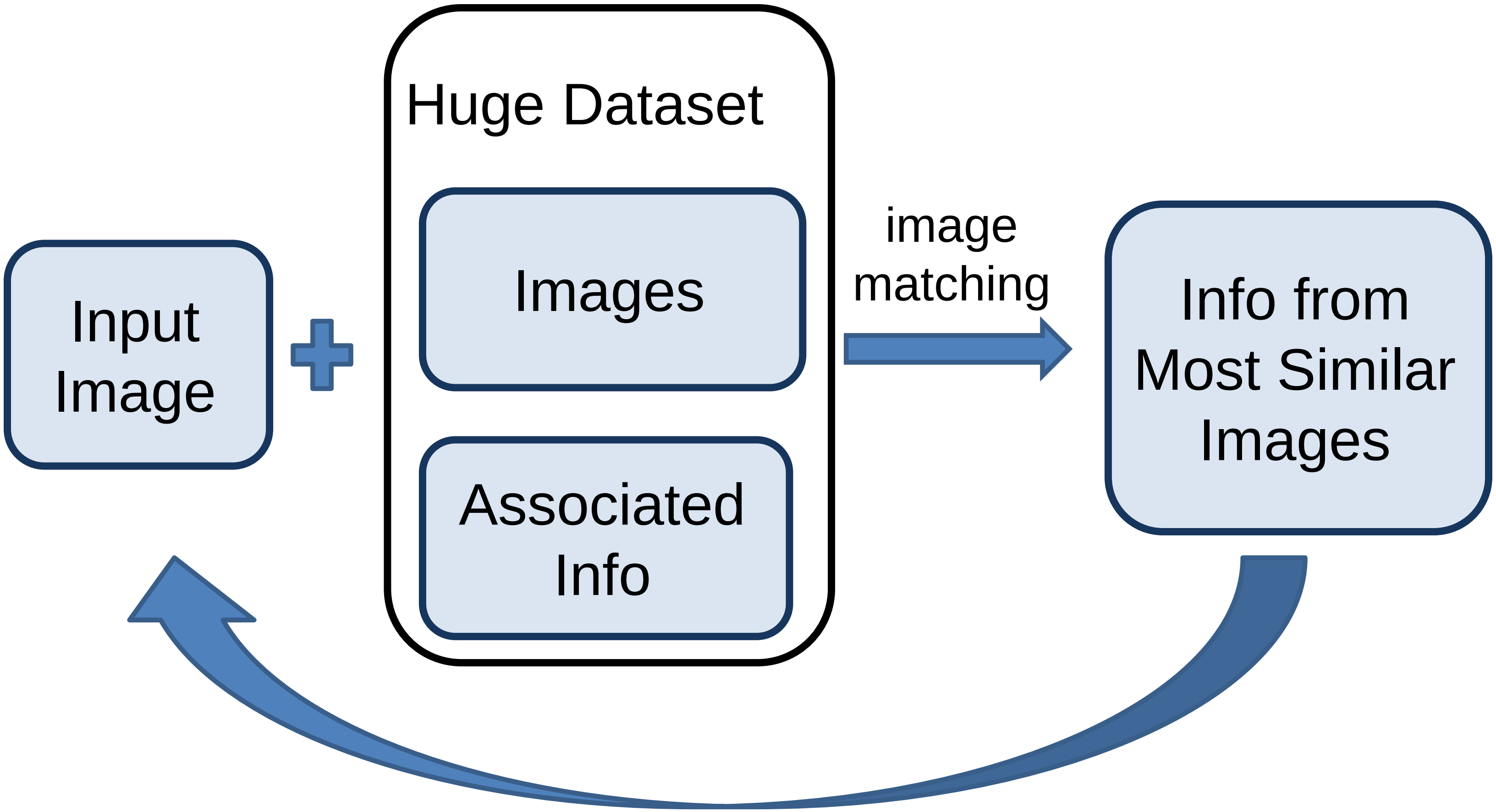




Gist scene descriptor
(Oliva and Torralba 2001)

Hays and Efros, SIGGRAPH 2007

Recap: Using lots of data!



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

Label Transfer

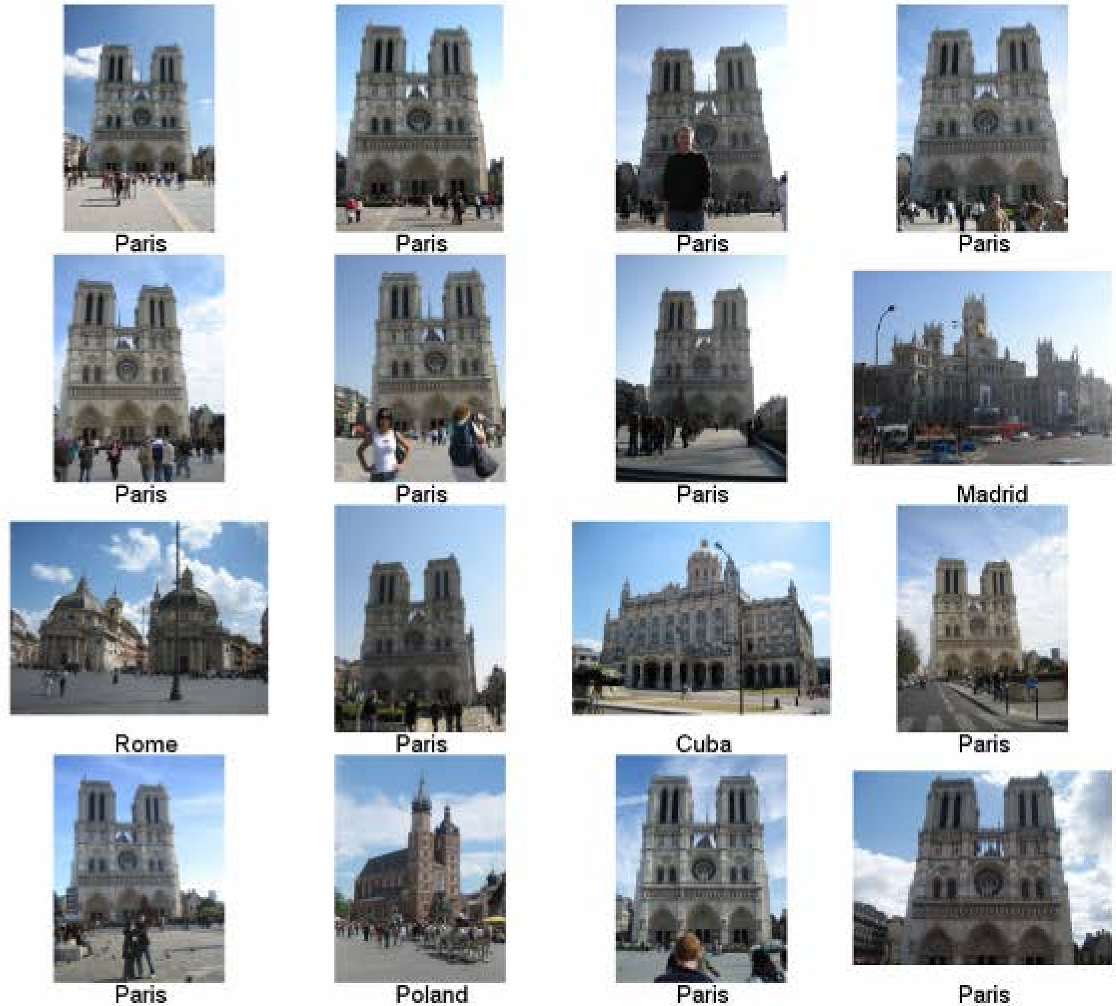
Tags: Sky, Water, Beach, Sunny, …
Time: 1pm, August, 2006, …
Location: Italy, Greece, Hawaii …
Photographer: Flickrbug21, Traveller2

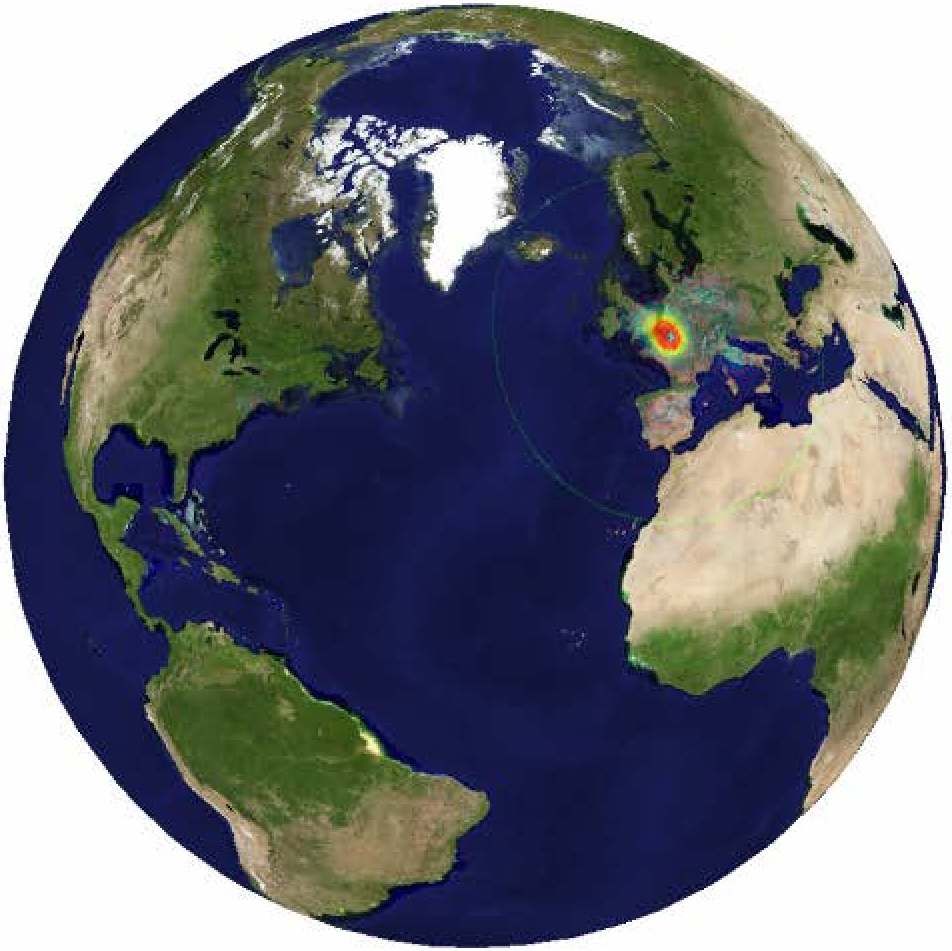
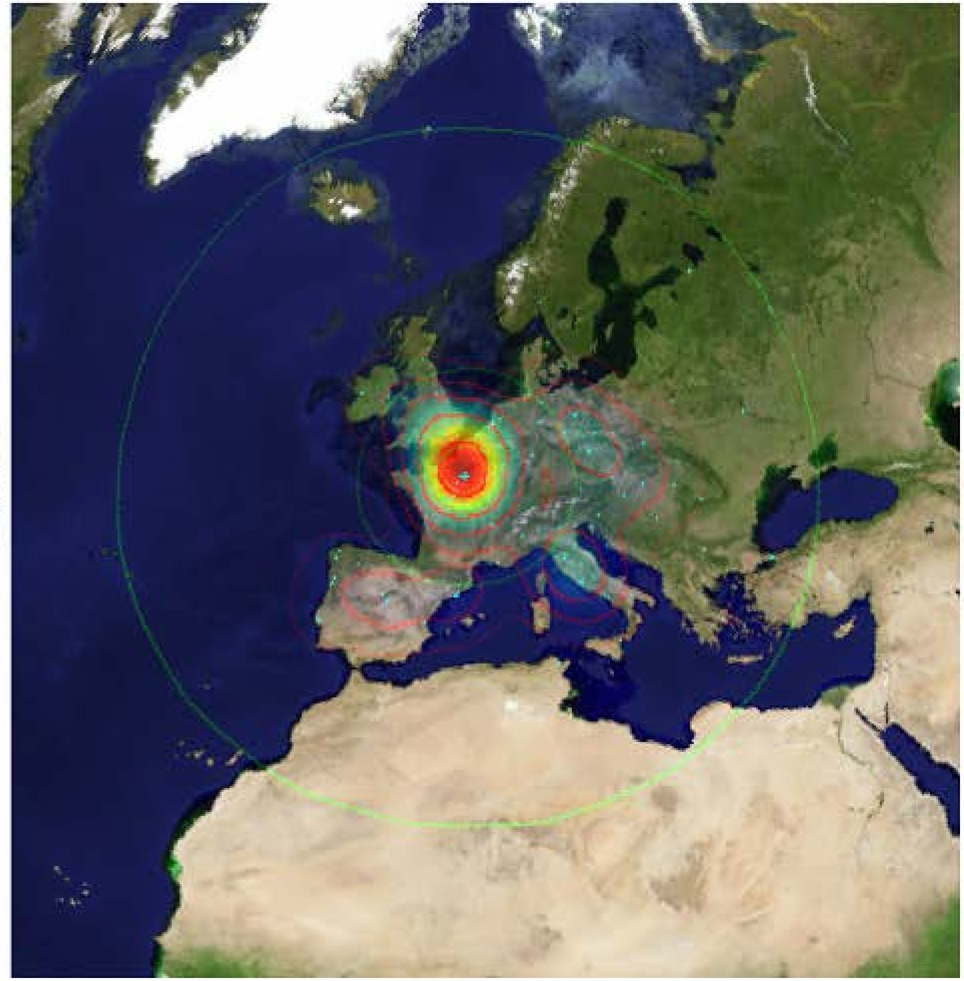
Label Transfer

im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

How much can an image tell about its geographic location?

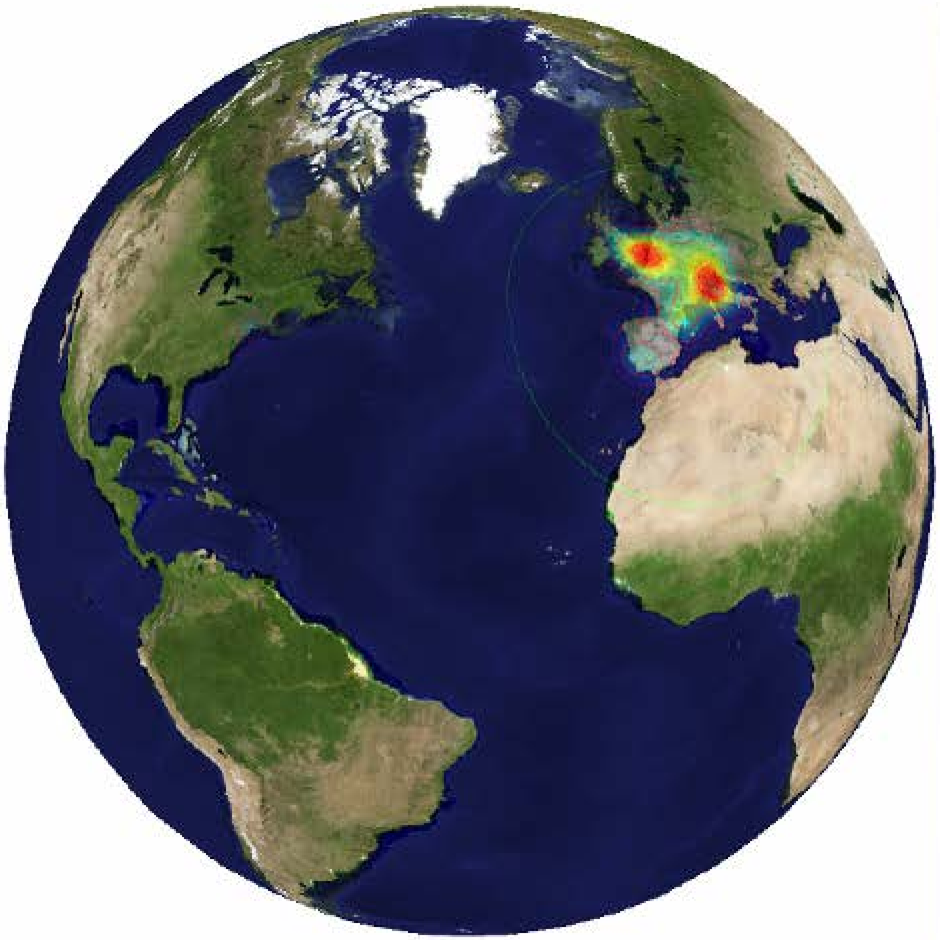

Paris
Paris
Paris
Paris
Paris
Paris
Paris
Madrid
Rome
Paris
Cuba
Paris
Paris
Poland
Paris
Paris

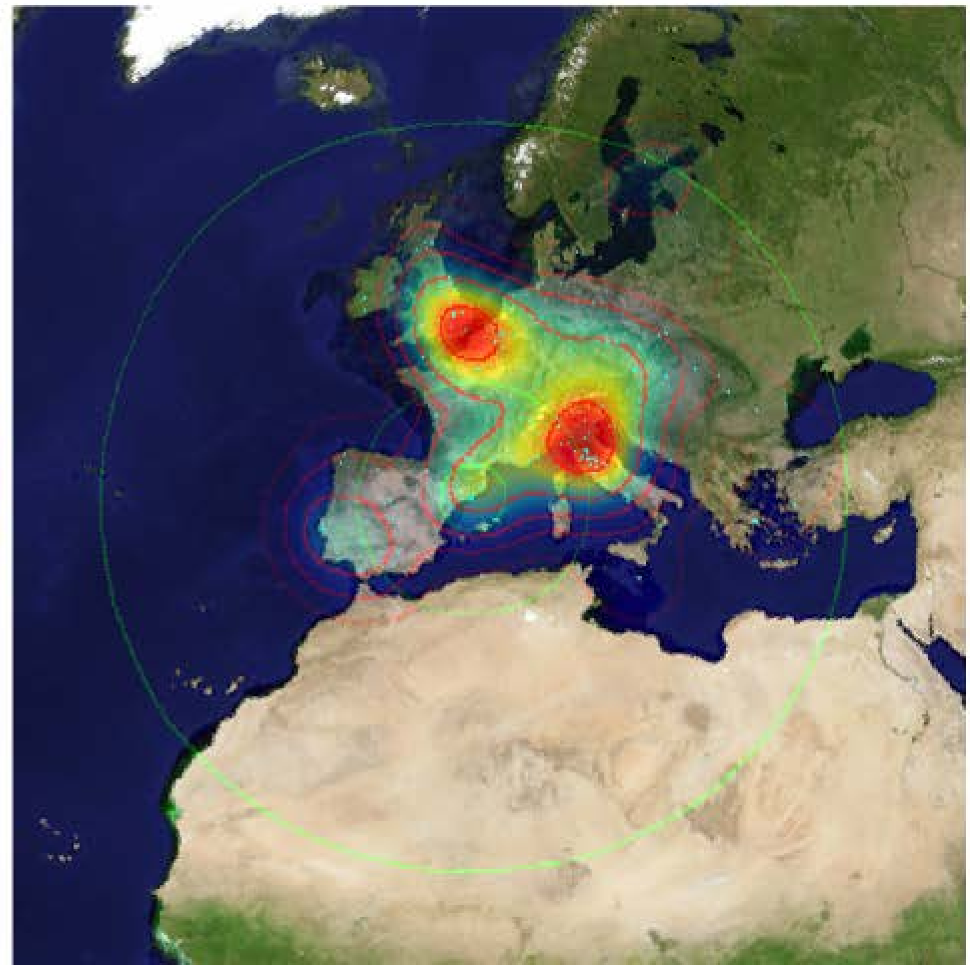



Im2gps

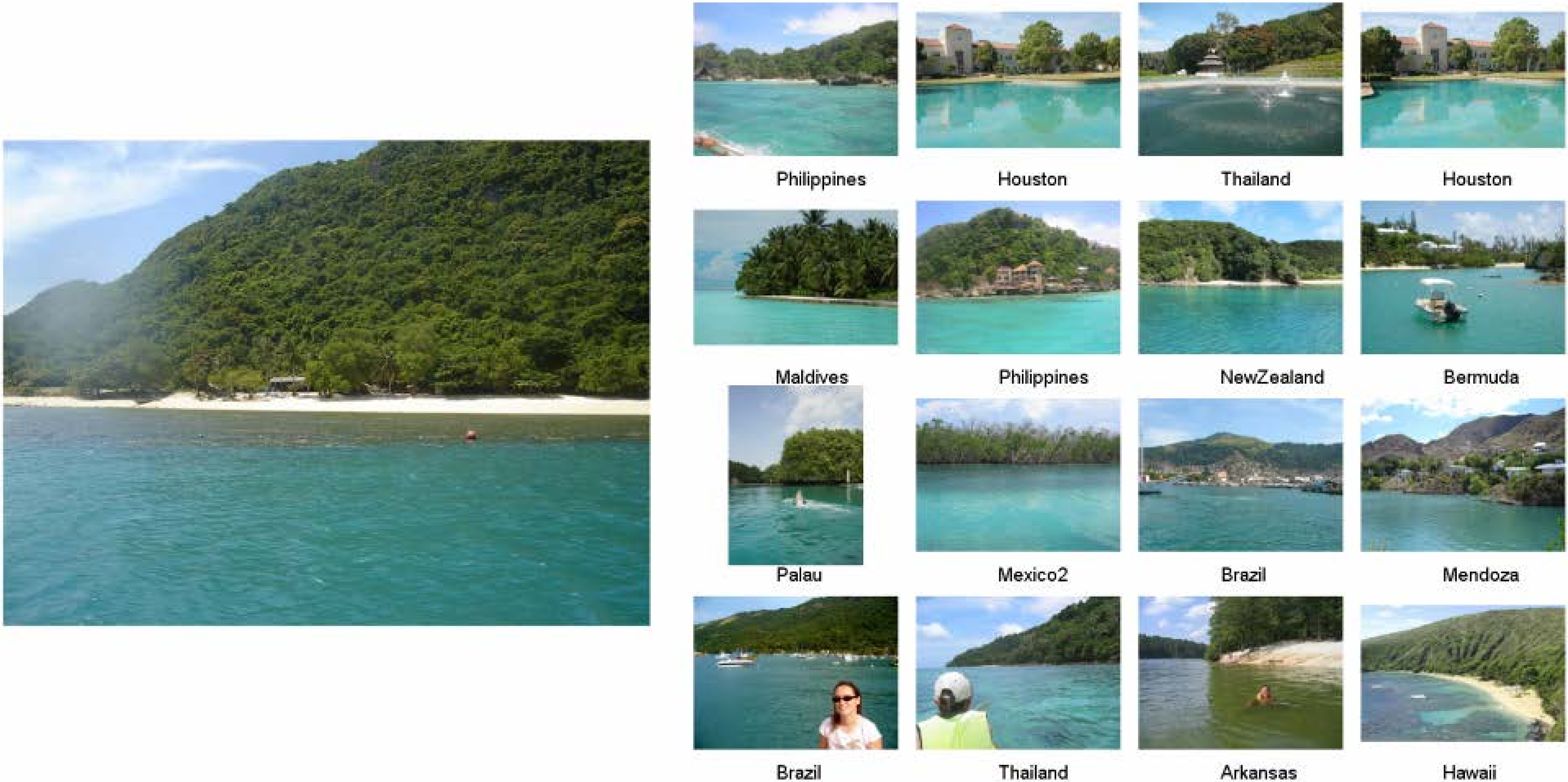
Example Scene Matches

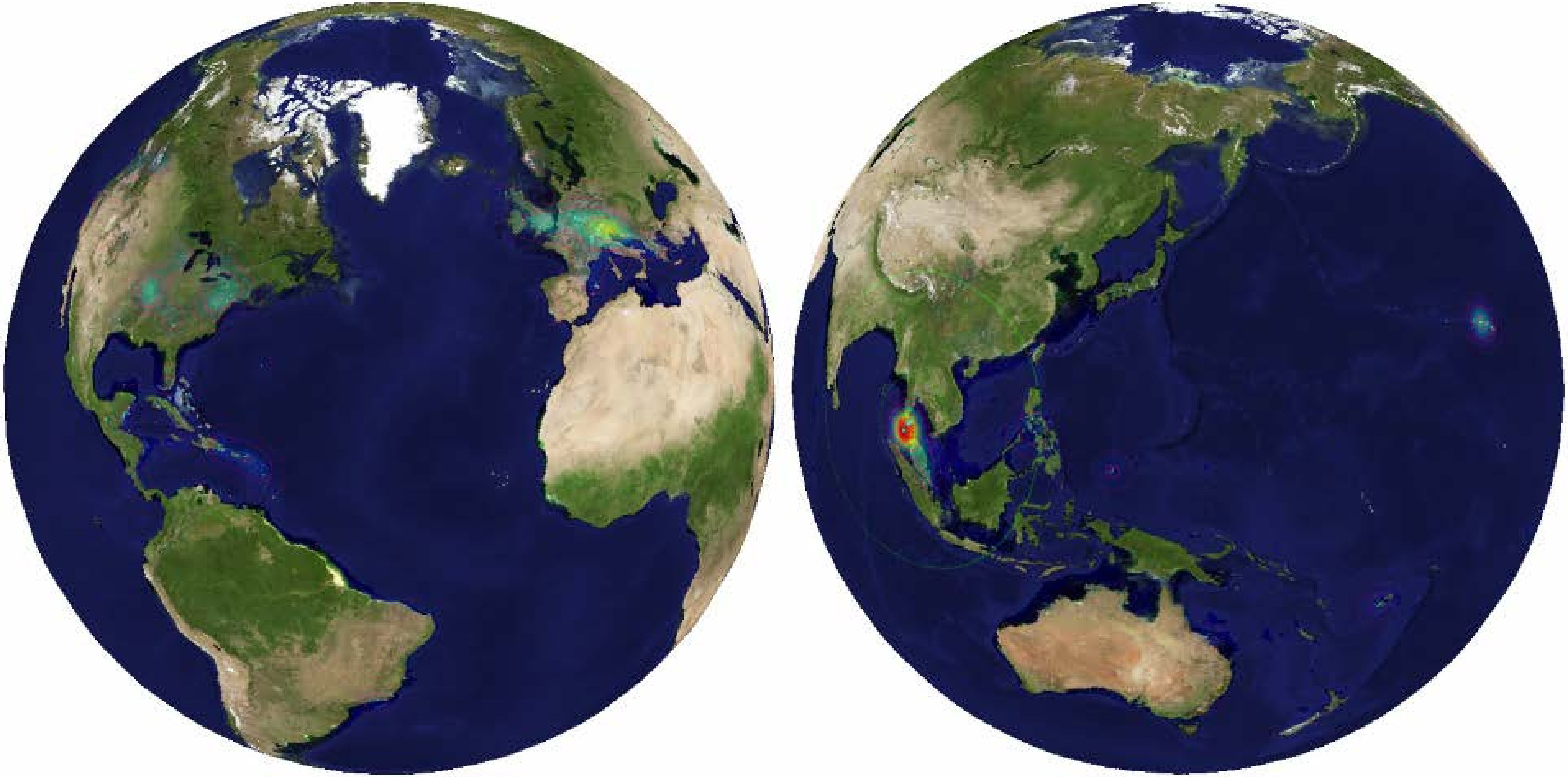
Voting Scheme

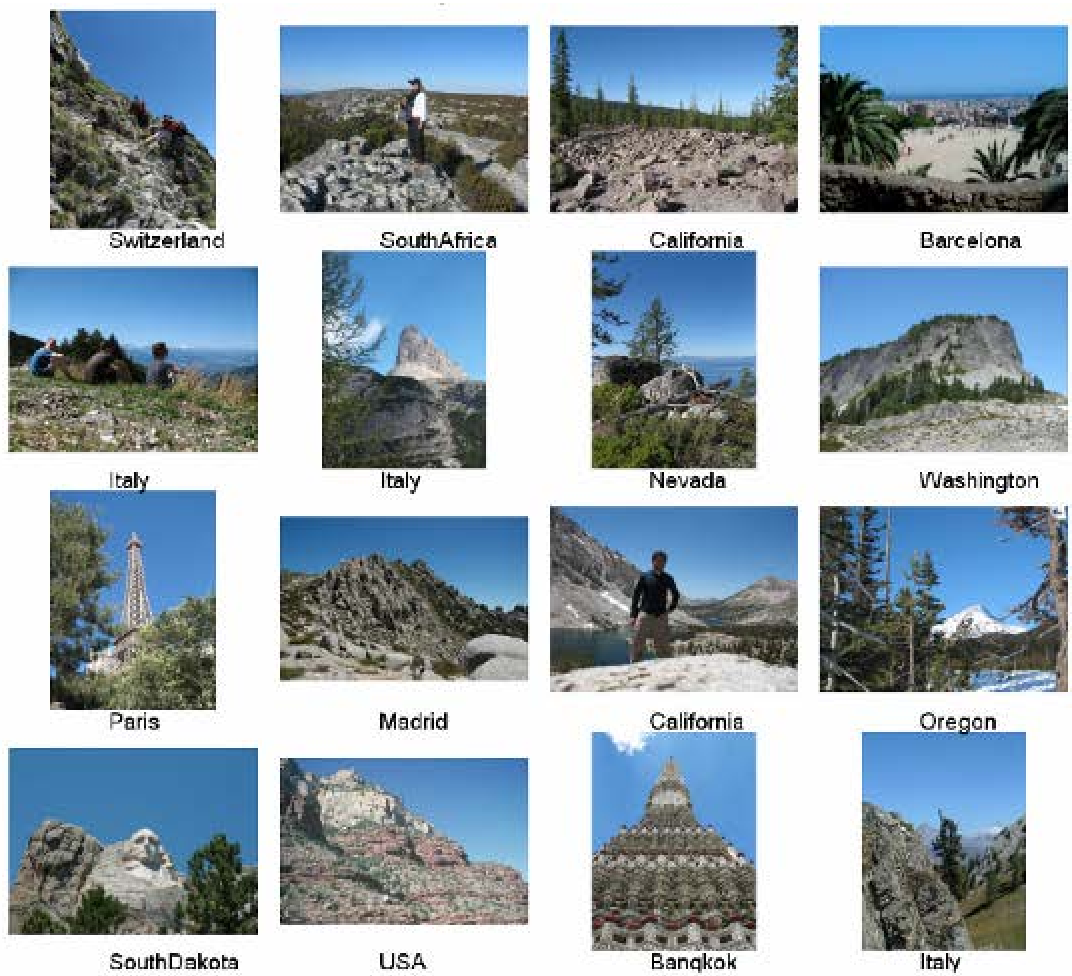


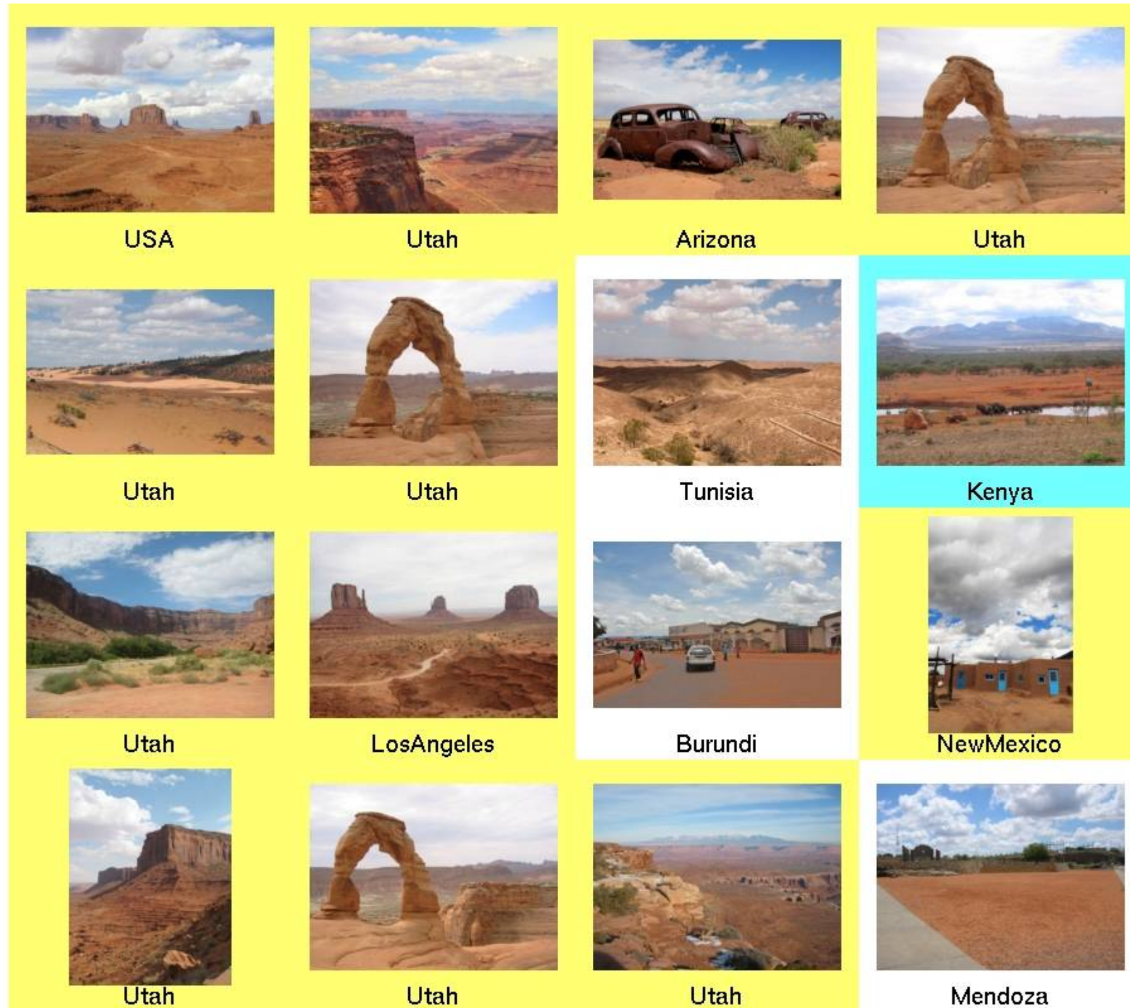


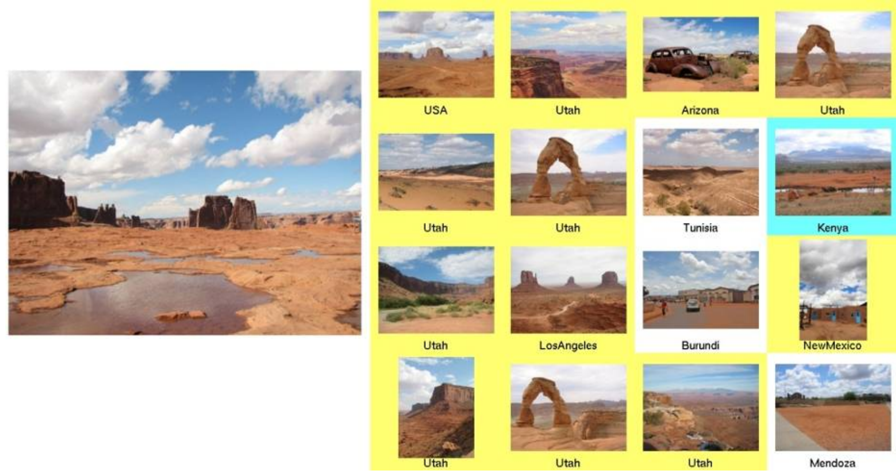
im2gps


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

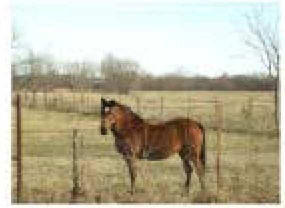



Switzerland
SouthAfrica
California
Barcelona
Italy
Italy
Nevada
Washington
Paris
Madrid
California
Oregon
SouthDakota
USA
Bangkok
Italy



USA
Utah
Arizona
Utah
Utah
Utah
Tunisia
Kenya
Utah
LosAngeles
Burundi
NewMexico
Utah
Utah
Utah
Mendoza


USA
Utah
Arizona
Utah
Utah
Utah
Tunisia
Kenya
Utah
LosAngeles
Burundi
NewMexico
Utah
Utah
Utah
Mendoza

California



Oklahoma

SouthAfrica

Zambia

Kenya

Hyderabad

Mongolia

SouthAfrica

Kenya

Kenya

Zambia

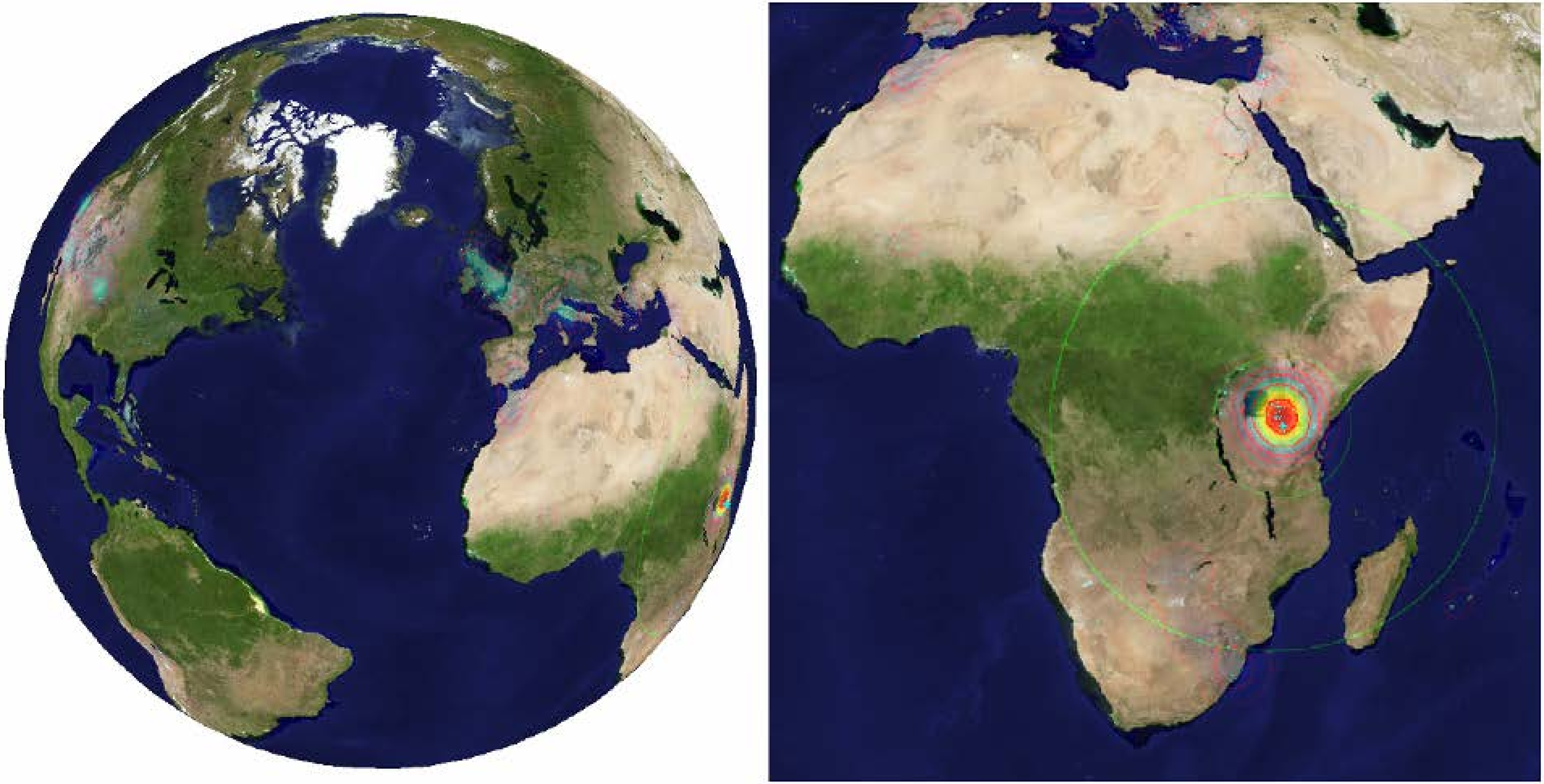
Ethiopia

Nevada

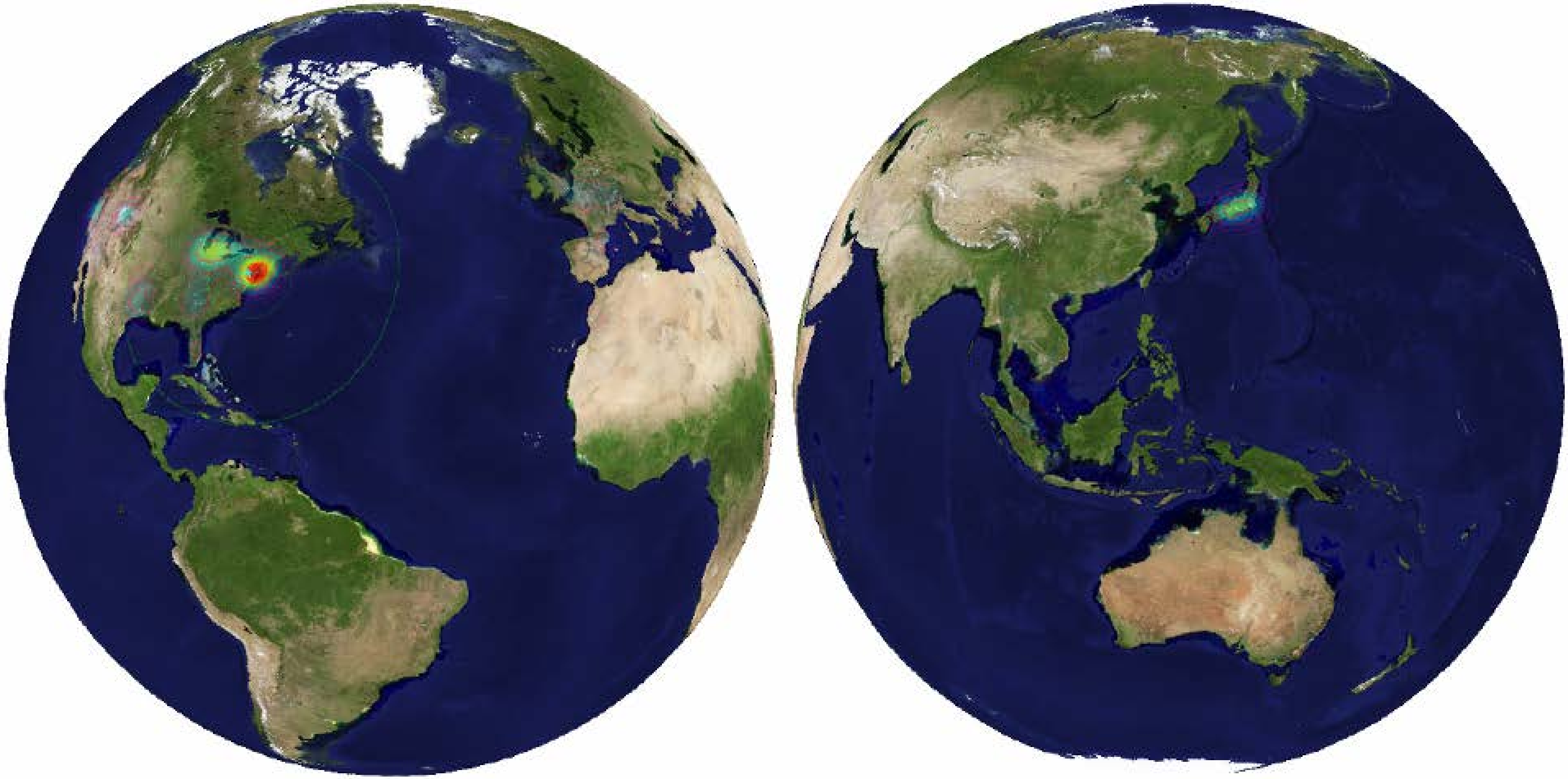
africa

Morocco

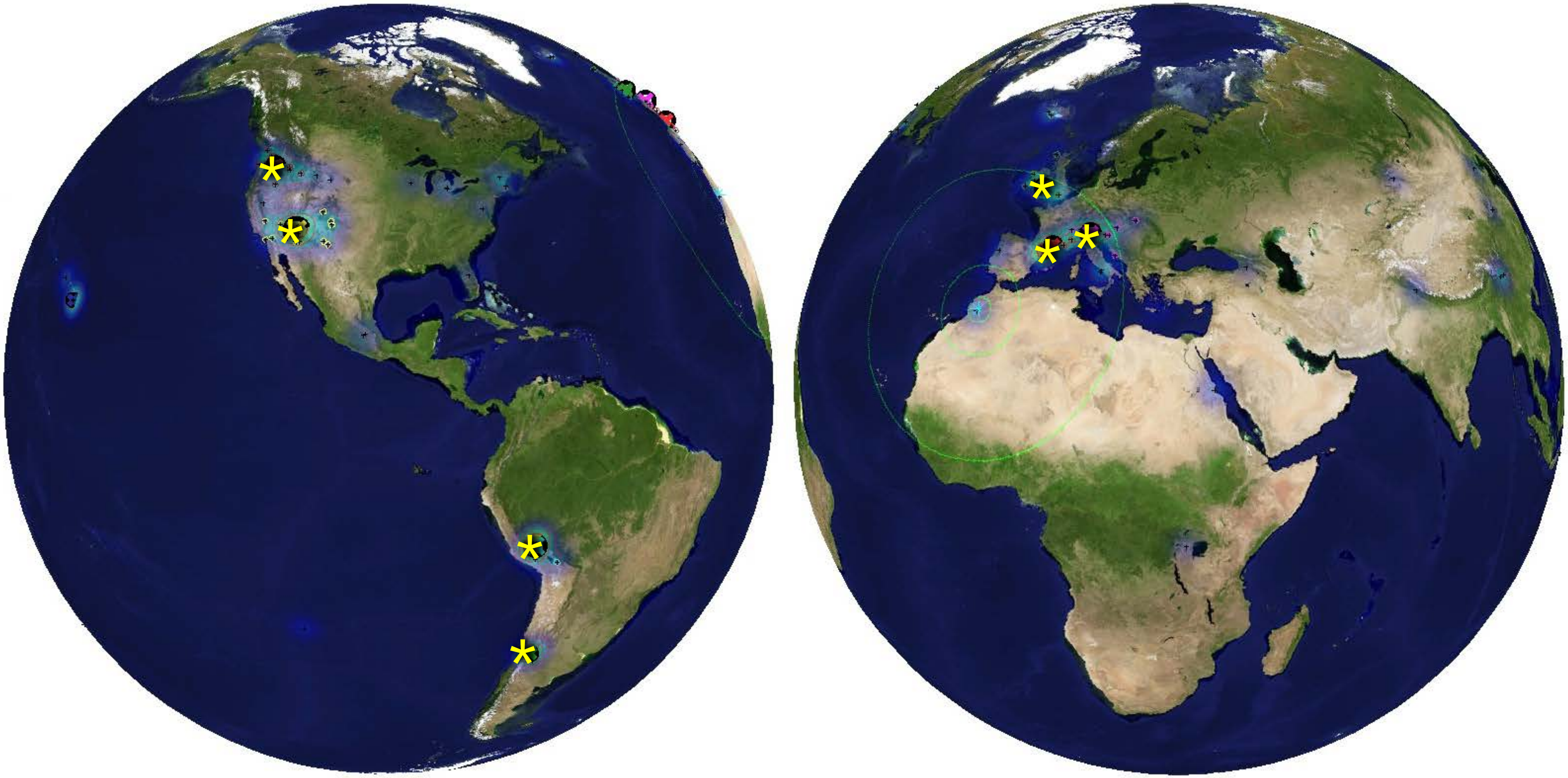
Tennessee



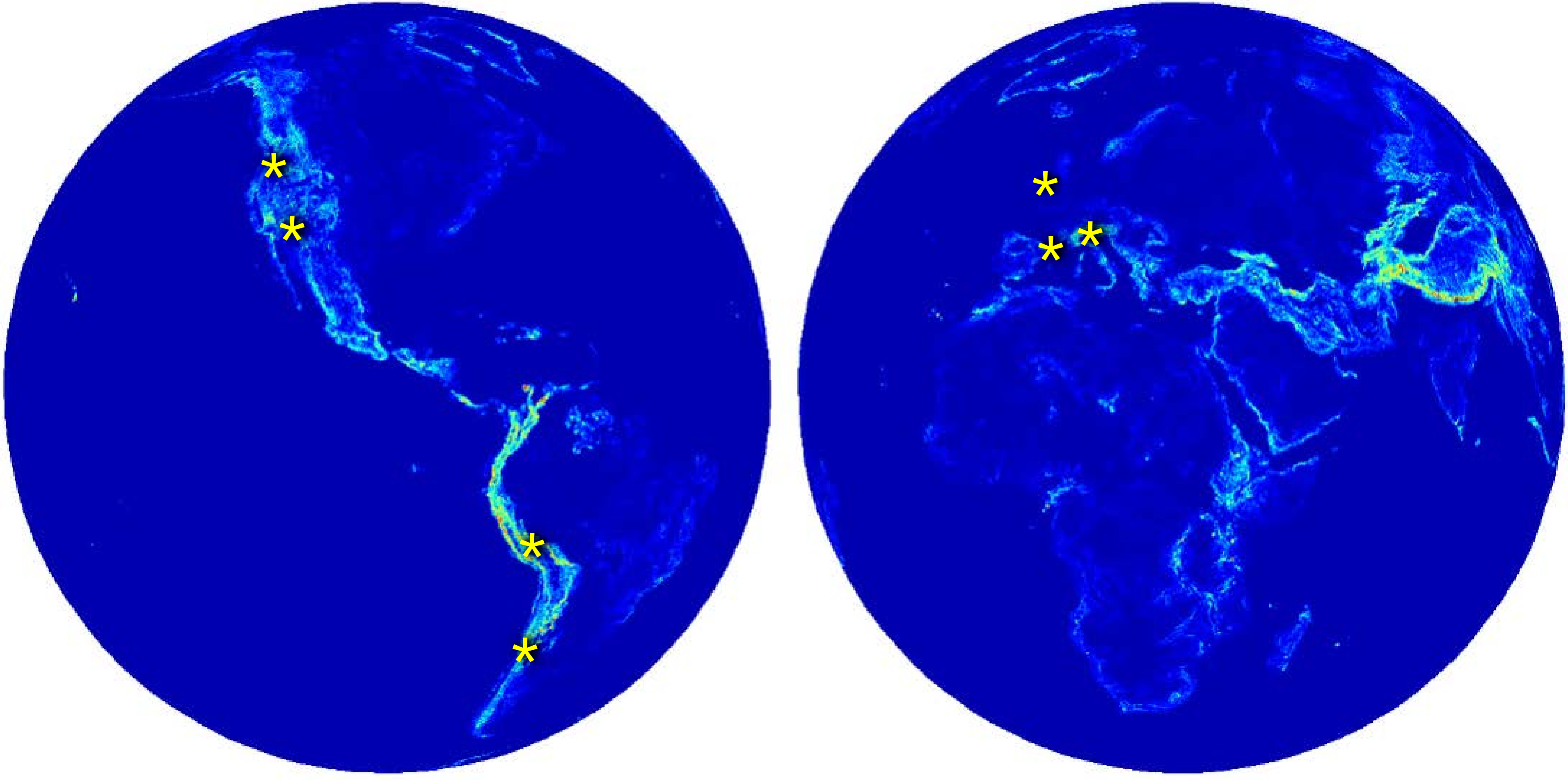
Toronto
Florida
NewYork
Boston
Boston
Oregon
Oregon
Oregon
NewYork
Barcelona
Oregon
Chicago
Ohio
Philadelphia
NewYorkCity
Boston



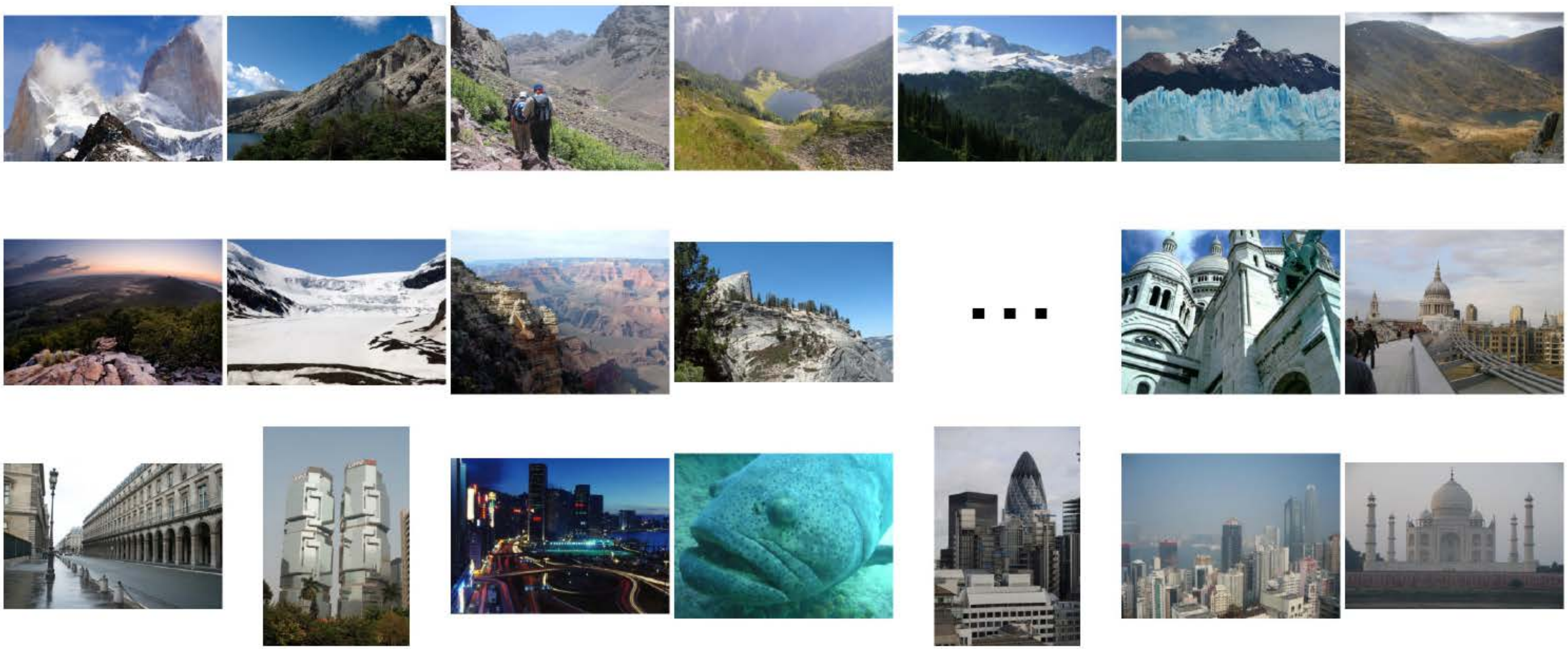
Data-driven categories



Elevation gradient =
112 m / km



Elevation gradient magnitude ranking



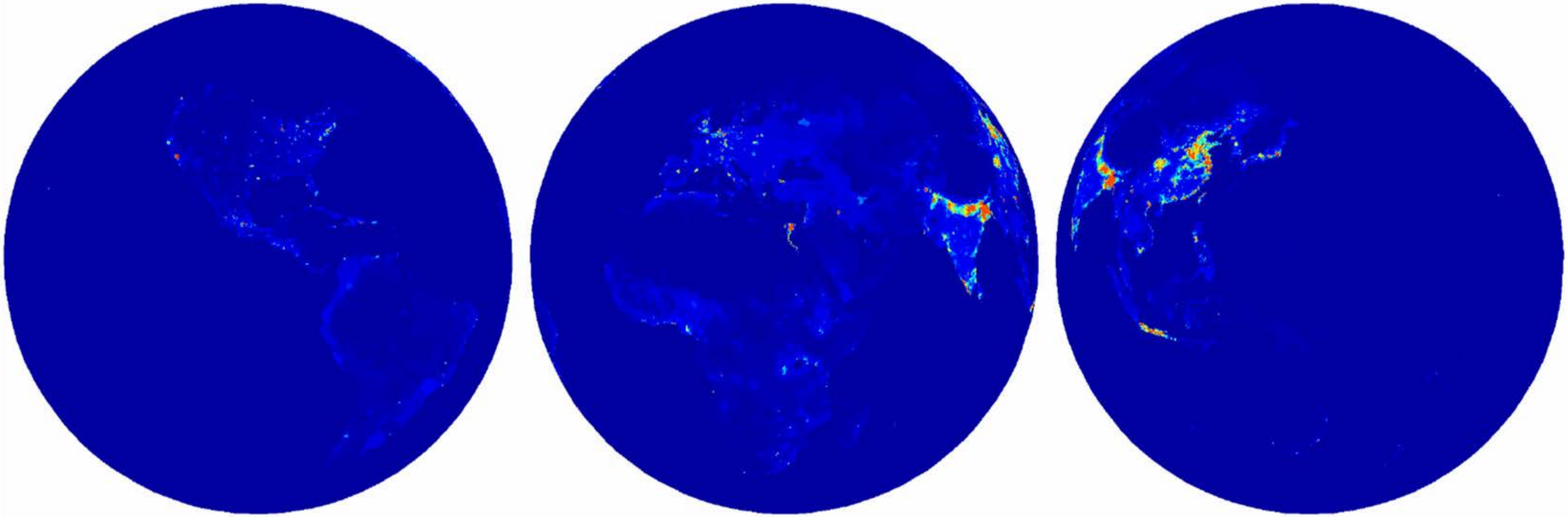
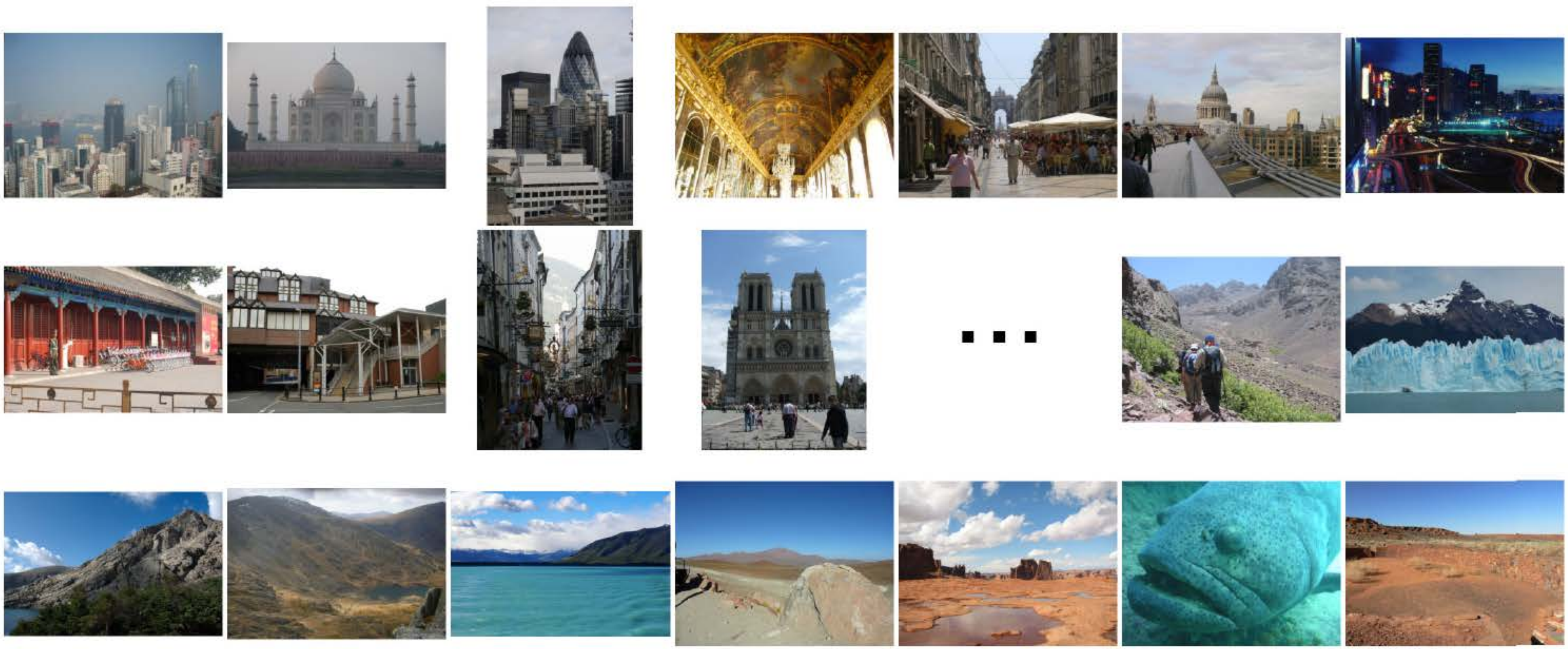


Figure 2. Global population density map.

Population density ranking



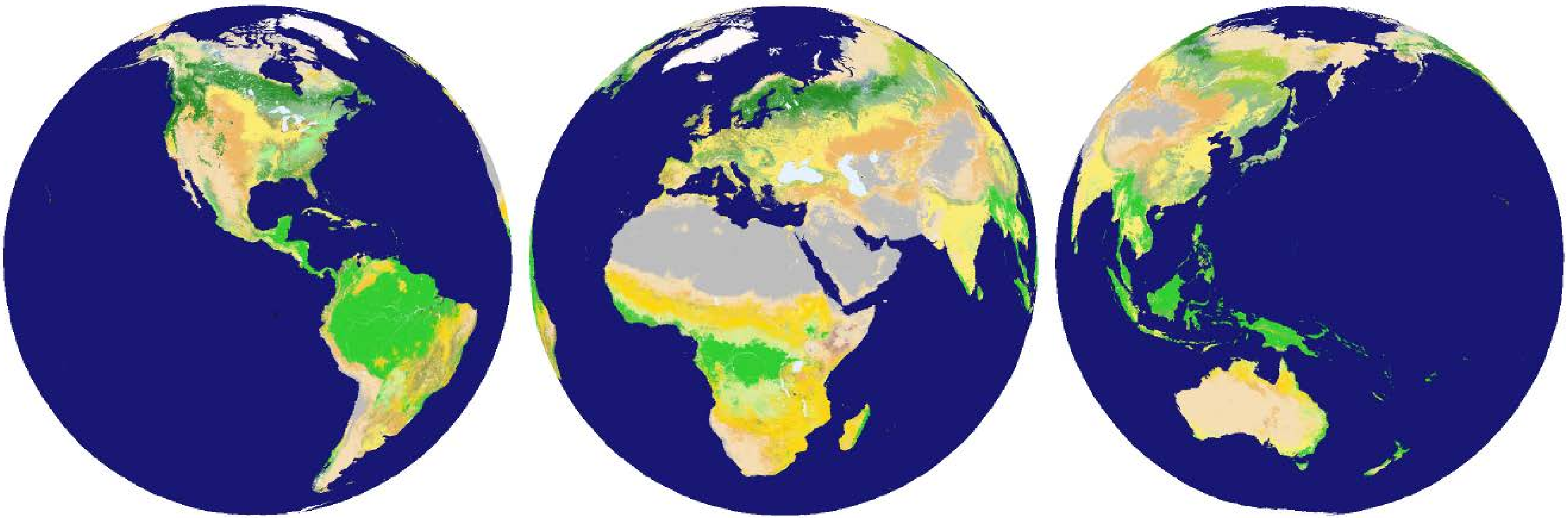
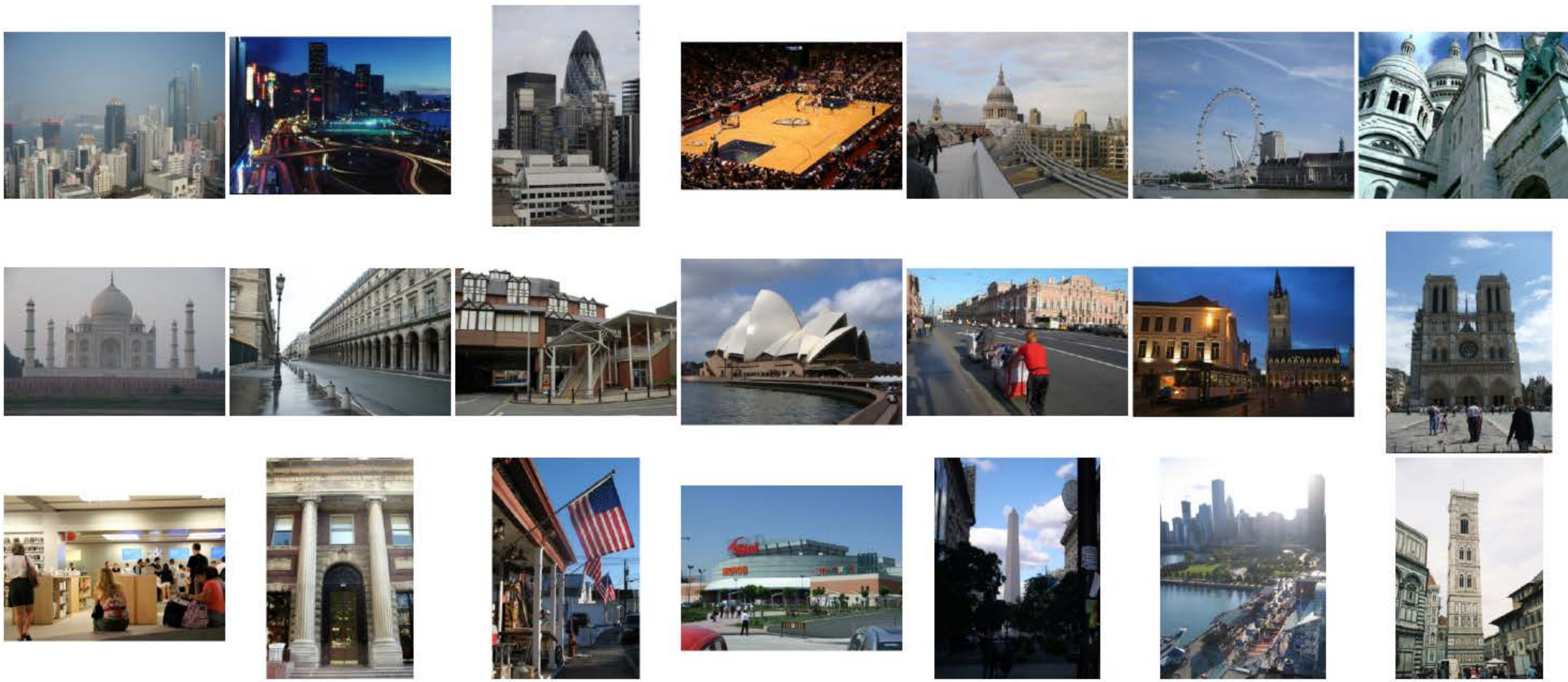


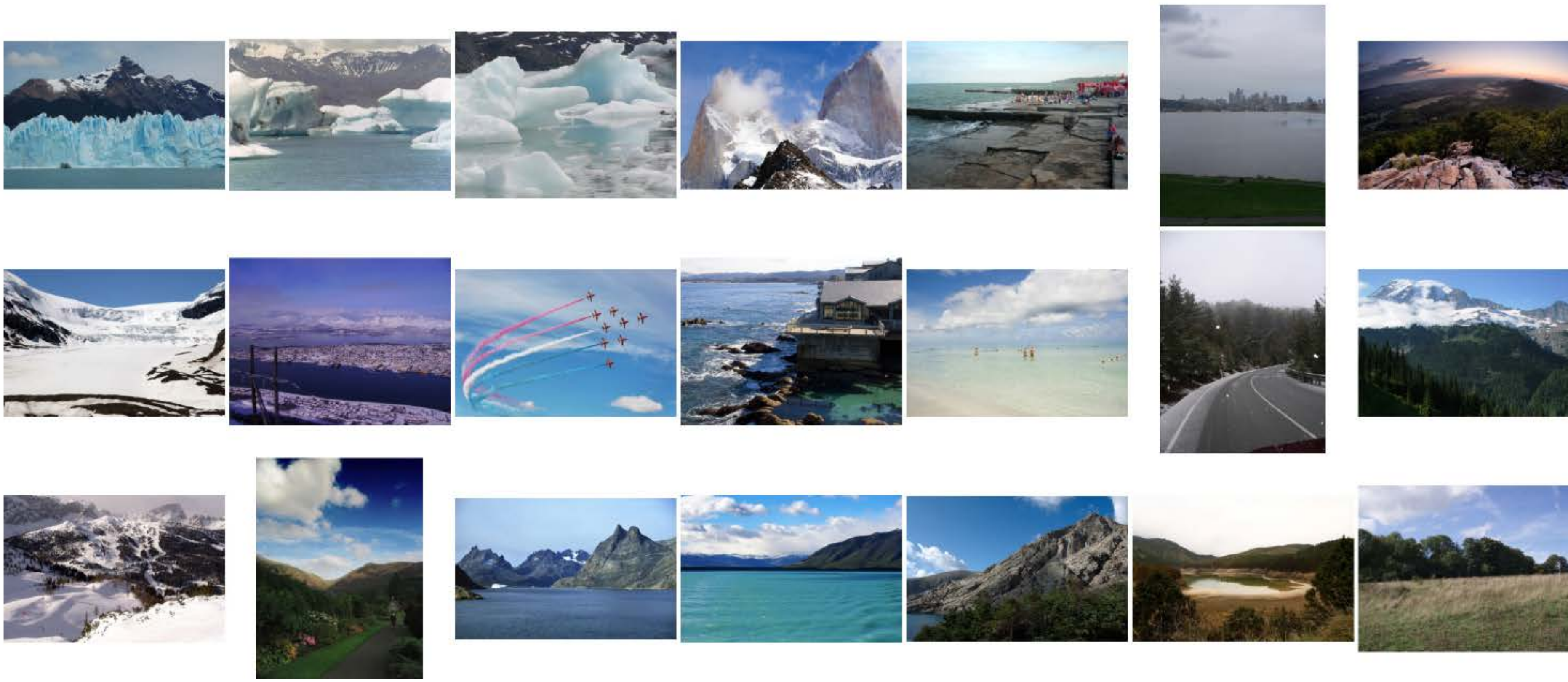
Figure 4. Global land cover classification map.

Barren or sparsely populated

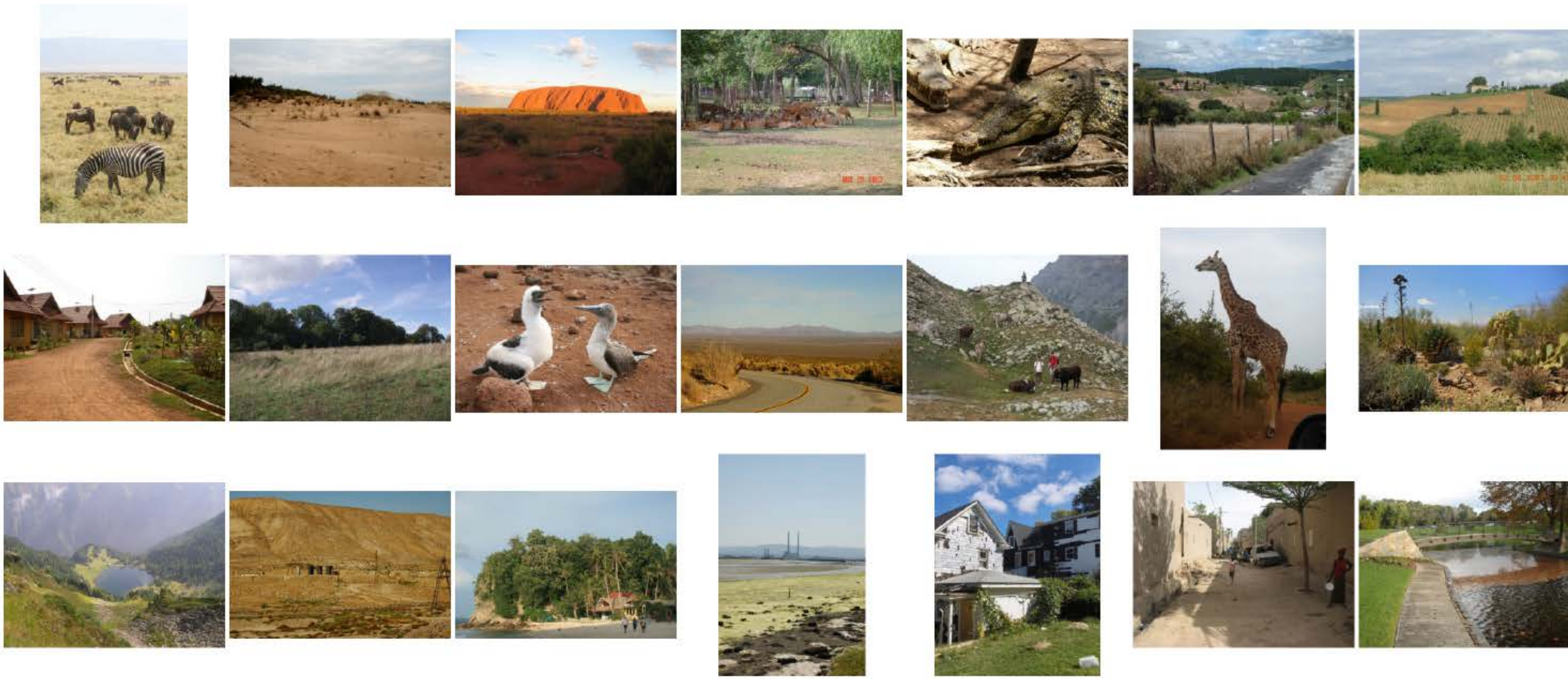
Urban and built up



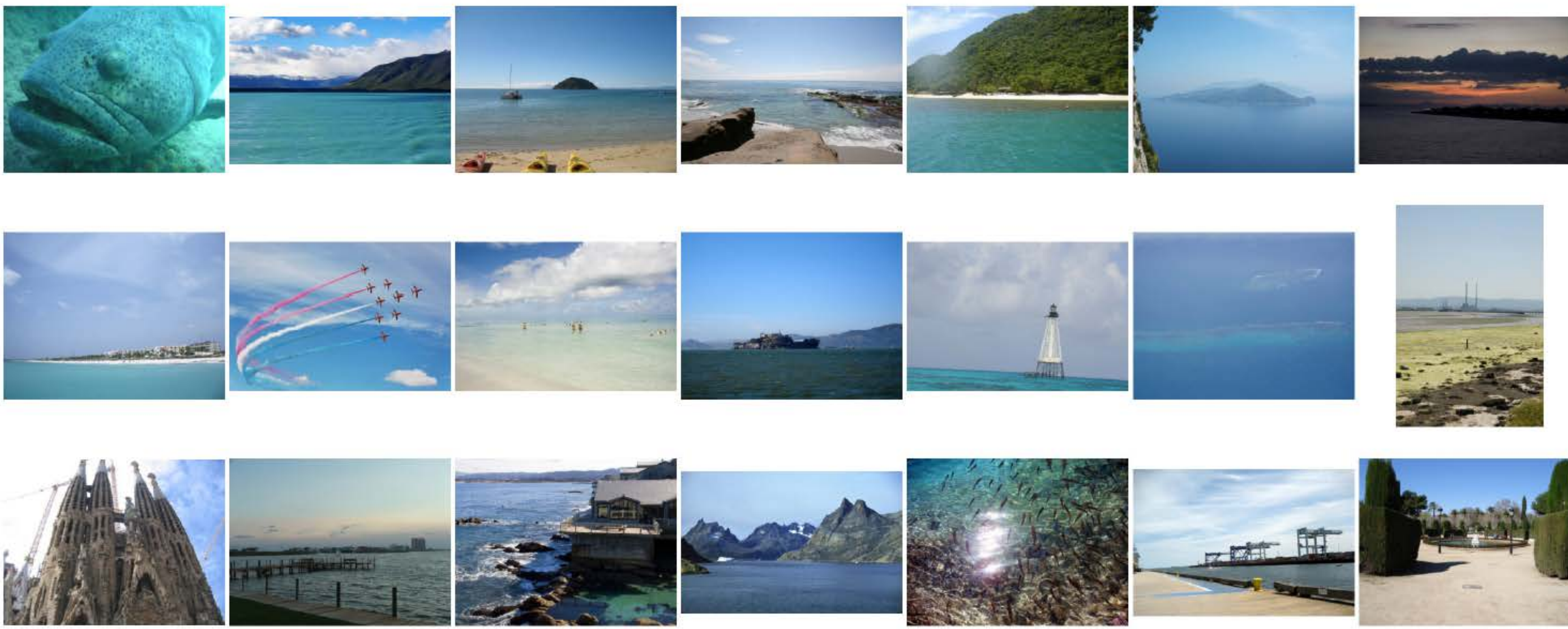
Snow and Ice



Savannah



Water



But surely the brain can’t remember this much!?

What's the Capacity of Visual Long Term Memory?

What we know…

Standing (1973)

10,000 images

83% Recognition

… people can remember thousands of images

What we don't know…

… what people are remembering for each item?

According to Standing

"Basically, my recollection is that we just separated the pictures into distinct thematic categories: e.g. cars, animals, single-person, 2-people, plants, etc.) Only a few slides were selected which fell into each category, and they were visually distinct."

Dogs
Playing Cards

"Gist" Only

Sparse Details

Highly Detailed

Slide by Aude Oliva

Massive Memory I: Methods

Showed 14 observers 2500 **categorically unique objects**

1 at a time, 3 seconds each

800 ms blank between items

Study session lasted about 5.5 hours

Repeat Detection task to maintain focus

Followed by 300 2-alternative forced choice tests

Slide by Aude Oliva

Massive Memory Experiment I

A stream of objects will be presented on the
screen for
~ 3 second each.

Your primary task:

Remember them ALL!

afterwards you will be tested with…

Completely different objects...

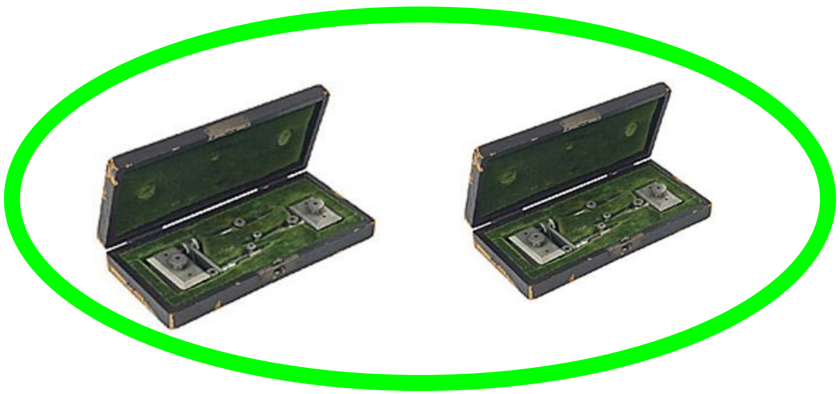
Different exemplars of the same kind of object...

Different states of the same object...

Massive Memory Experiment I

Your other task:

Detect exact repeats anywhere in the stream



Ready?

(Seriously, get ready to clap. The images go by fast…)

<clap!>

<clap!>

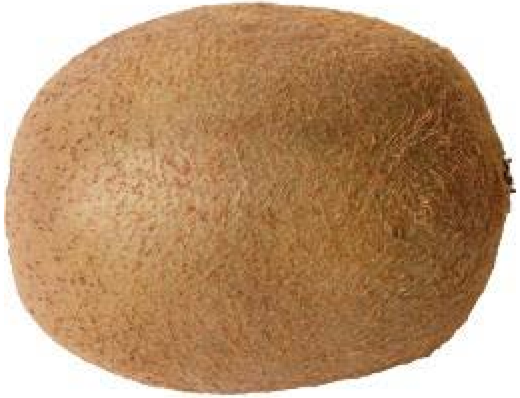
10 Minutes Later...

<clap!>

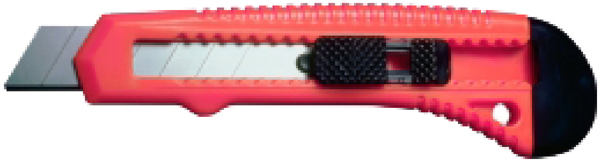
<clap!>

30 Minutes Later...

1 Hour Later...

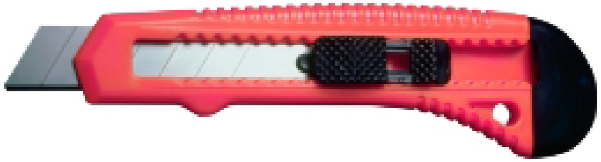



POLICE



<clap!>

2 Hours Later...



<clap!>

4 Hours Later...


PURE
MAPLE SYRUP
540 mL

<clap!>

5:30 Hours Later...

Which one did you see?

(go ahead and shout out your answer)

-A-

-B-

-A-

-B-

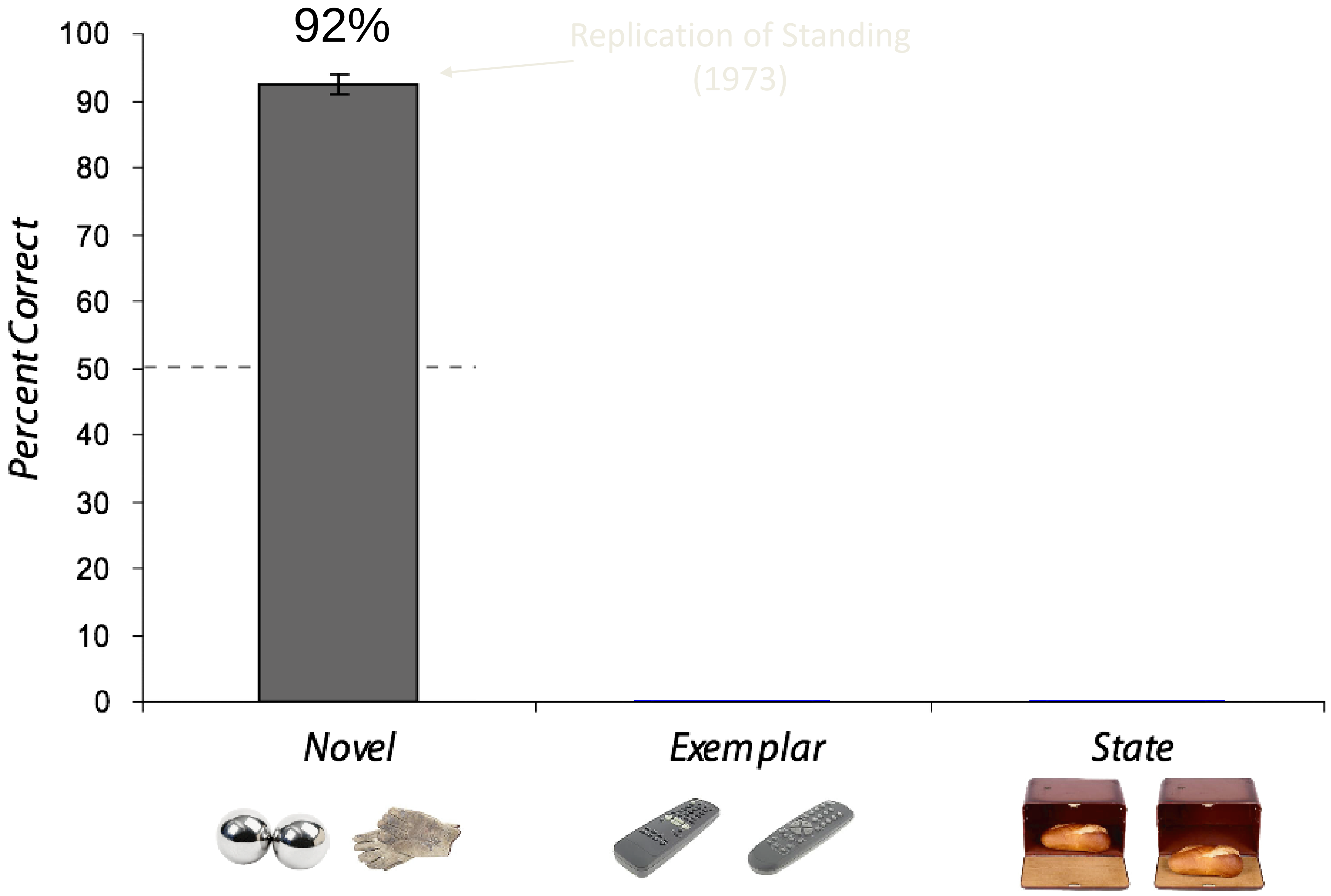
-A-

-B-

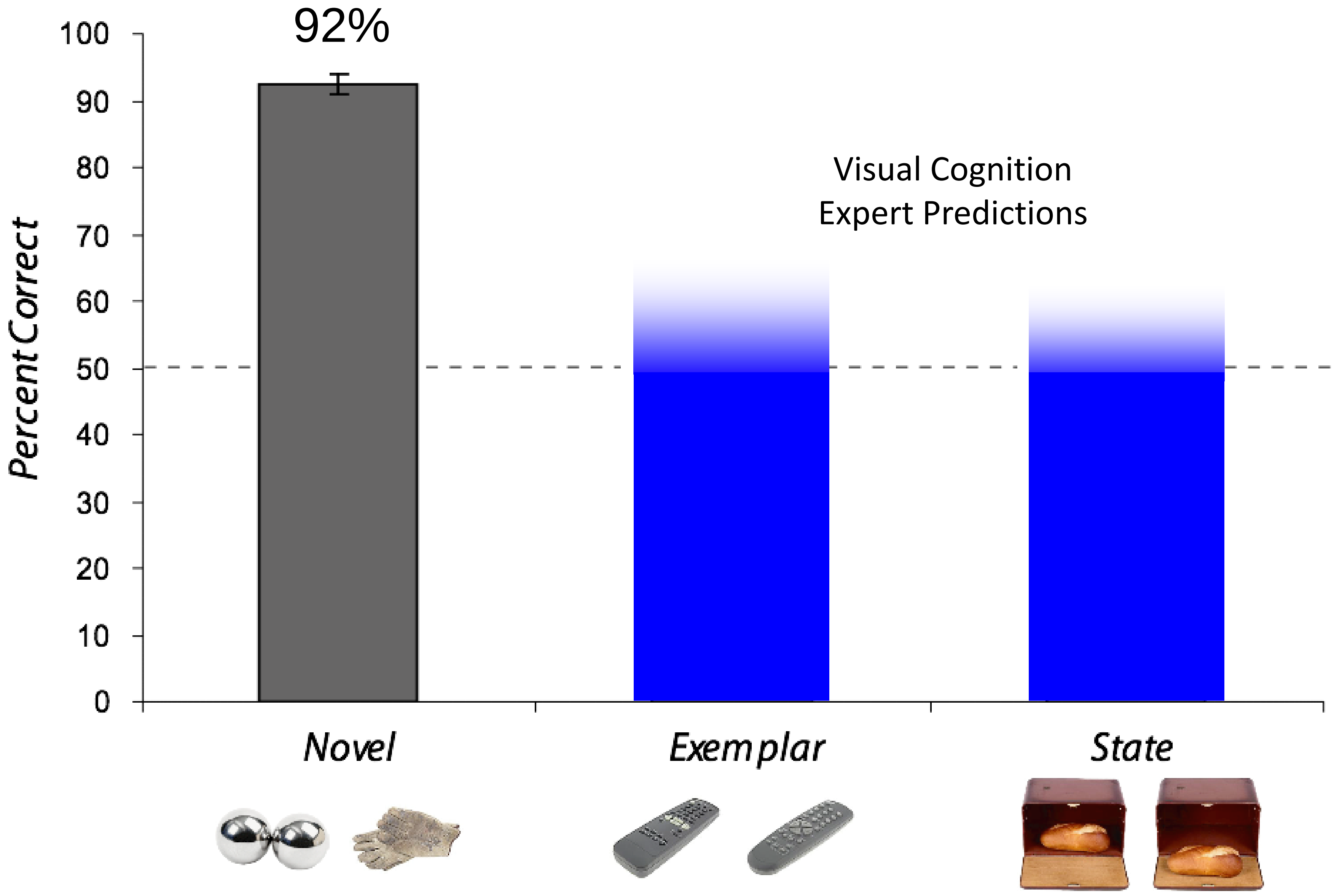
Examples of **State** memory test

All stimuli available at: cvcl.mit.edu/MM

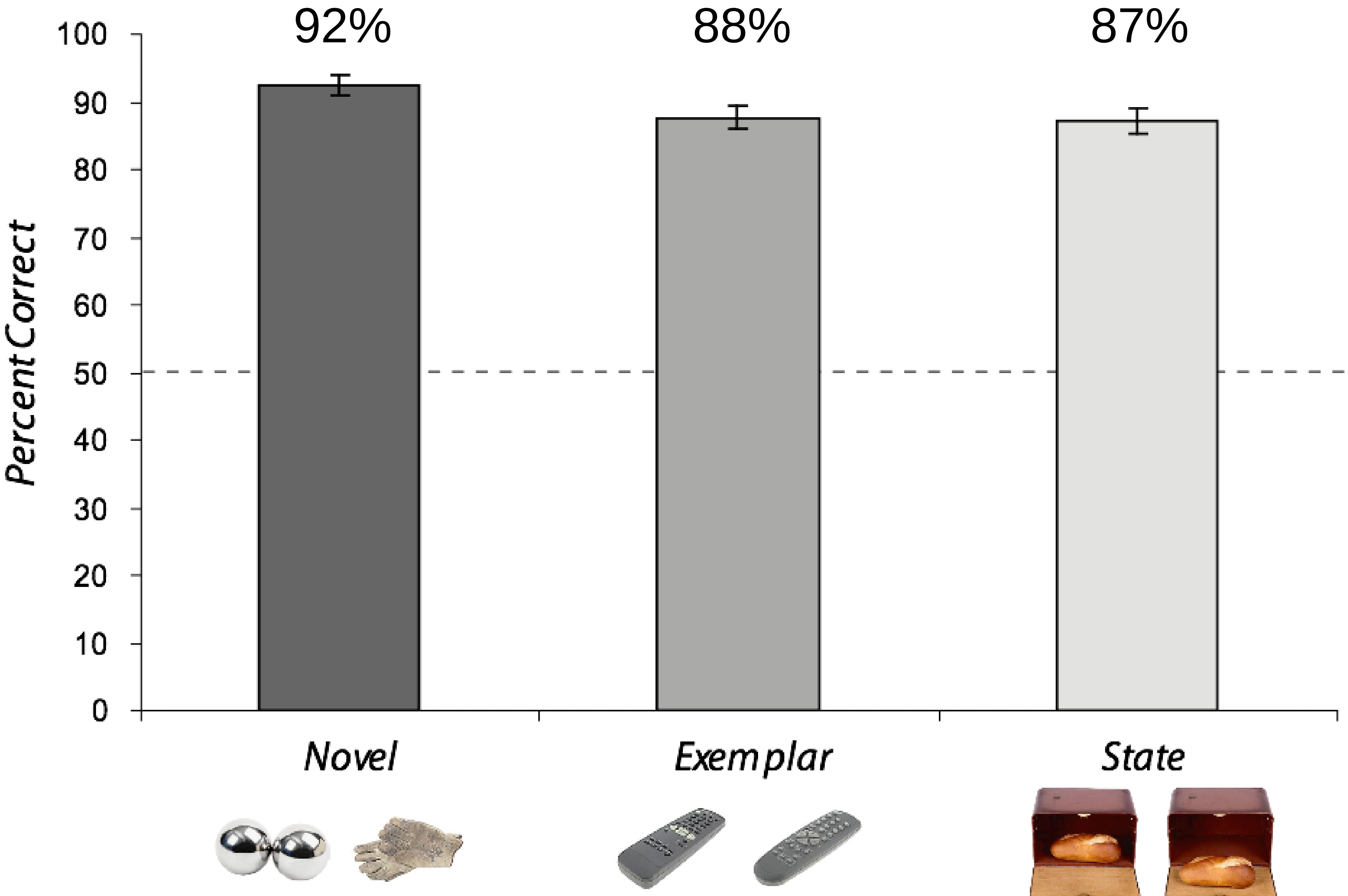
Recognition Memory Results



Recognition Memory Results



Recognition Memory Results



Brady, et al. (2008), *PNAS*

Using Data for Image Creation…

Michel Gondry, *Je Danse la Mia*

https://www.youtube.com/watch?v=7ceNf9qJjgc

Scene matching with camera transformations

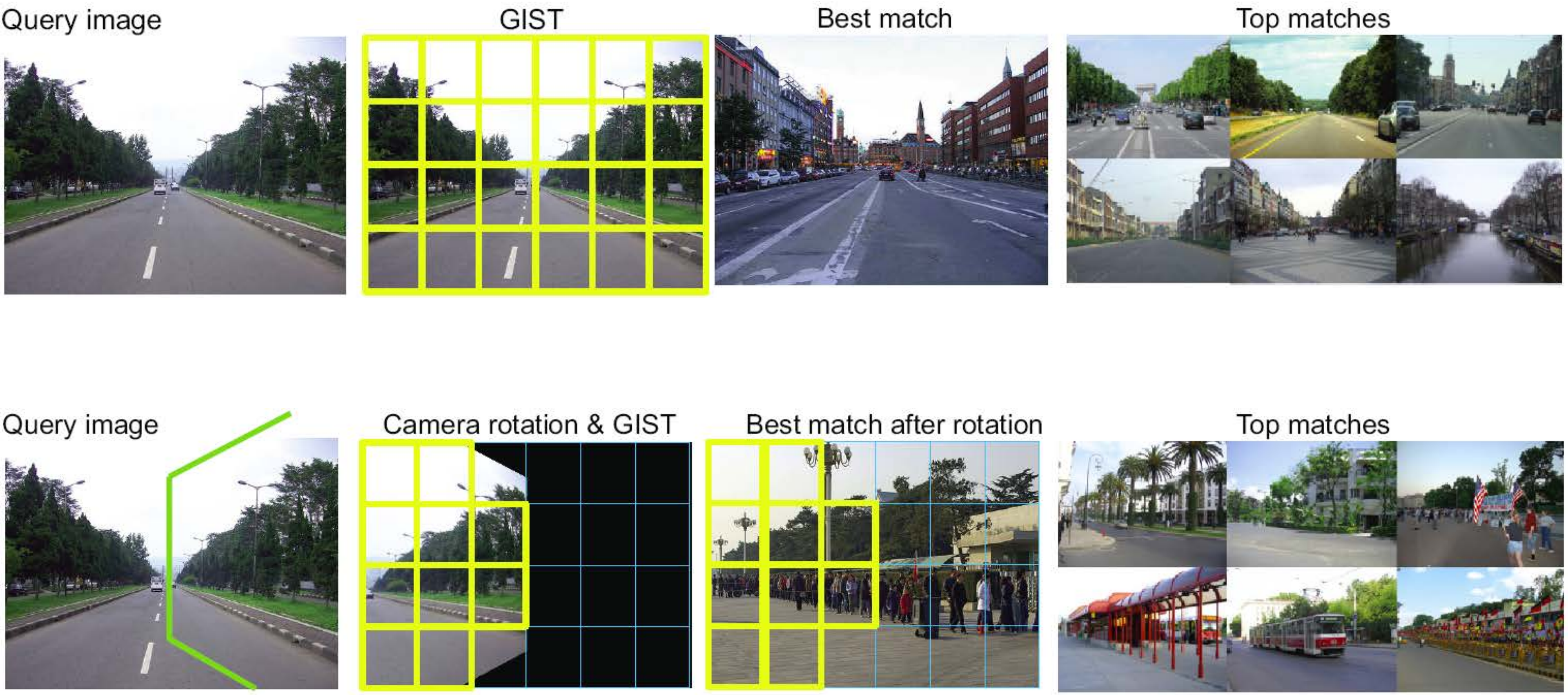
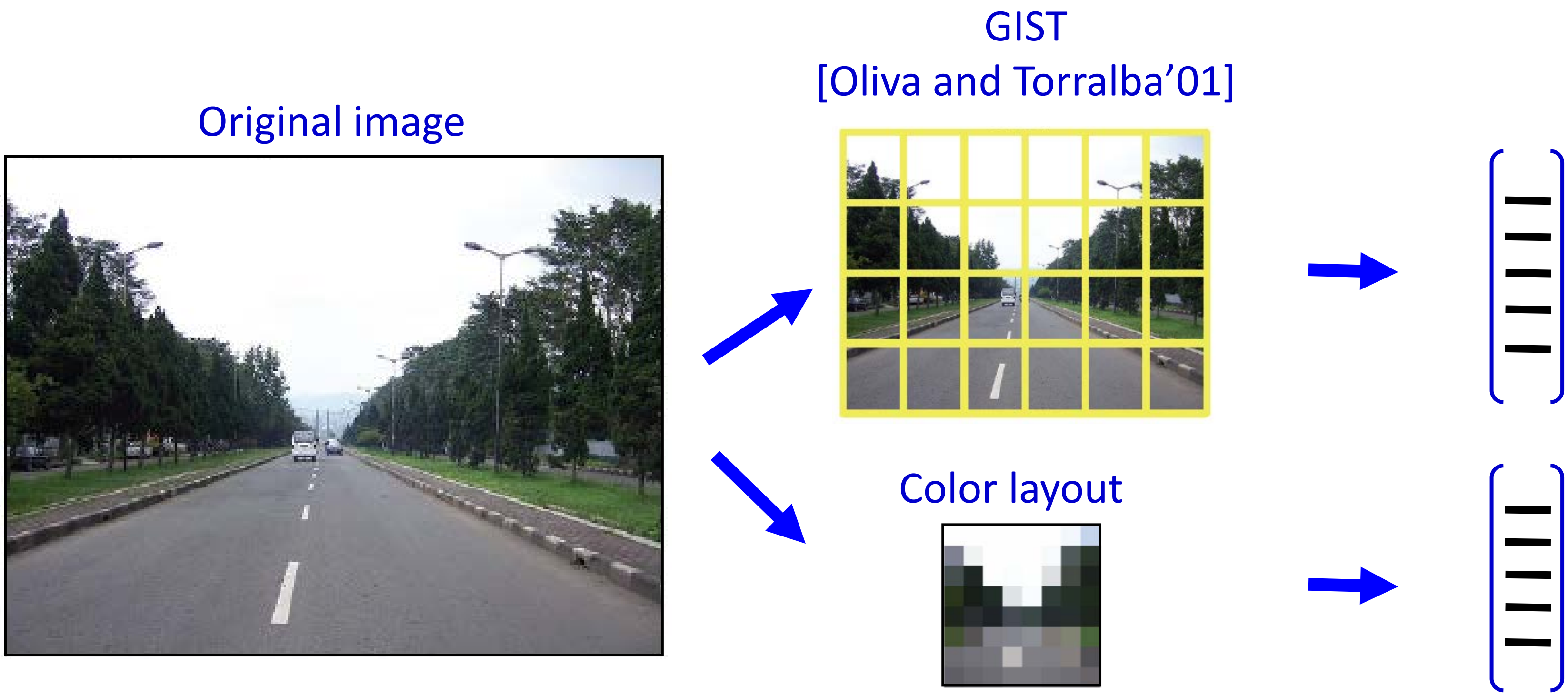
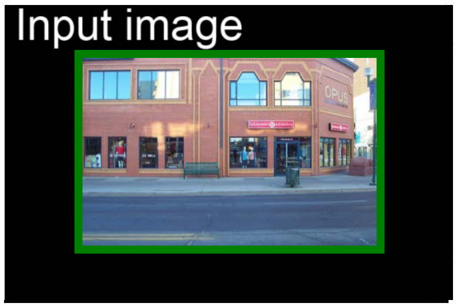


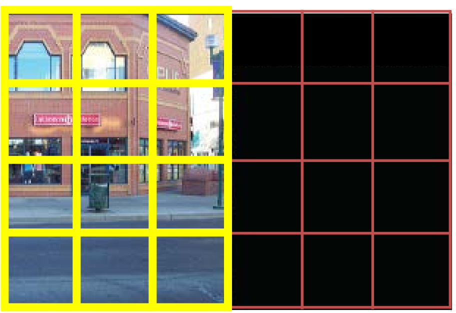
Image representation



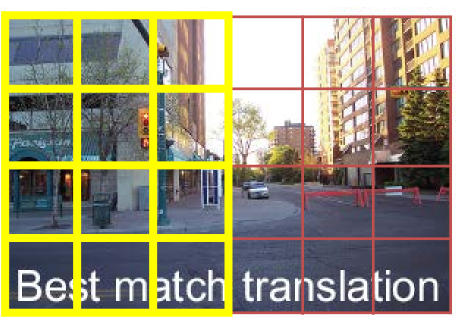
Scene matching with camera view transformations: Translation



1. Move camera



2. View from the virtual camera



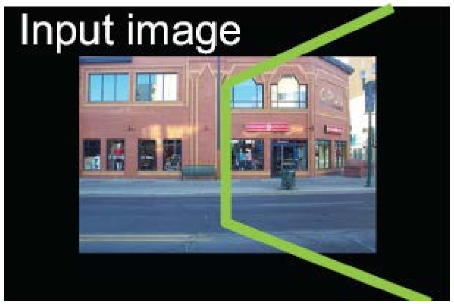
3. Find a match to fill the missing pixels

4. Locally align images

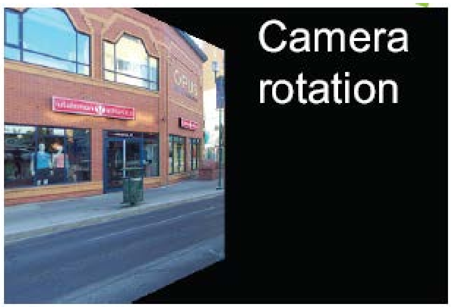
5. Find a seam

6. Blend in the gradient domain

Scene matching with camera view transformations: Camera rotation

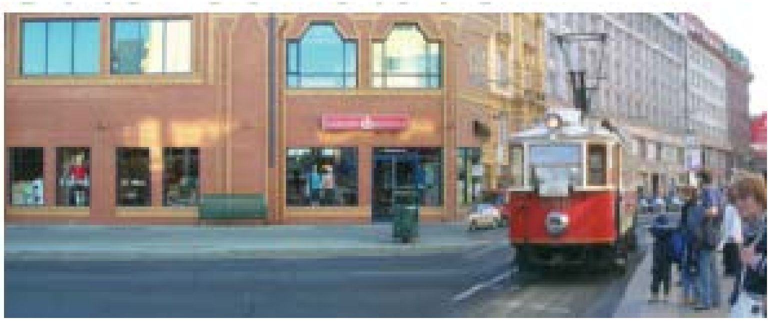


1. Rotate camera



2. View from the virtual camera

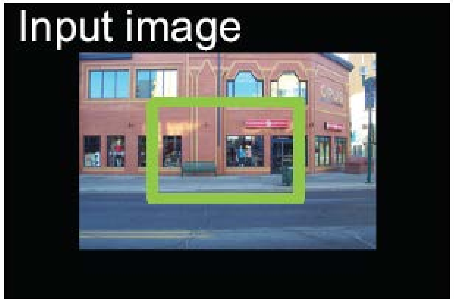
3. Find a match to fill-in the missing pixels



4. Stitched rotation

5. Display on a cylinder

Scene matching with camera view transformations: Forward motion



1. Move camera

2. View from the virtual camera

3. Find a match to replace pixels

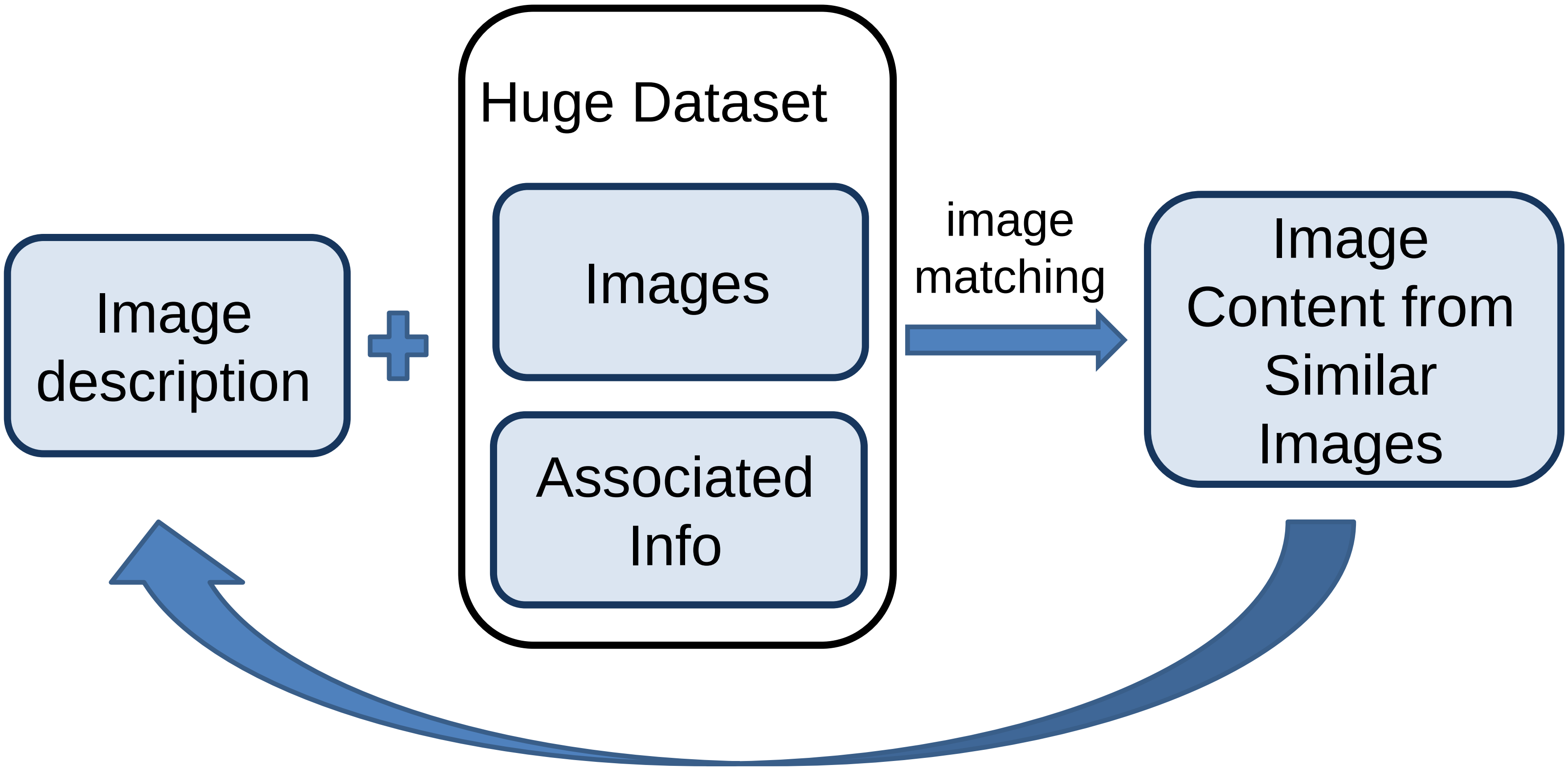
Tour from a single image

Navigate the virtual space using intuitive motion controls

Video

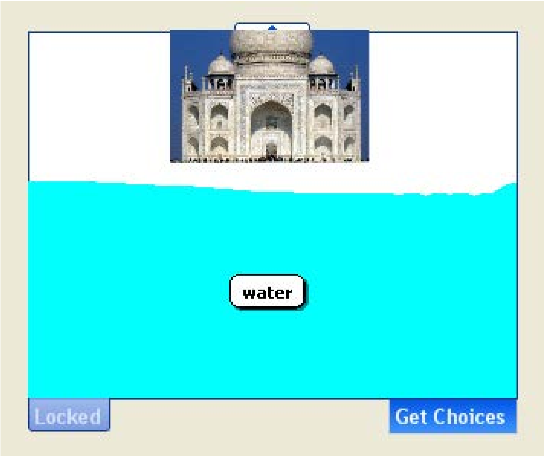
http://www.youtube.com/watch?v=E0rboU10rPo

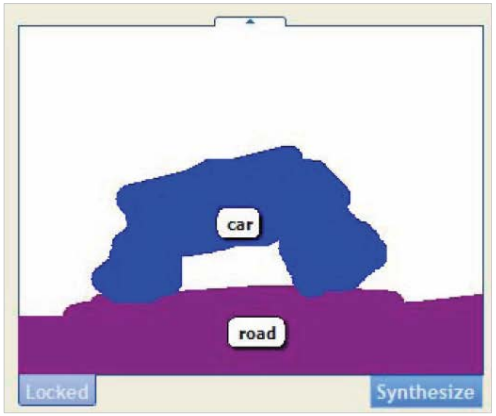
Semantic Photo Synthesis



M. Johnson, G. Brostow, J. Shotton, O. A. c, and R. Cipolla, “Semantic Photo Synthesis,” Computer Graphics Forum Journal (Eurographics 2006), vol. 25, no. 3, 2006.

Semantic Photo Synthesis [EG'06]





Johnson, Brostow, Shotton, Arandjelovic, Kwatra, and Cipolla. Eurographics 2006.

Semantic Photo Synthesis

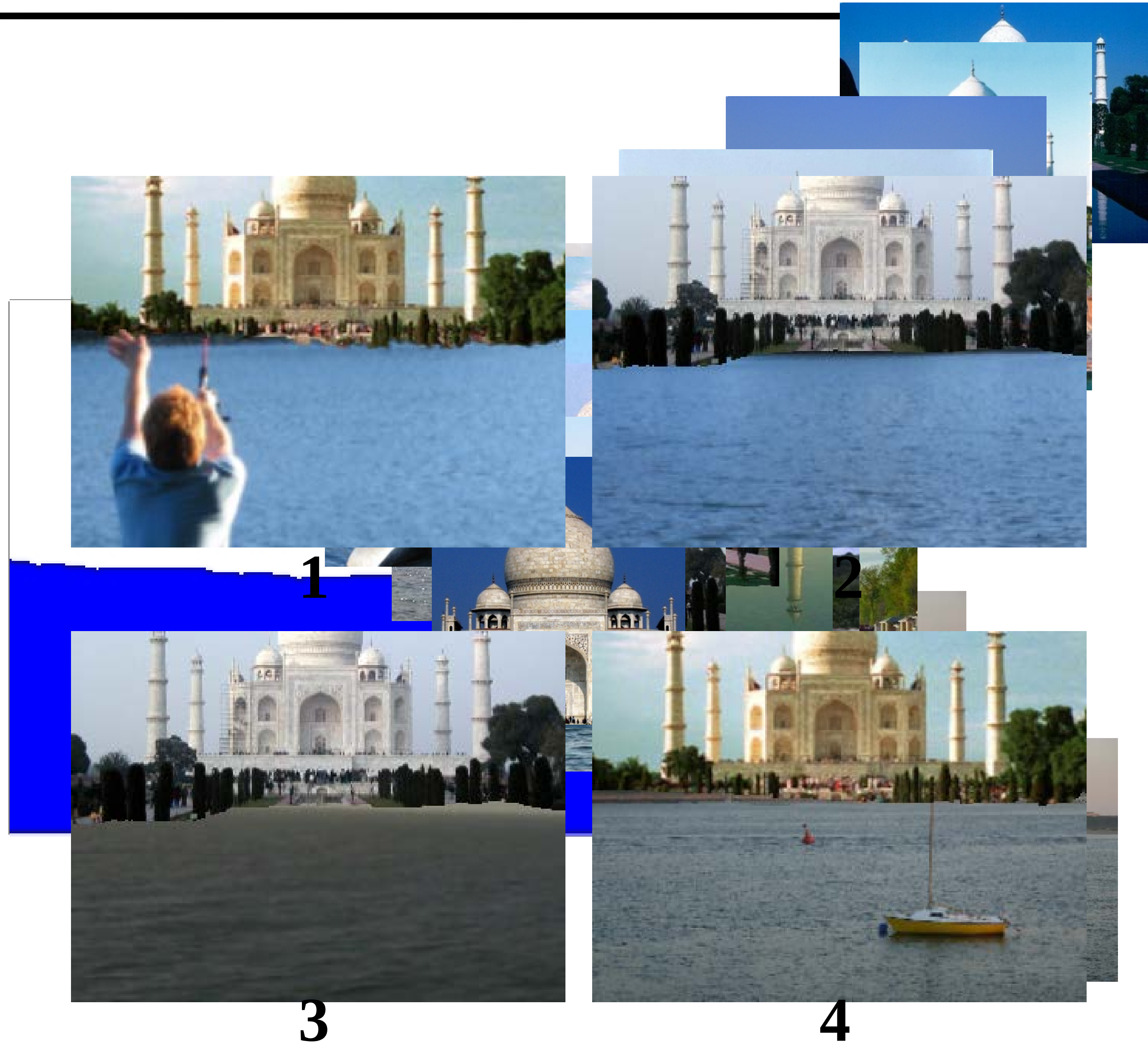
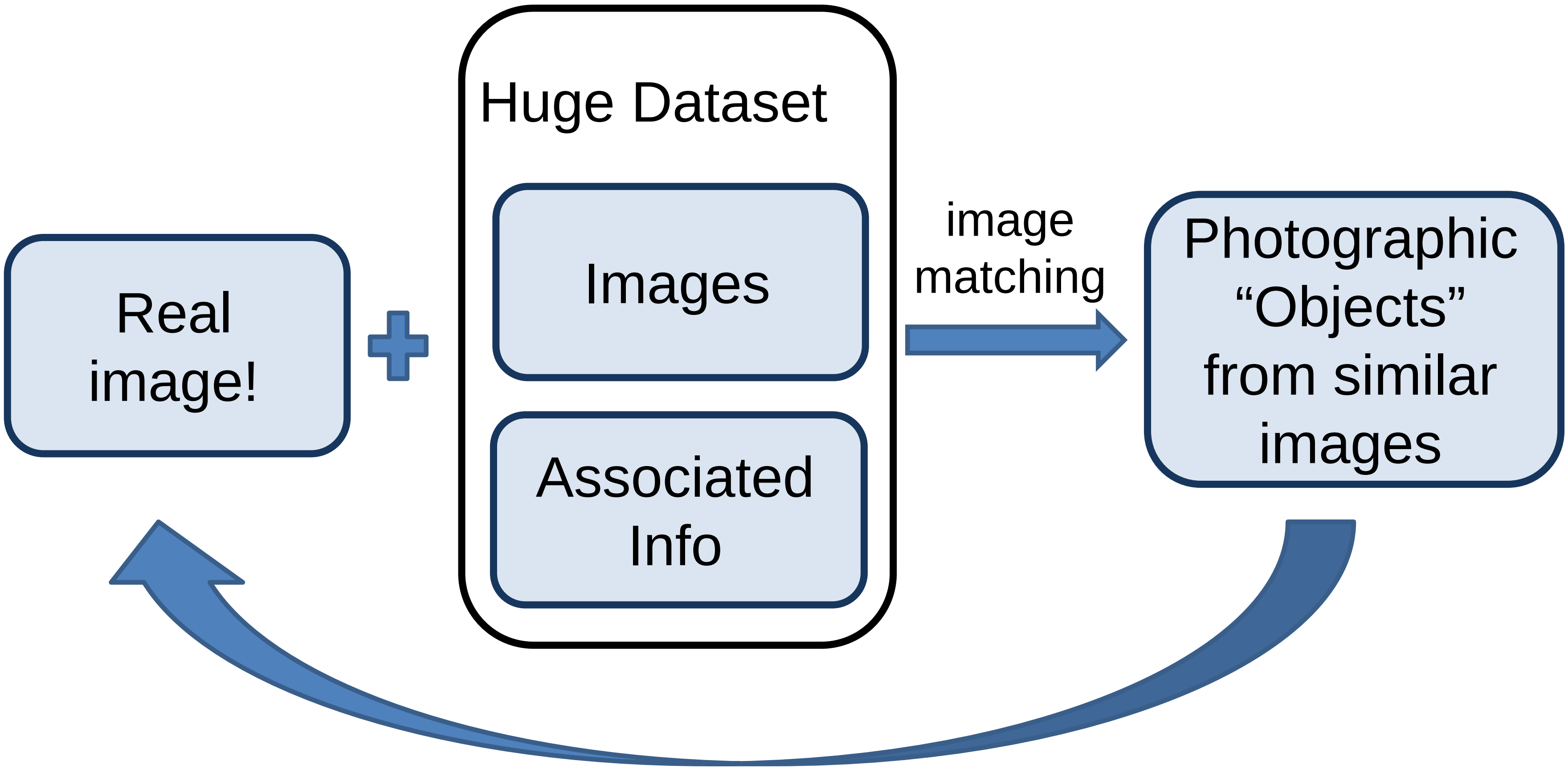


Photo Clip Art



J.-F. Lalonde, D. Hoiem, A. A. Efros, C. Rother, J. Winn, and A. Criminisi, "Photo Clip Art," ACM Transactions on Graphics (SIGGRAPH 2007), vol. 26, no. 3, Aug. 2007.

Photo Clip Art [SG'07]

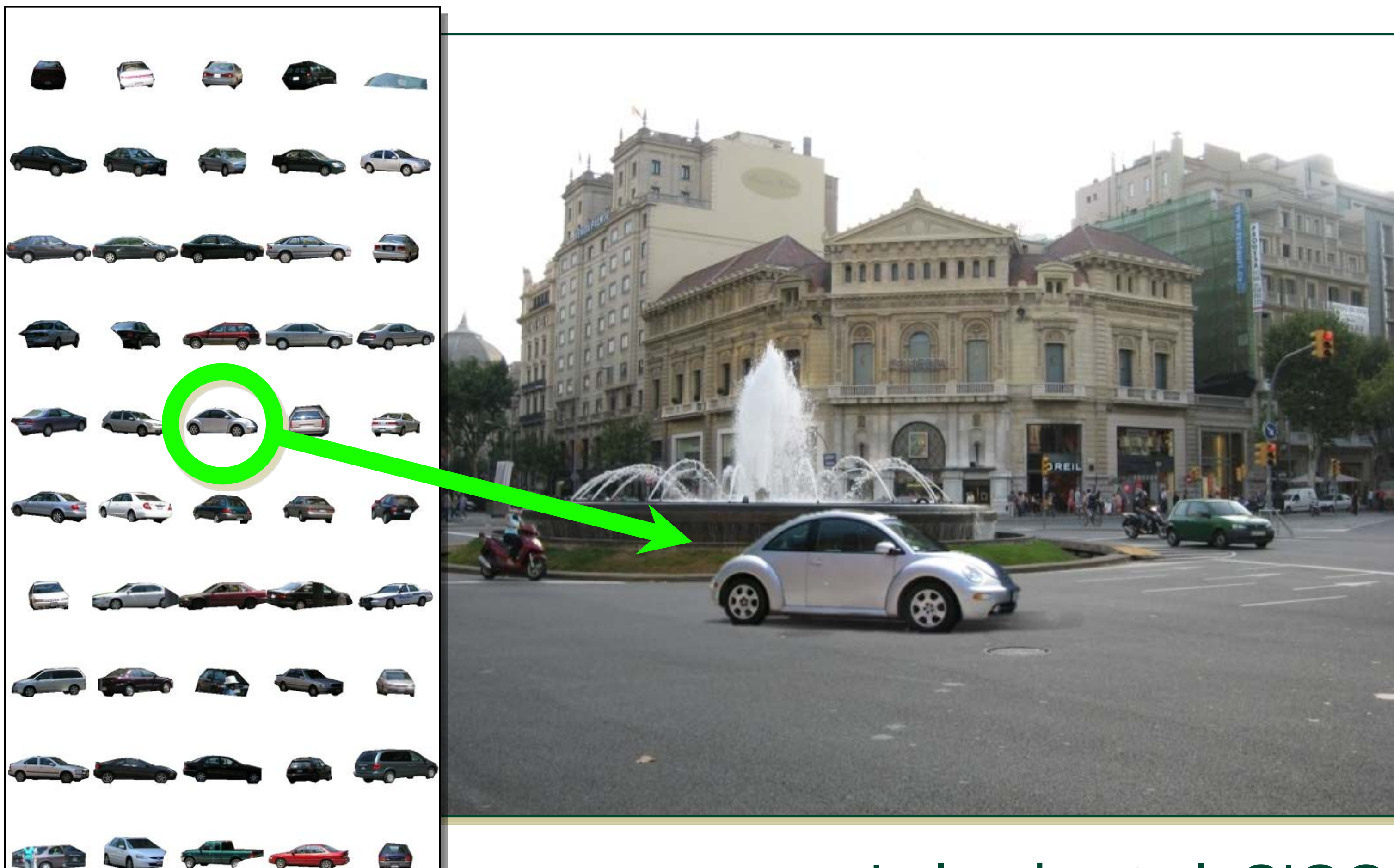
Inserting a single object -- still very hard!

- object size, orientation
- scene illumination

Lalonde et al, SIGGRAPH 2007

Photo Clip Art [SG'07]

Use database to find well-fitting object



Lalonde et al, SIGGRAPH 2007

Geometry is not enough

Illumination context

- Exact environment map is impossible
- Approximations [Khan et al., ‘06]

Database image

Environment map rough approximation

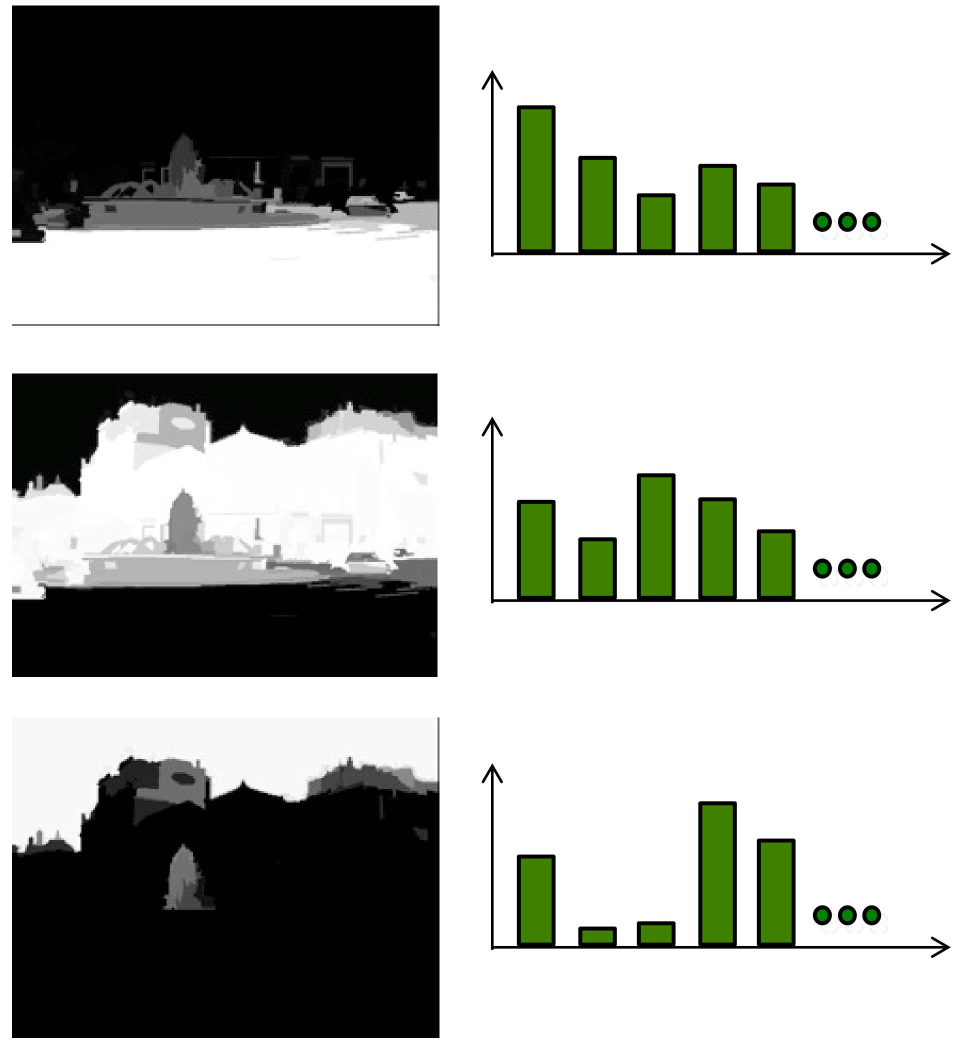
Illumination context

Database image

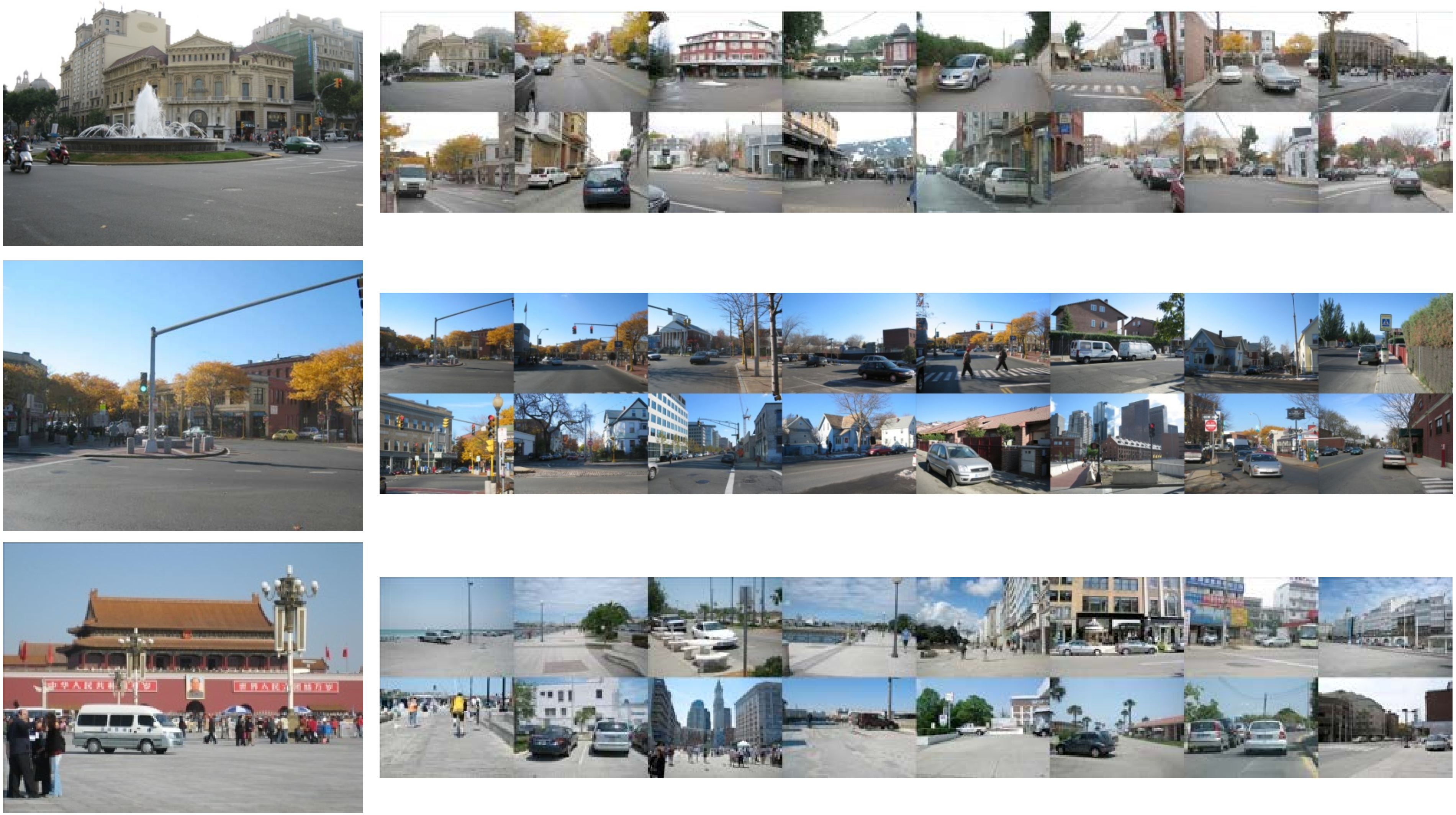
Automatic Photo Popup
Hoiem et al., SIGGRAPH '05

P(pixel|class)

CIE L*a*b* histograms



Illumination nearest-neighbors



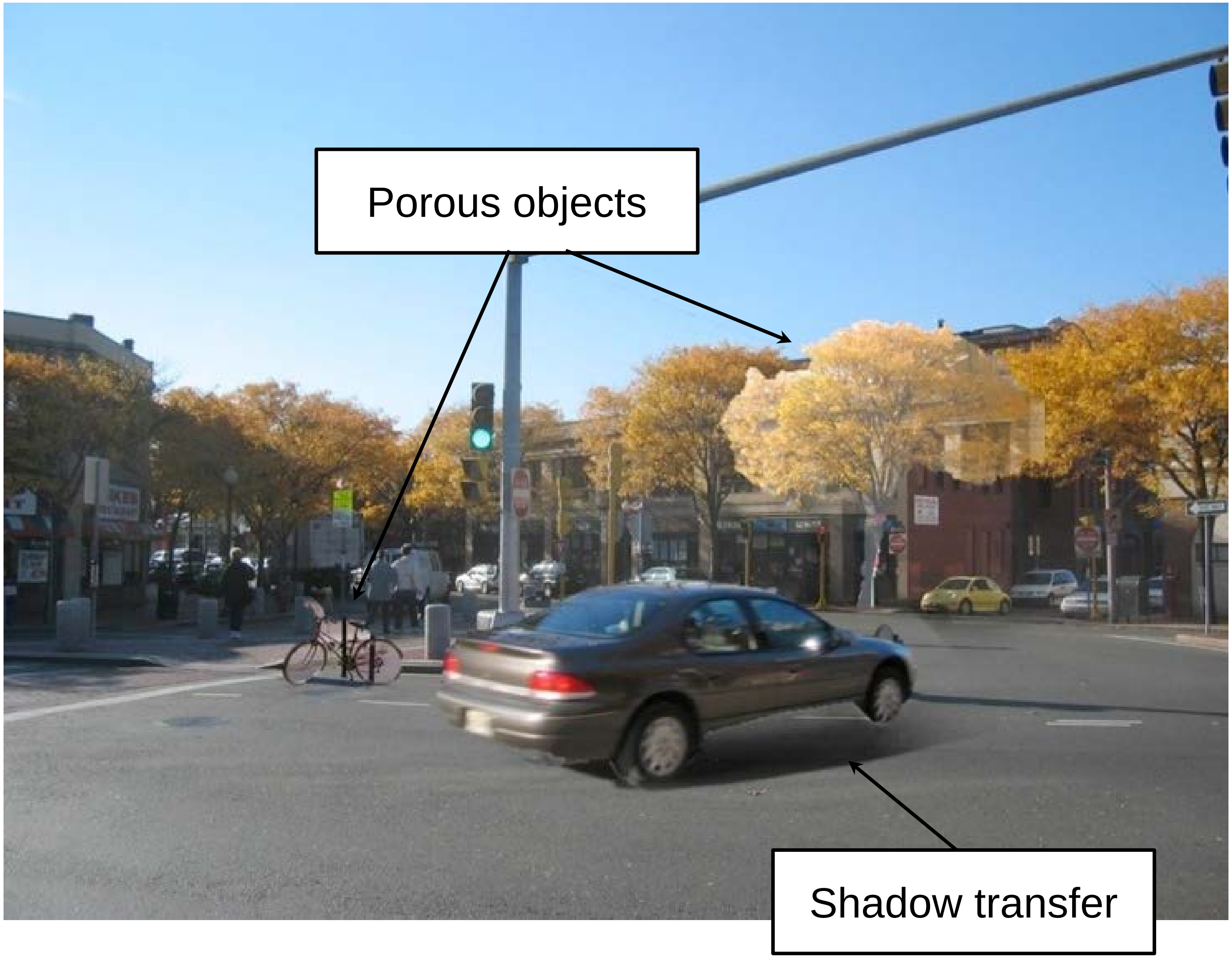
Street accident

Bridge

Painting

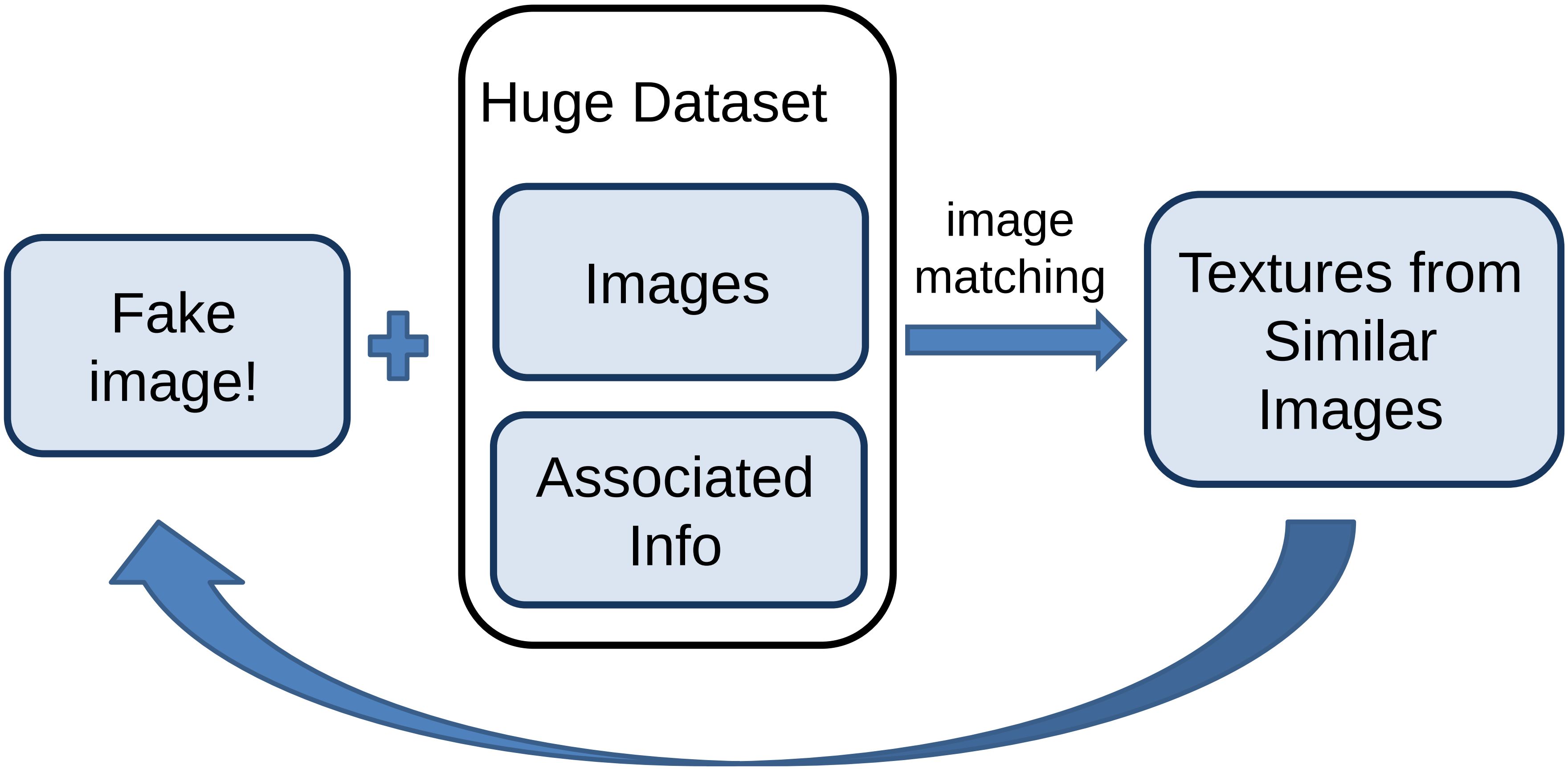
Alley

Failure cases



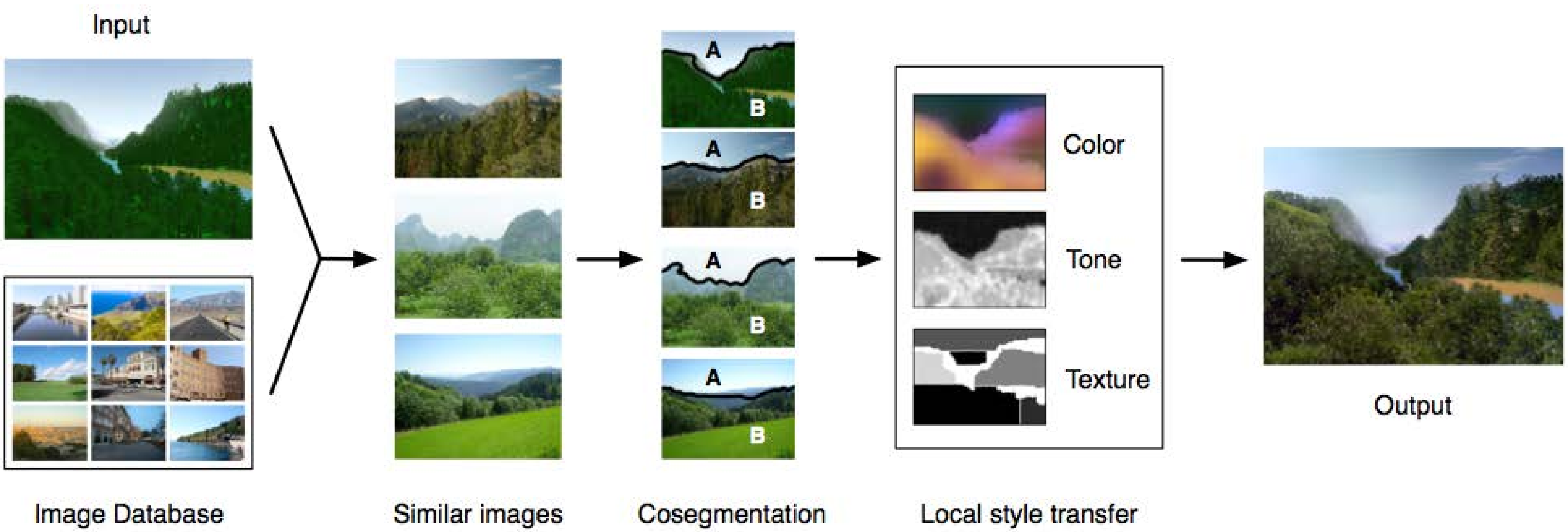
Failure cases

CG2Real



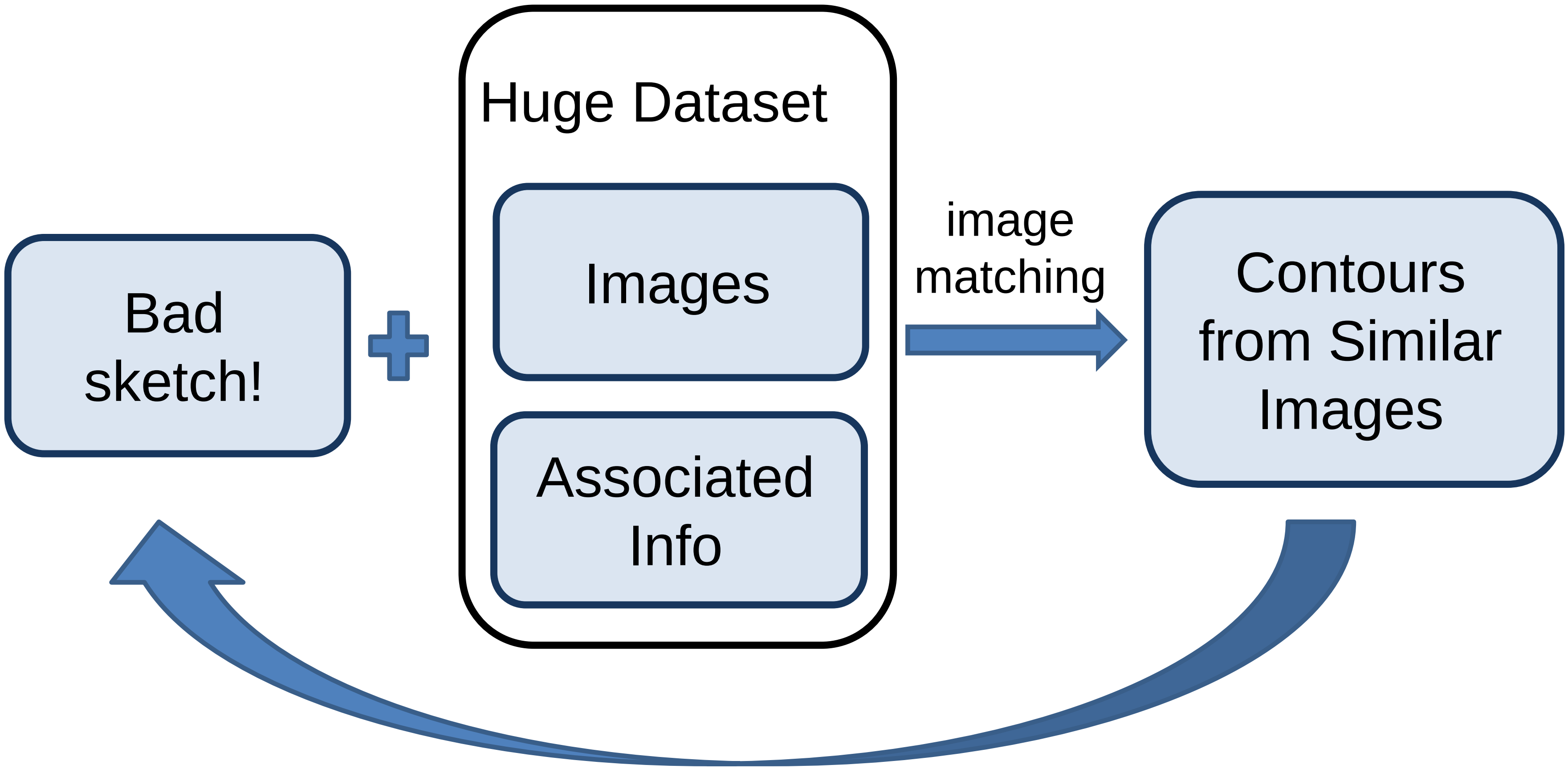
M. K. Johnson, K. Dale, S. Avidan, H. Pfister, W. T. Freeman, and W. Matusik, “CG2Real: Improving the realism of computer generated images using a large collection of photographs,” IEEE Transactions on Visualization and Computer Graphics, 2010.

CG2Real



M. K. Johnson, K. Dale, S. Avidan, H. Pfister, W. T. Freeman, and W. Matusik, “CG2Real: Improving the realism of computer generated images using a large collection of photographs,” IEEE Transactions on Visualization and Computer Graphics, 2010.

ShadowDraw


Huge Dataset
Bad sketch!
Images
Associated Info
image matching
Contours from Similar Images

ShadowDraw

http://www.youtube.com/watch?v=zh_-HUdQwow

Explore Visual Data

AverageExplorer

http://www.youtube.com/watch?v=1QgL_aPPCpM

The Dangers of Data

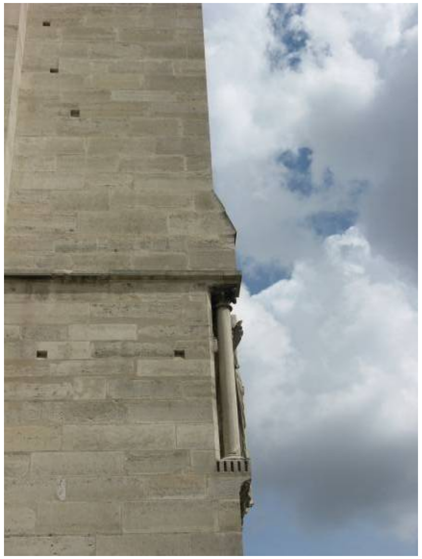
Bias

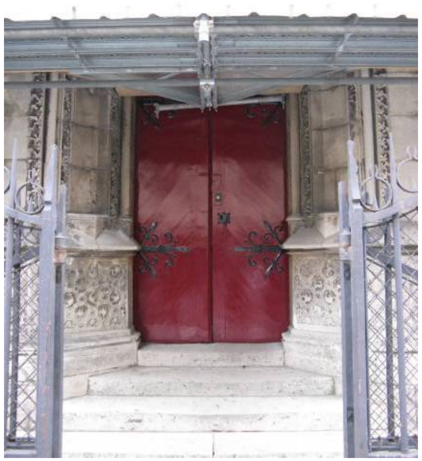
- Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
- But it's <u>not</u> random samples of visual world
- Many sources of bias:
  - Sampling bias
  - Photographer bias
  - Social bias

Flickr Paris

Real Paris

Real Notre Dame



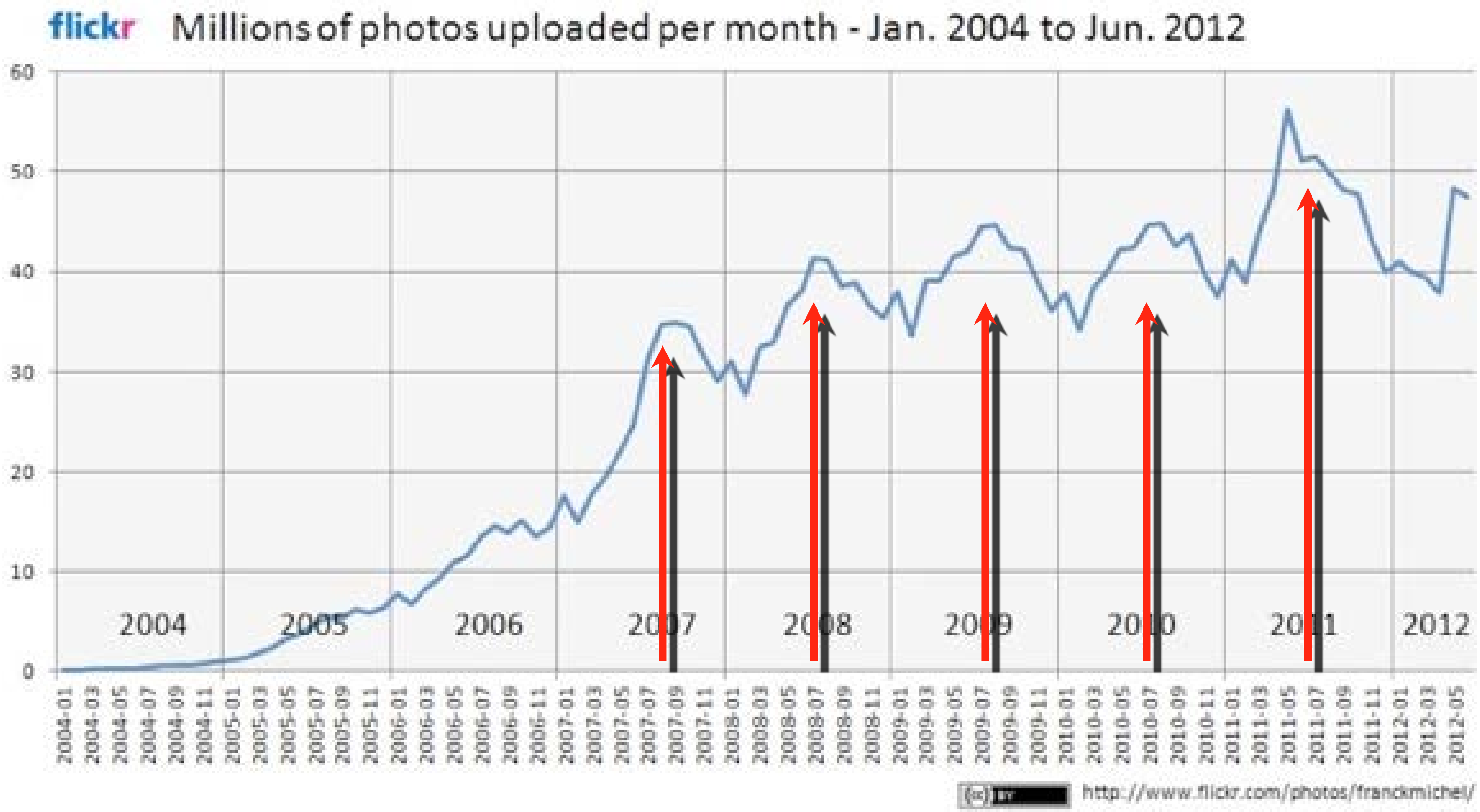


Sampling Bias

- People like to take pictures on vacation

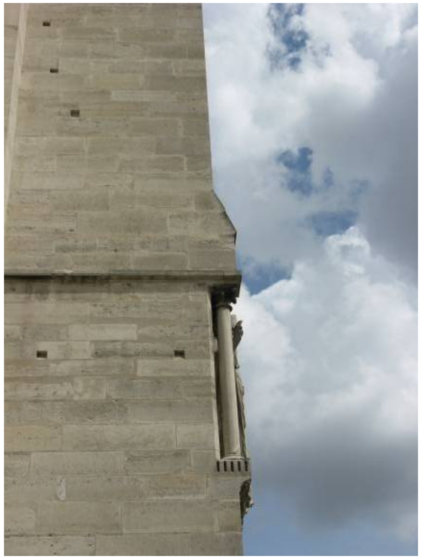
Sampling Bias

People like to take pictures on vacation



Photographer Bias

- People want their pictures to be recognizable and/or interesting



vs.

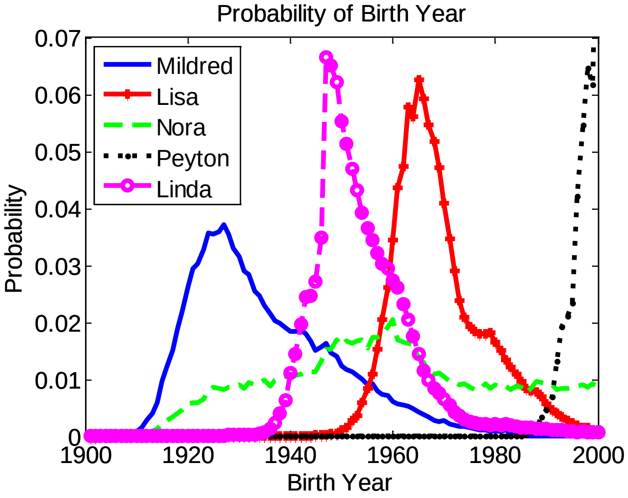
Photographer Bias

- People follow photographic conventions

vs.

Social Bias

Mildred and Lisa



Source: U.S. Social Security Administration

Gallagher et al CVPR 2008

Social Bias

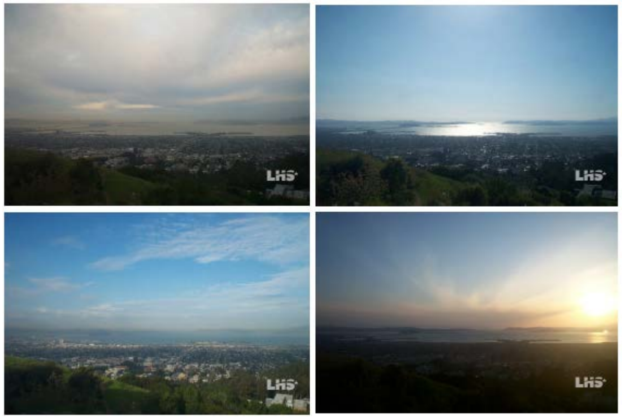
Gallagher et al CVPR 2008

Gallagher et al, CVPR 2009

Reducing / Changing Bias

Street side
Google StreetView

Satellite
google.com



Webcams

- Autonomous capture methods can reduce / change bias
  - But it won't go away completely
- Sometimes you can just pick your data to suit your problem, but not always…