What can we learn from a single image?

© Quint Buchholz

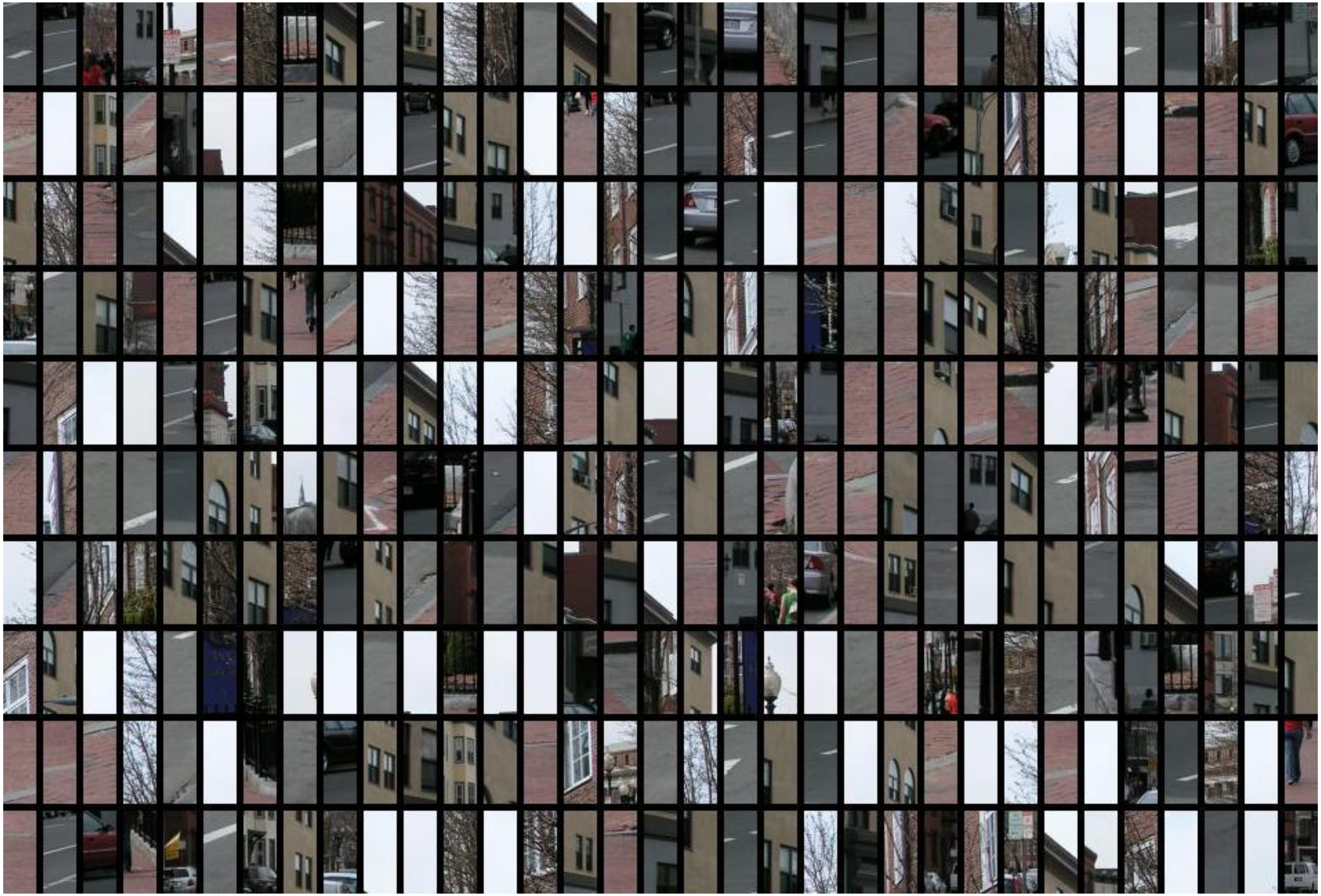
Alexei (Alyosha) Efros

Carnegie Mellon University

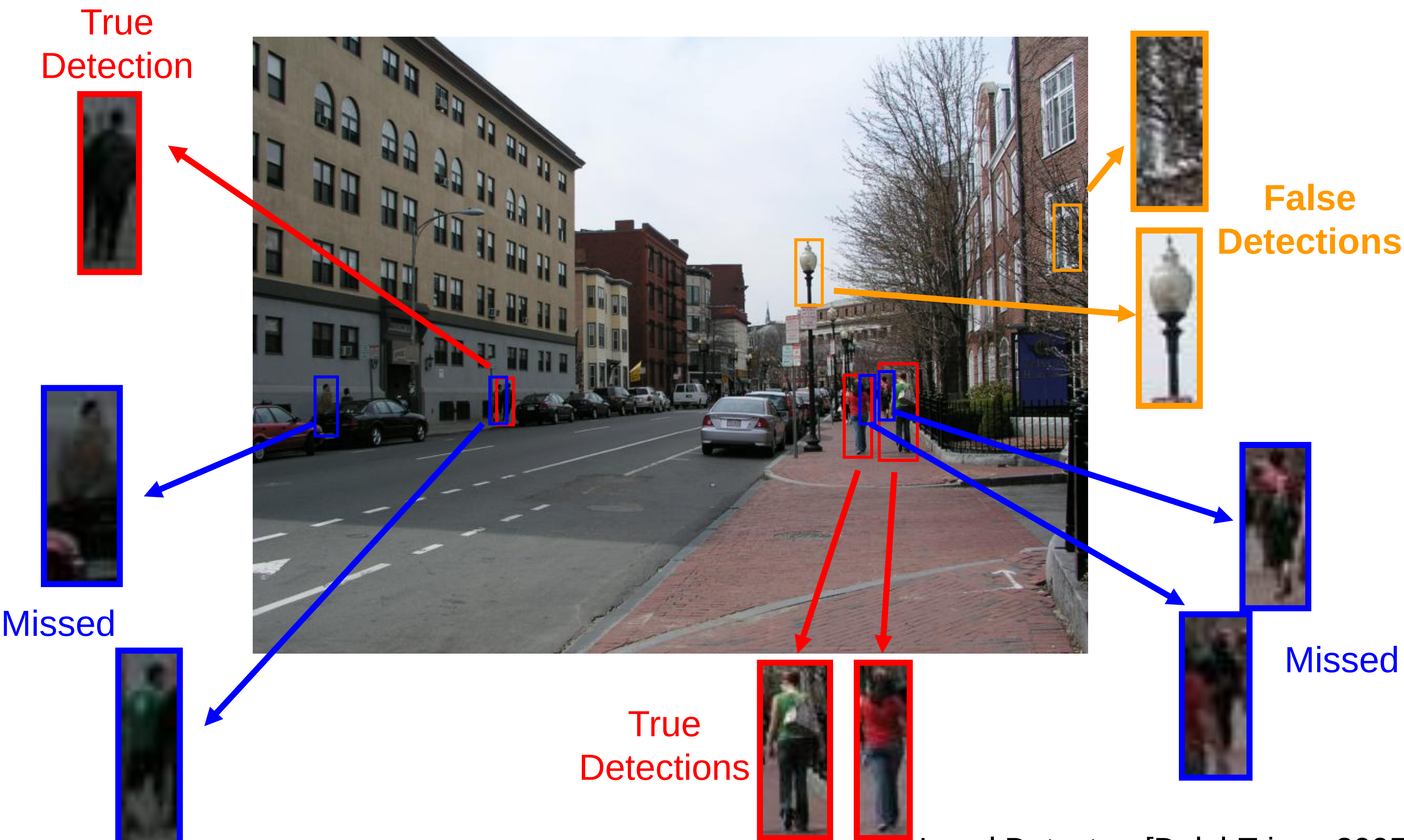
What do we see?

The Miserable Life of an Object Detector

What the Detector Sees



State-of-the-Art Pedestrian Detection



Local Detector: [Dalal-Triggs 2005]

Importance of Looking Globally

Claude Monet
Gare St.Lazare
Paris, 1877

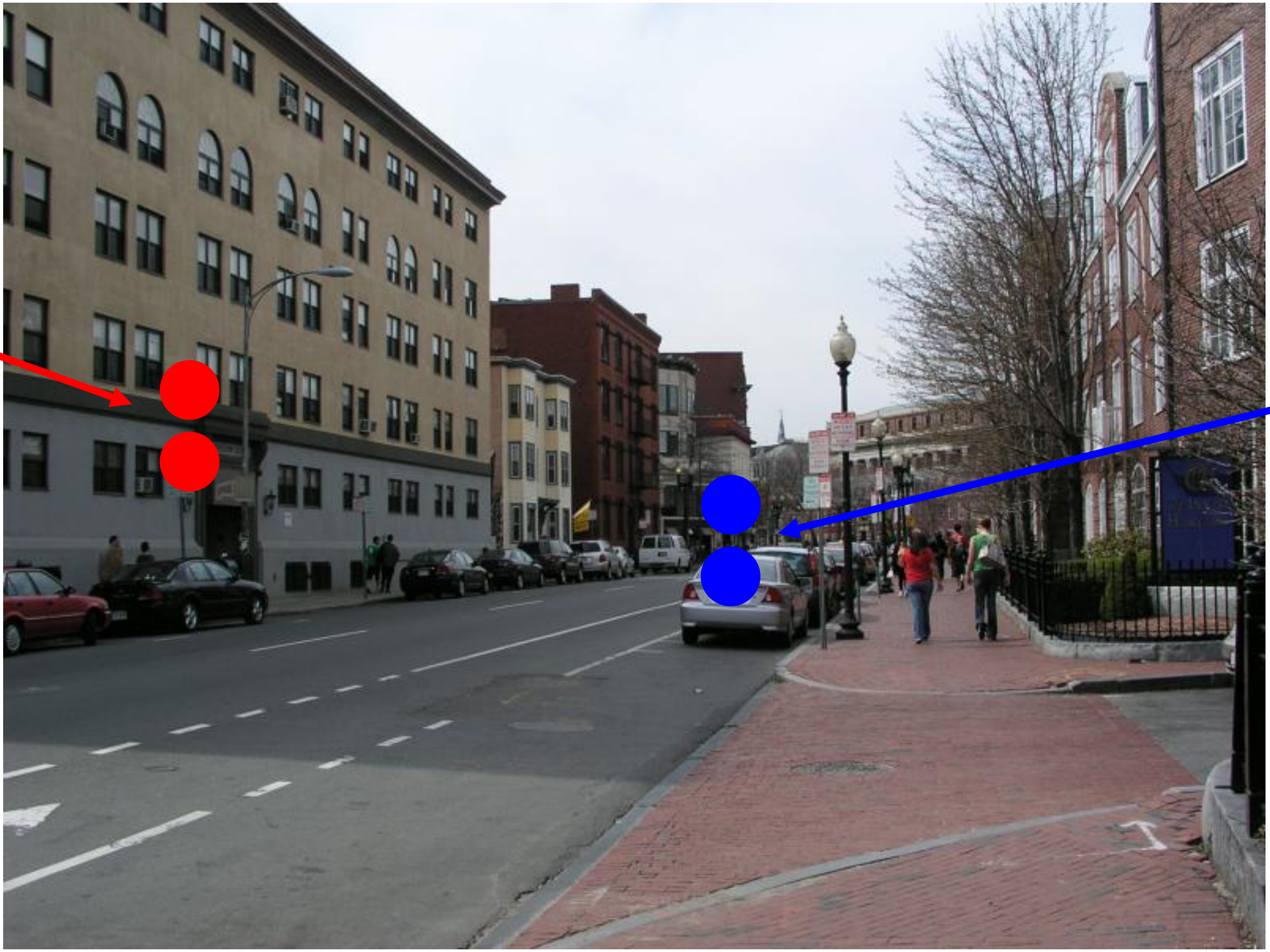
There is almost nothing inside!

Seeing less than you think…

Seeing less than you think…

Need to think "outside the box"

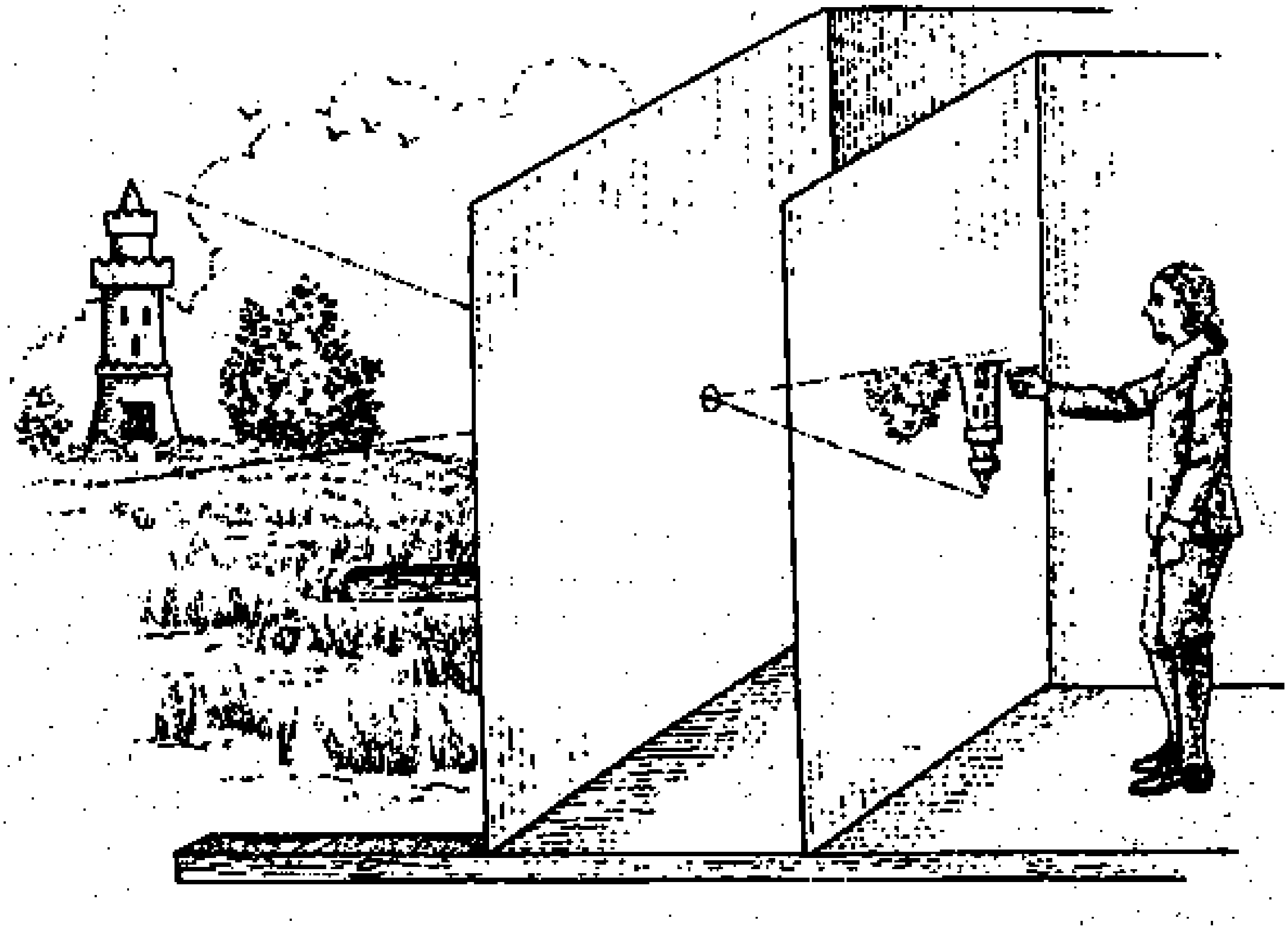
Real Relationships are 3D



Close

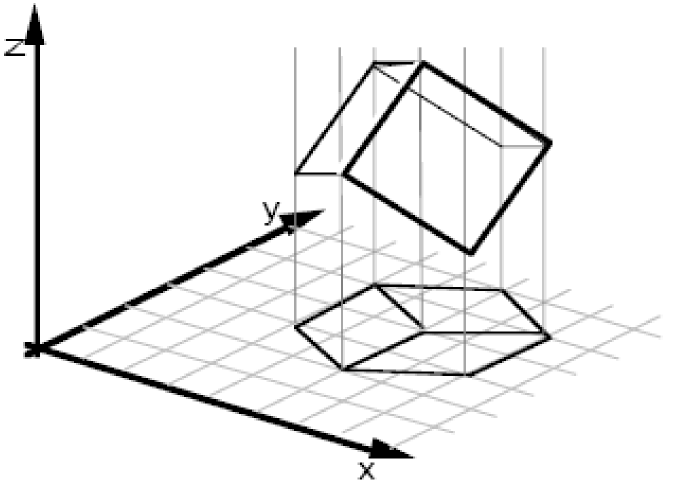
Not Close

Imaging Process



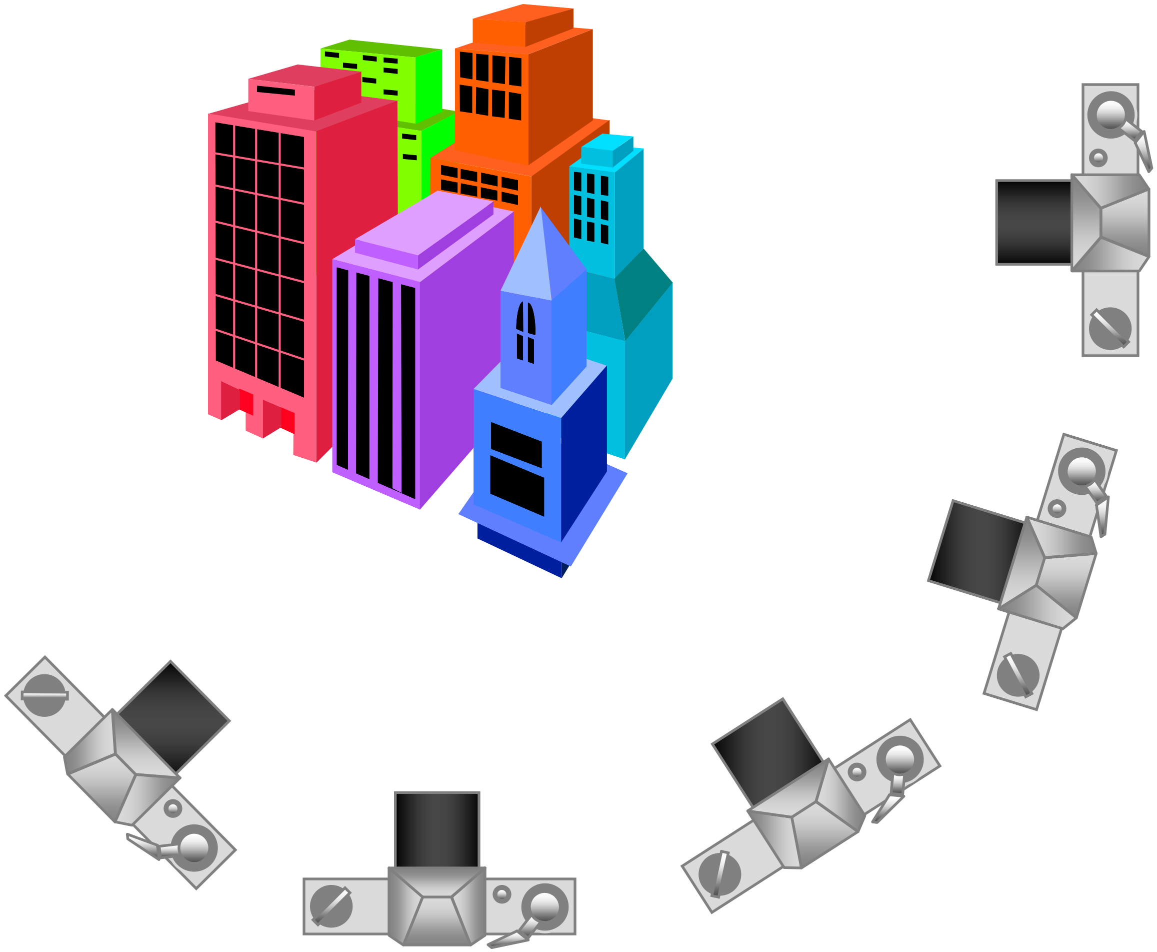
Unsolvable Problem

- Recovering 3D geometry from **single** 2D projection
- Infinite number of possible solutions!



from [Sinha and Adelson 1993]

Ecological Optics



J.J. Gibson's
"actively exploring organism"

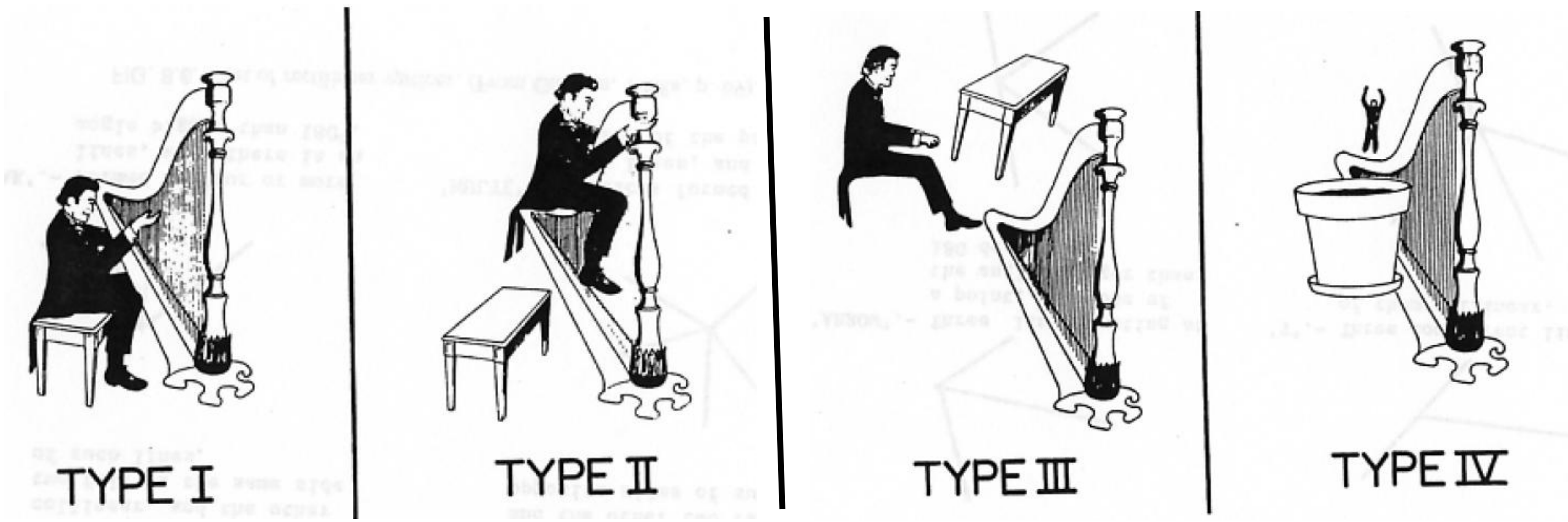
Our World is Structured

Abstract World

Our World

Image Credit (left): F. Cunin
and M.J. Sailor, UCSD

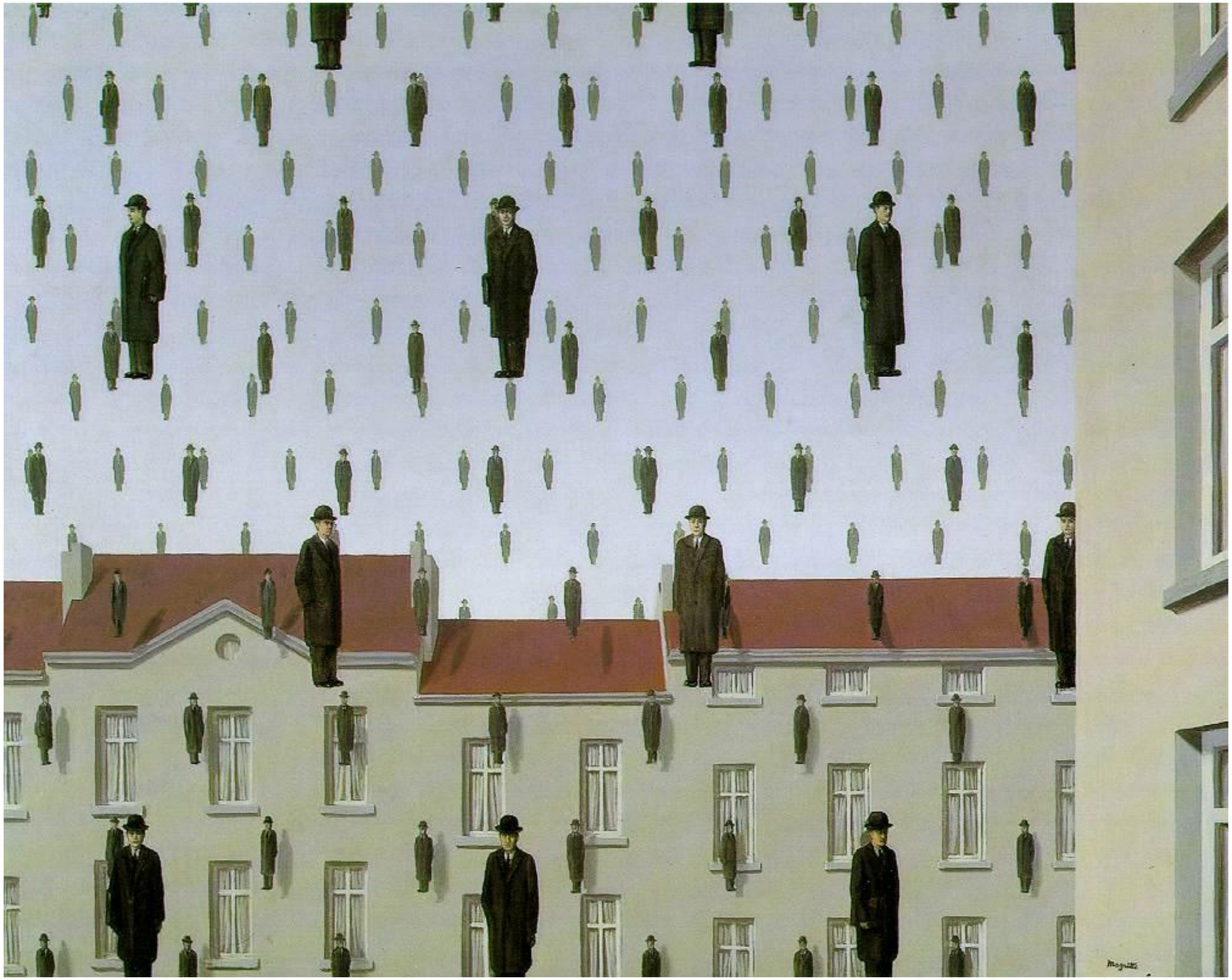
Understanding Scenes



Hock, Romanski, Galie, & Williams 1978

- Biederman's Relations among Objects in a Well-Formed Scene (1981):
  - Support
  - Size
  - Position
  - Interposition
  - Likelihood of Appearance

Support



Rene Magritte, *Golconde*

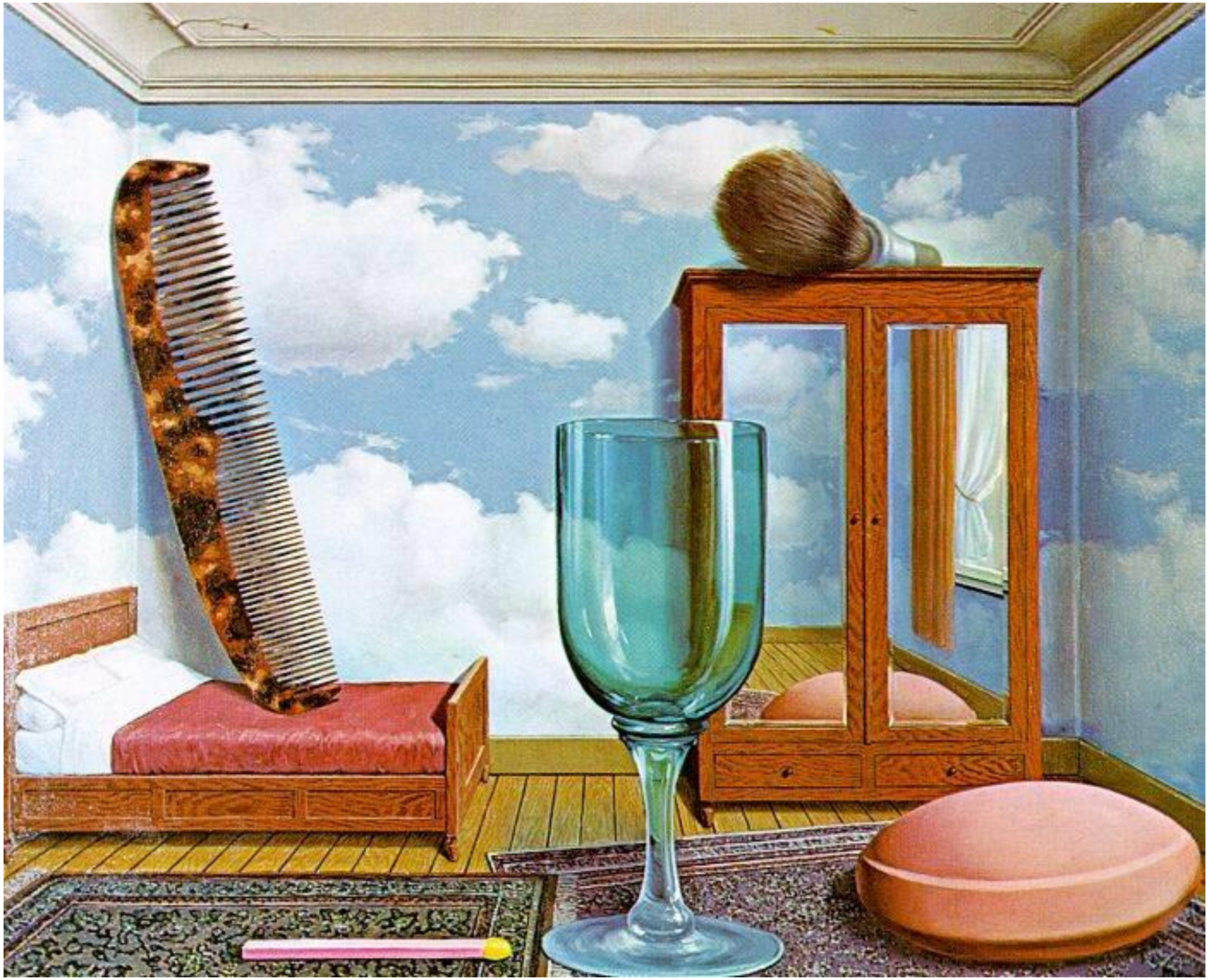
Size

Rene Magritte, *The Listening Room*

Interposition

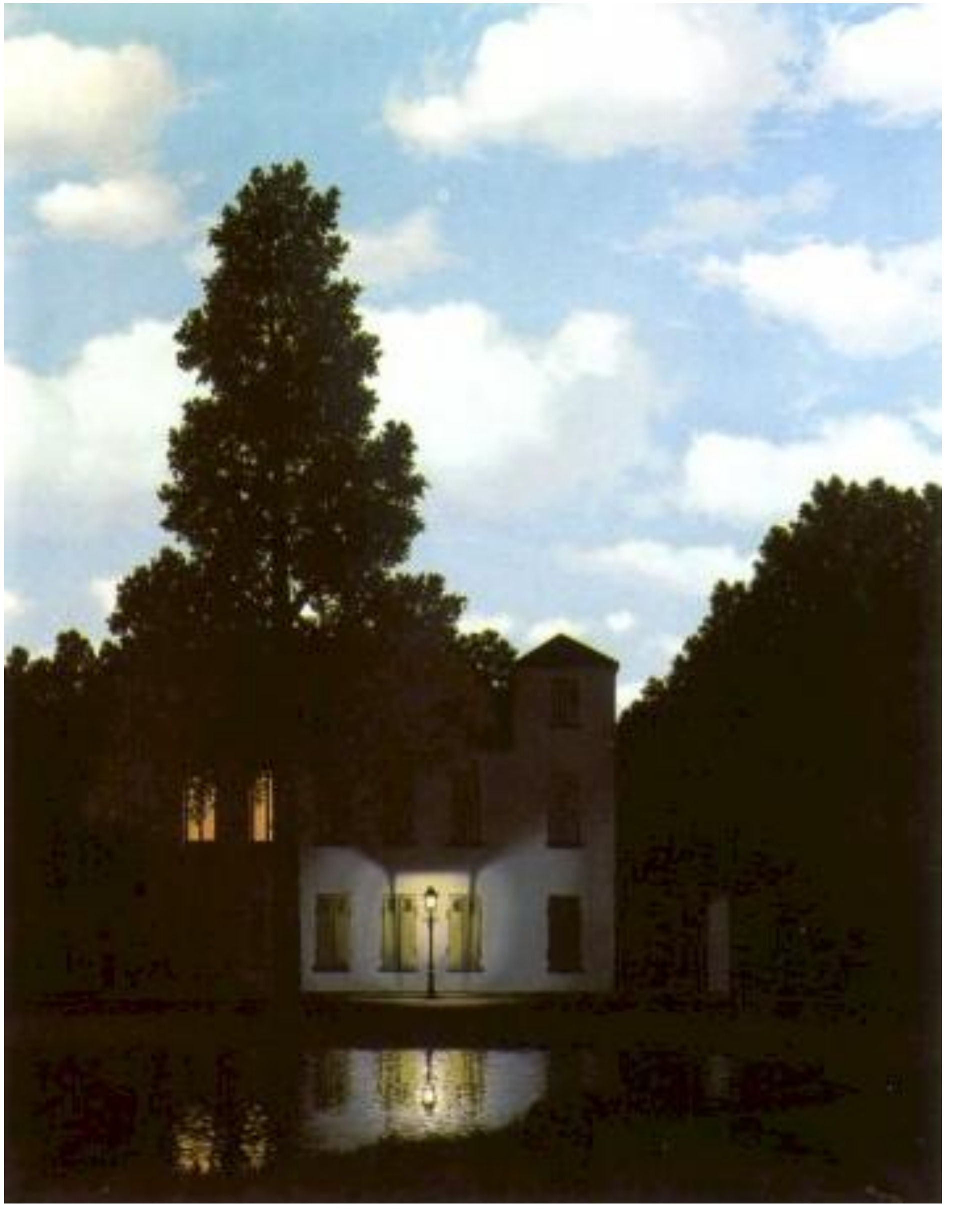
Rene Magritte, *Black Check*

Position, Probability, Size



Rene Magritte, *Personal Values*

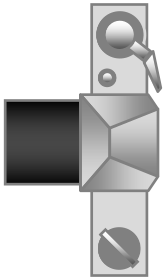
+ *illumination*



Our Goal: Scene Understanding

- scene layout
- occlusions
- camera viewpoint
- scale
- illumination
- location semantics

Ecological *Statistics*



Labeled Data

LabelMe, Caltech 101, PASCAL, etc.

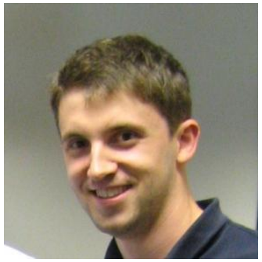
Unlabelled Data

Flickr, Google, YouTube, etc.

Collaborators

- Derek Hoiem
  - (PhD 2007, now assistant professor at UIUC)
  - co-advised with **Martial Hebert**

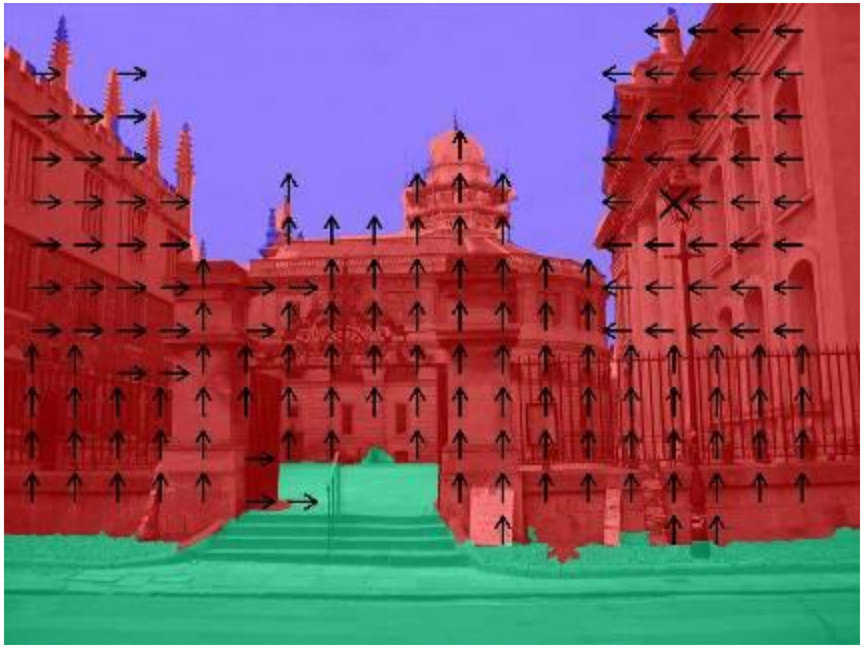
- James Hays
  - (PhD 2009, now assistant professor at Brown University)



- Jean-Francois Lalonde
  - (PhD 2010 ??)
  - co-advised with **Srinivas Narasimhan**

Thanks for many great discussions while at Oxford:
Mark Everingham, Josef Sivvic, Fred Schaffalitsky
Andrew Fitzgibbon, and, of course, AZ.

Scene Layout



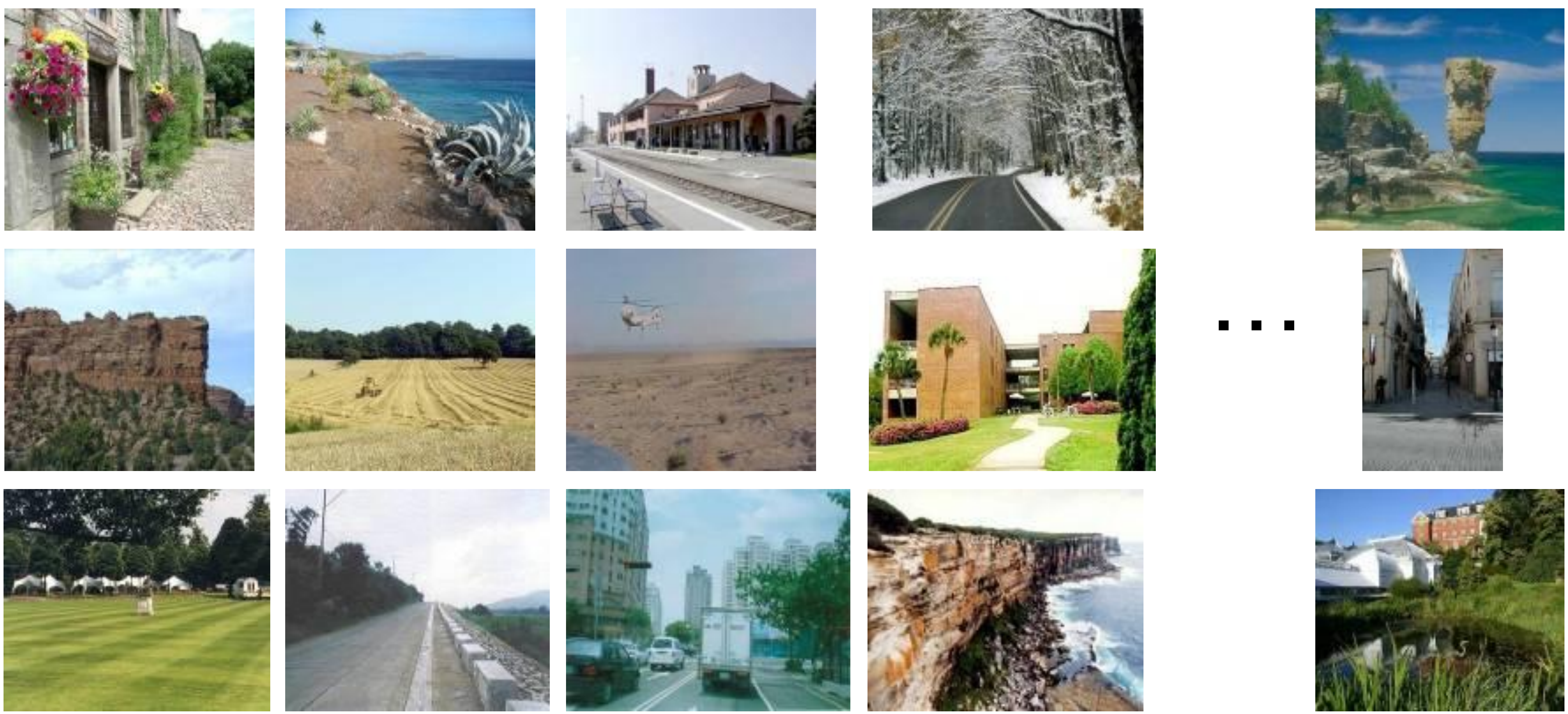
Goal: learn labeling of image into 7 <u>Geometric Classes</u>:

- **Support (ground)**
- **Vertical**
  - Planar: facing **Left** **(←)**, **Center** **(↑), Right** **(→)**
  - Non-planar: **Solid** **(X)**, **Porous** or wiry **(O)**
- **Sky**

Hoiem et al 2005

Learn from labeled data

- 300 outdoor images from Google Image Search



What cues to use?

Vanishing points, lines

Color, texture, image location

Texture gradient

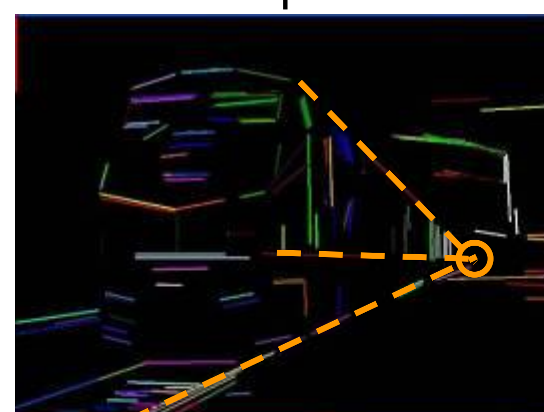
Weak Geometric Cues

Color

Texture

Location

Perspective



Need Good Spatial Support

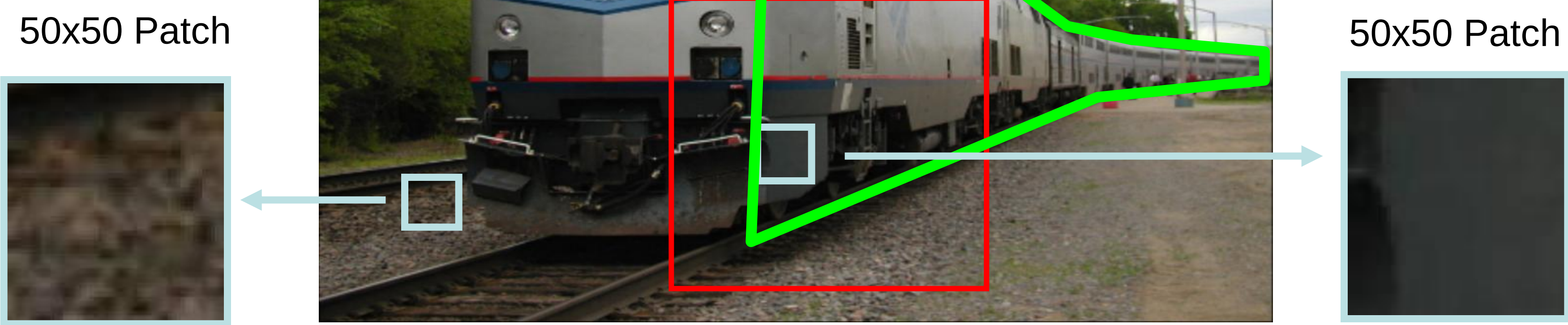
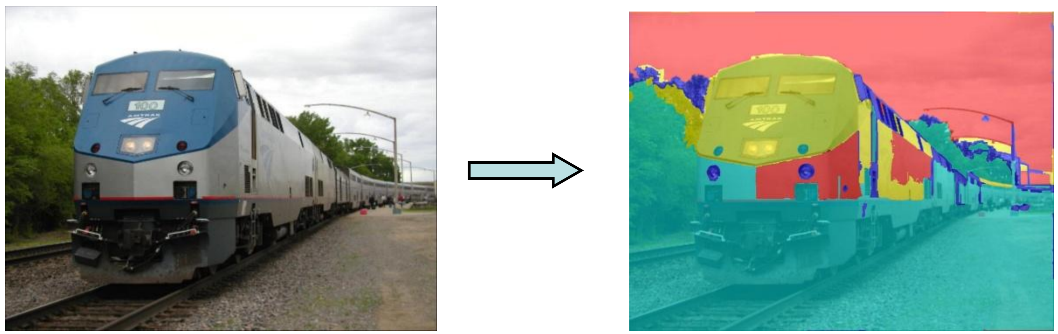


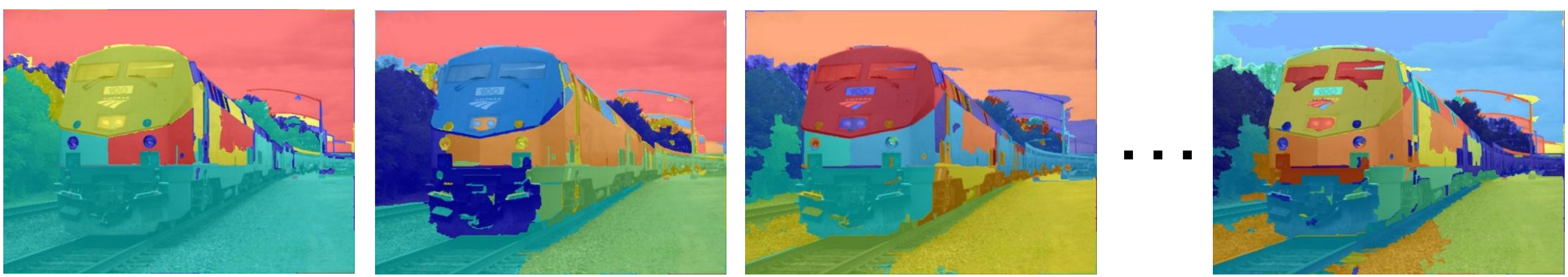
Image Segmentation

- Naïve Idea #1: segment the image



  – Chicken & Egg problem

- Naïve Idea #2: <u>multiple</u> segmentations



  – Decide later which segments are good

Estimating surfaces from segments

- We want to know:
  - Is this a good (coherent) segment?

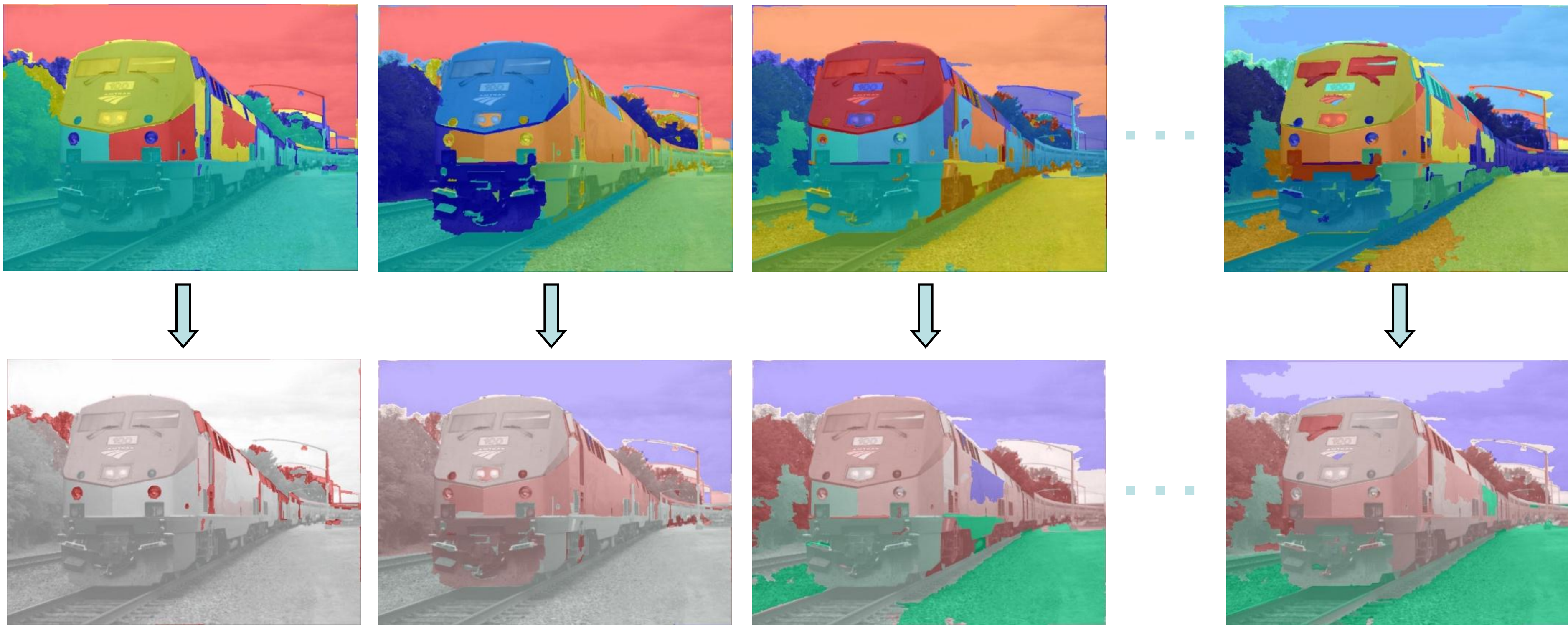
    P(*good segment* | *data*)

  - If so, what is the surface label?

    P(*label* | *good segment*, *data*)

- *Learn* these likelihoods from training images
  - we use Boosted Decision Trees

Labeling Segments

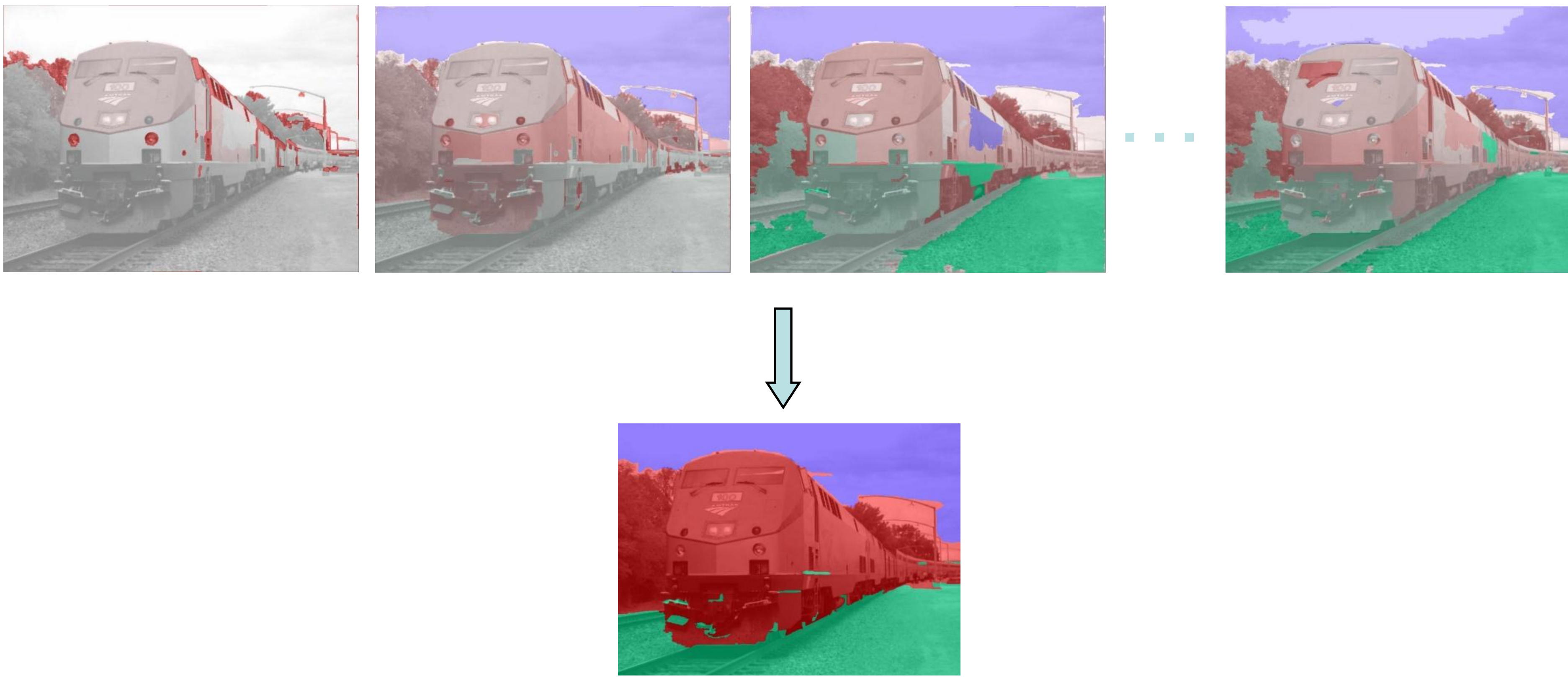


For each segment:

- Get P(*good segment* | *data*) P(*label* | *good segment*, *data*)

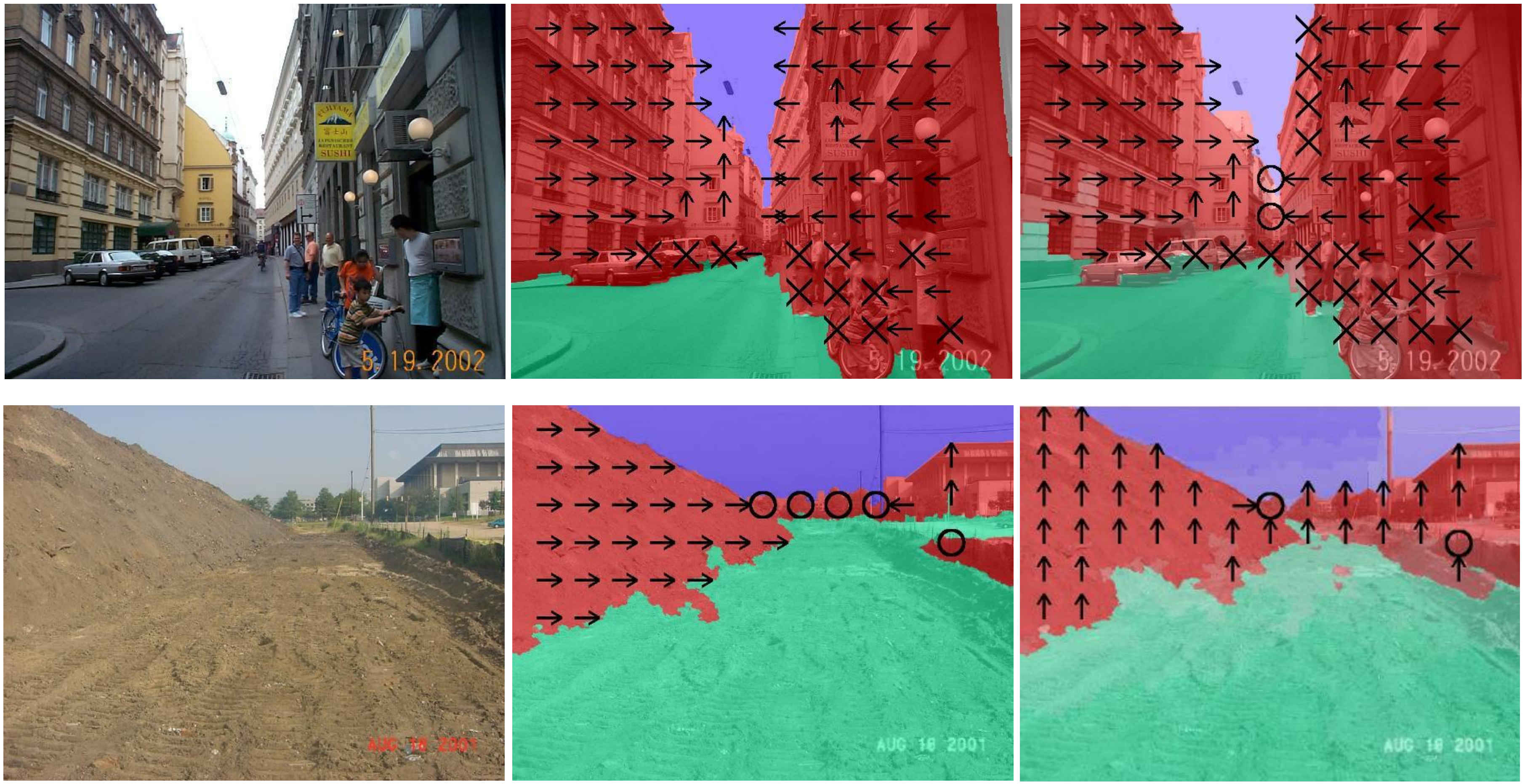
Image Labeling

Labeled Segmentations



Labeled Pixels

Results

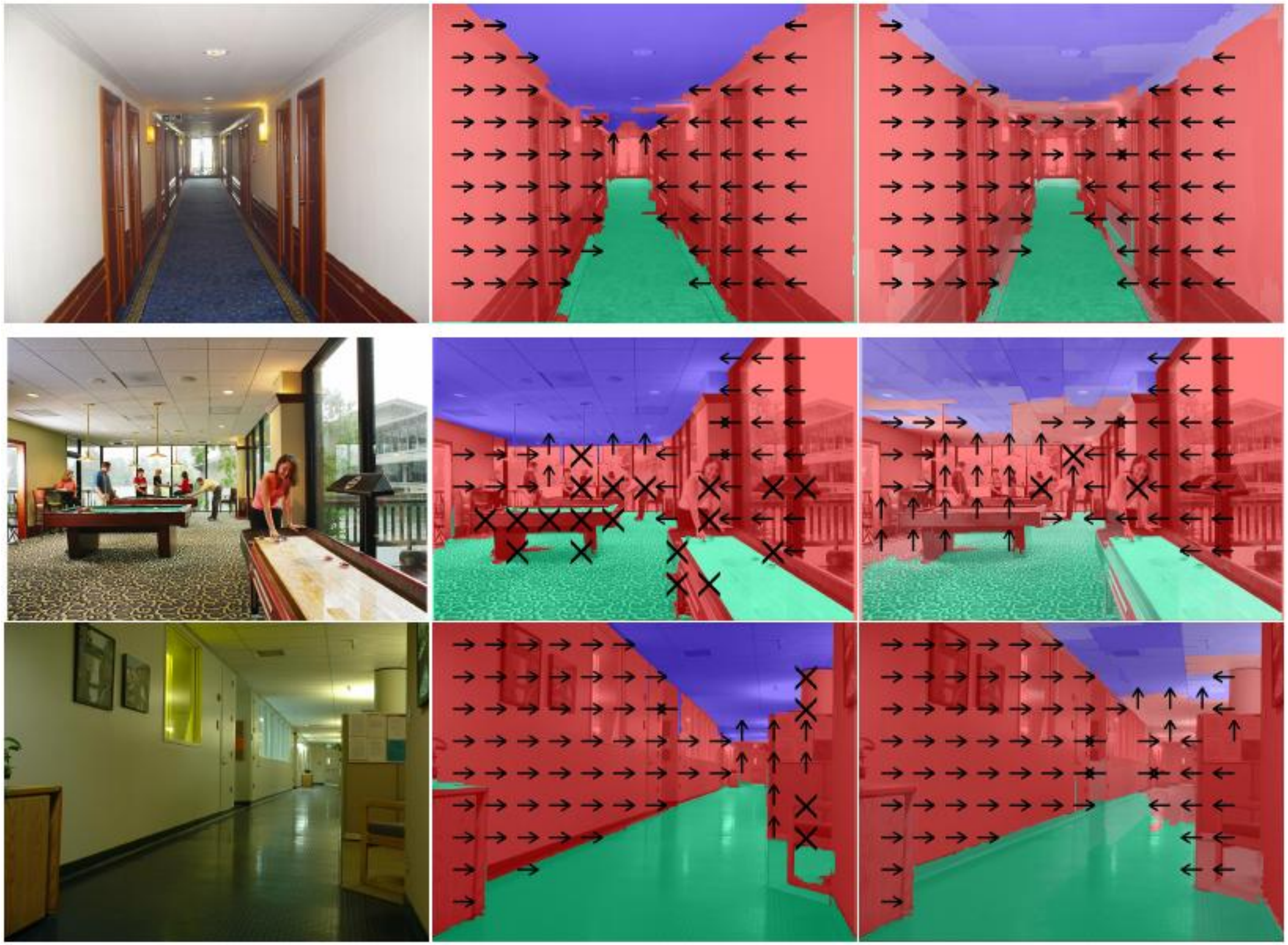


Input Image **Ground Truth** **Our Result**

No Hard Decisions

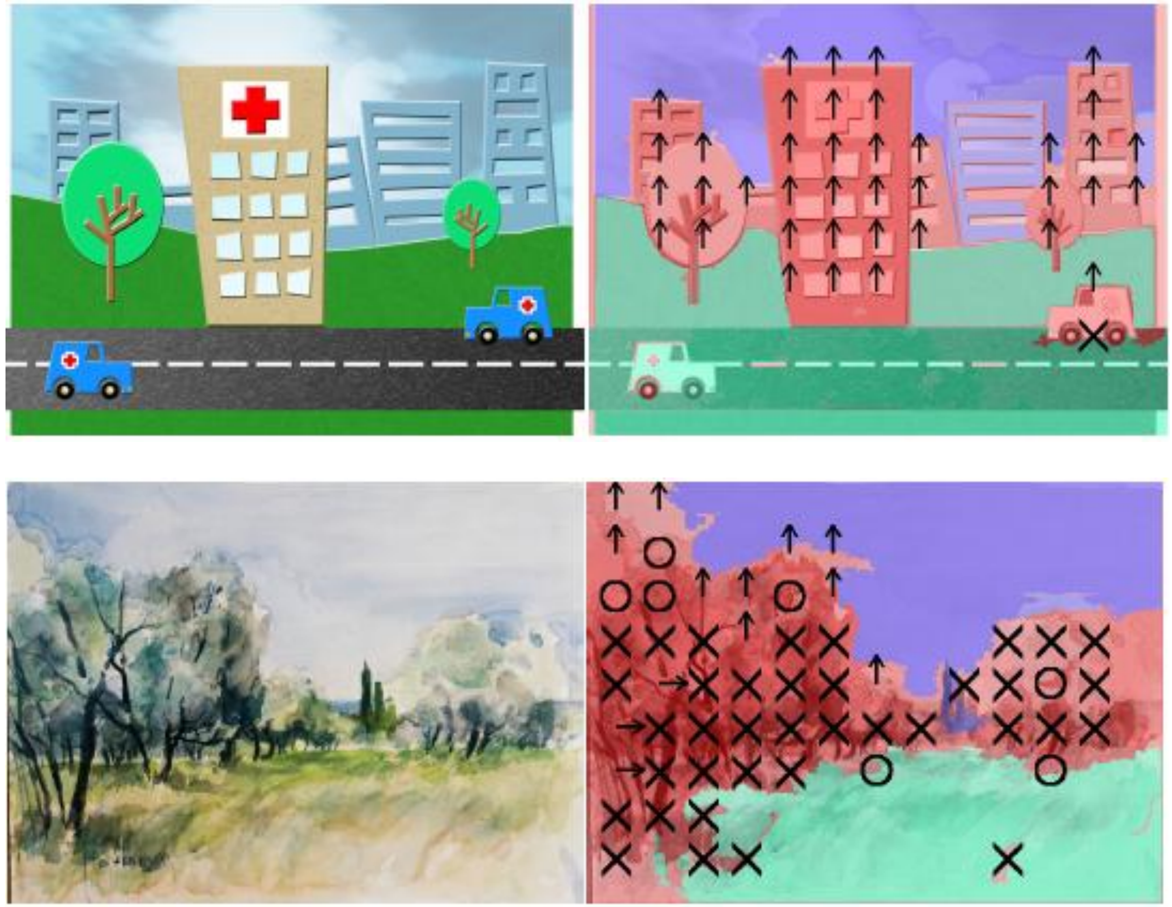
See also: Make3D, Saxena et al, 2007

Indoor Images



Input Image   Ground Truth   Our Result

Paintings



Input Image Our Result

Graphics application:
Automatic Photo Pop-up (SIGGRAPH'05)

Original Image

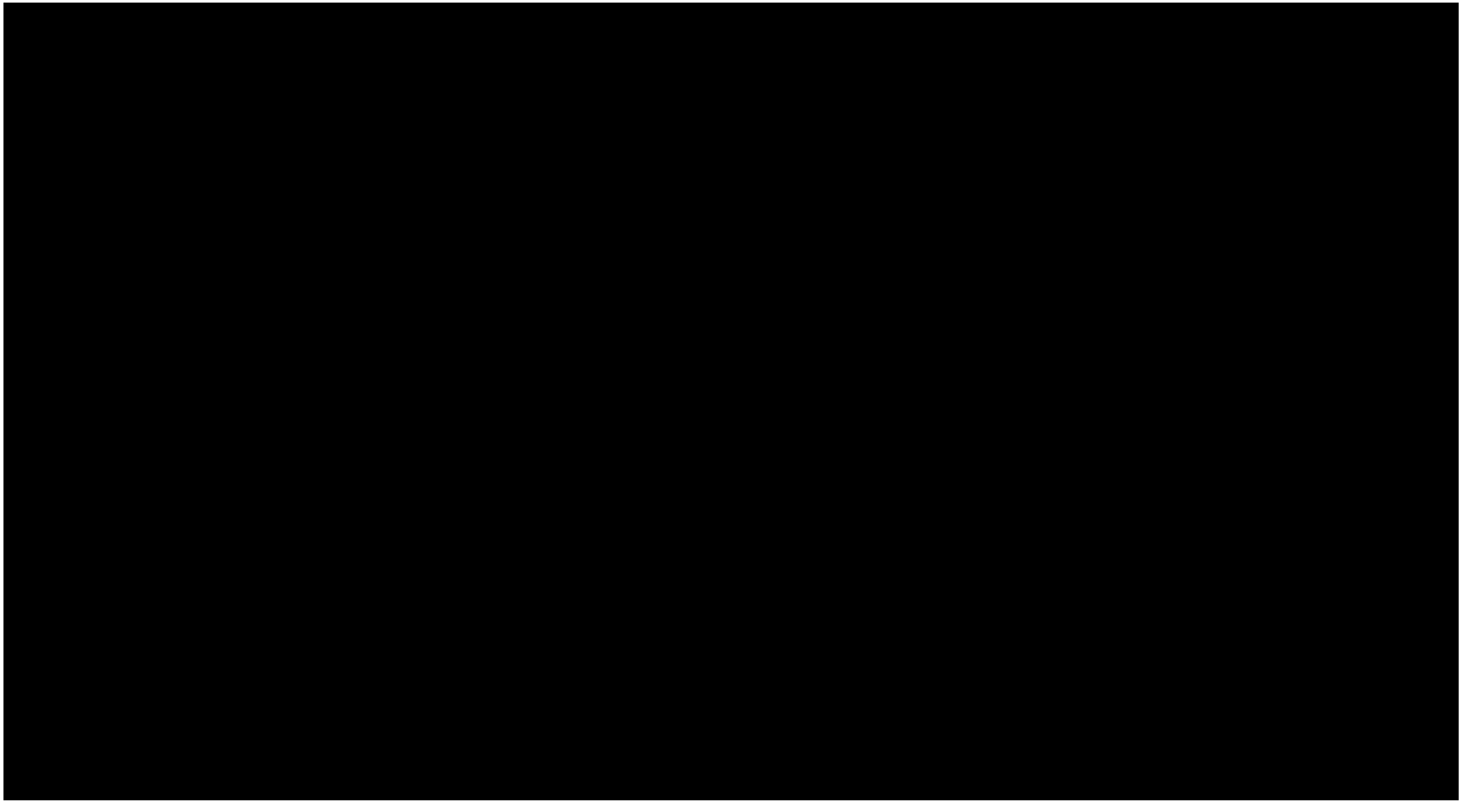
Geometric Labels

Fit Segments

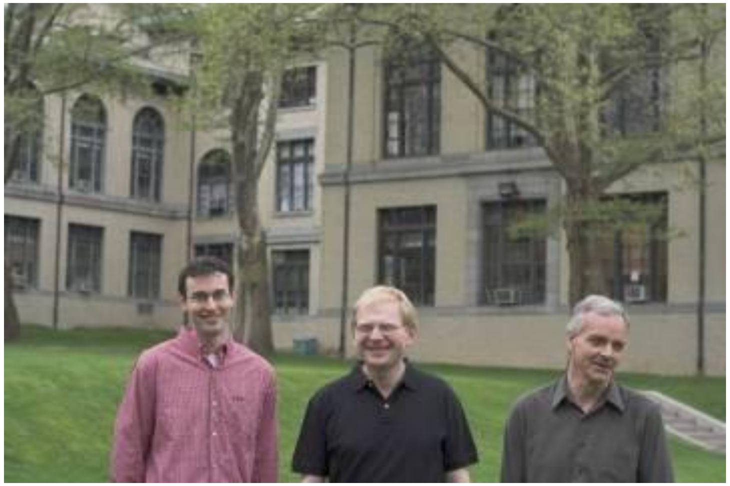
Cut and Fold

Novel View

Automatic Photo Pop-up

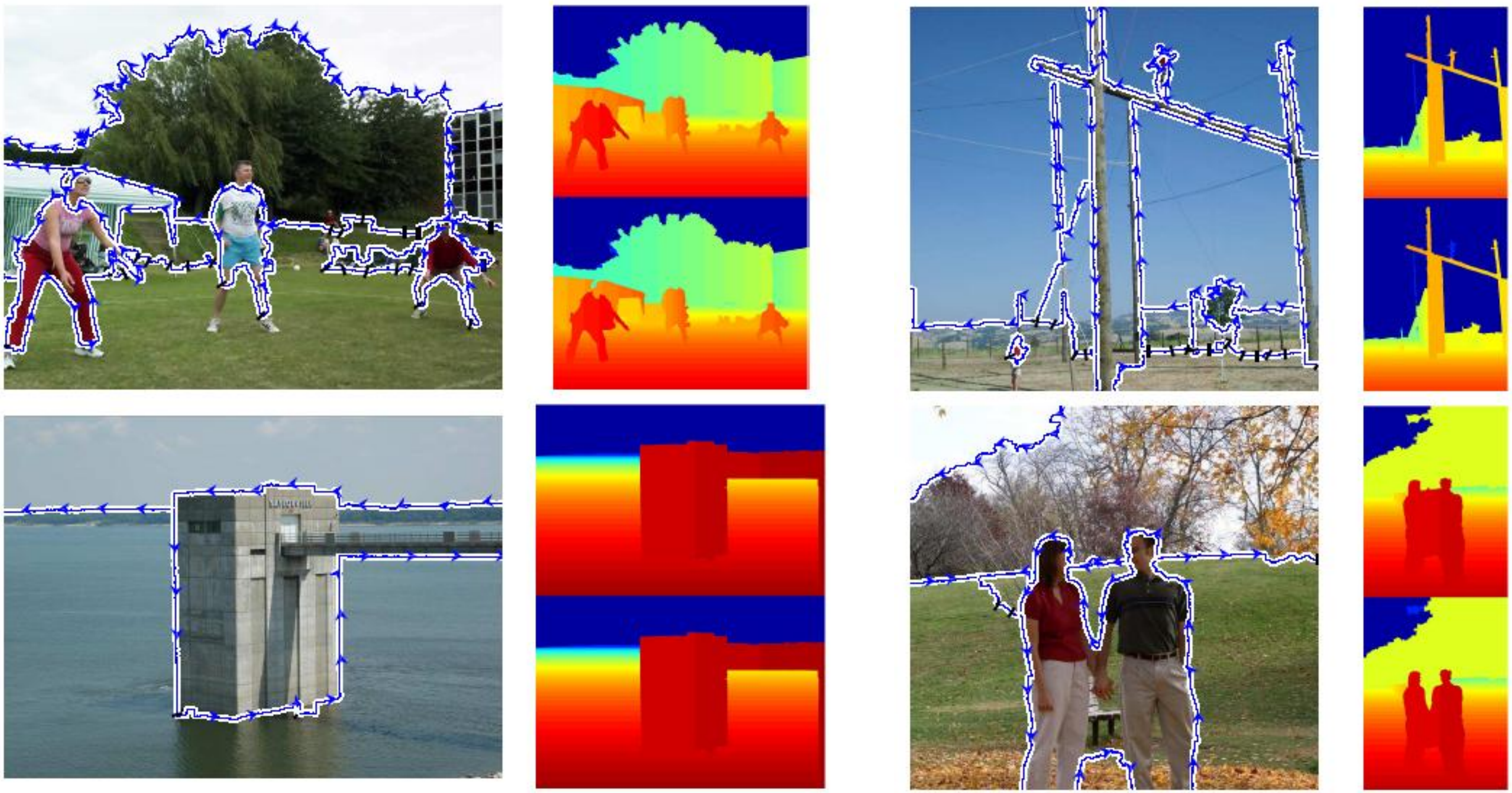


Failures

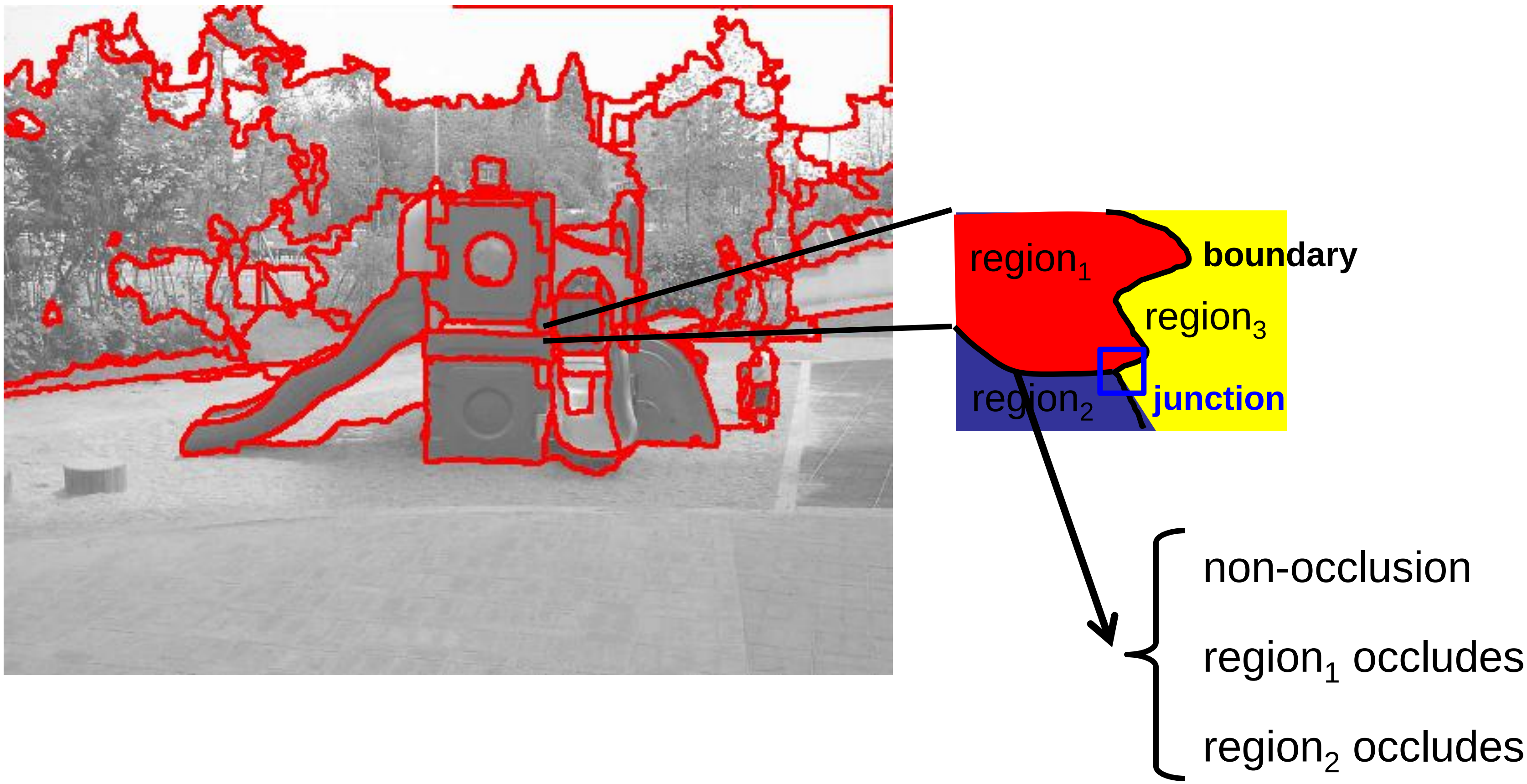


Occlusions are everywhere!

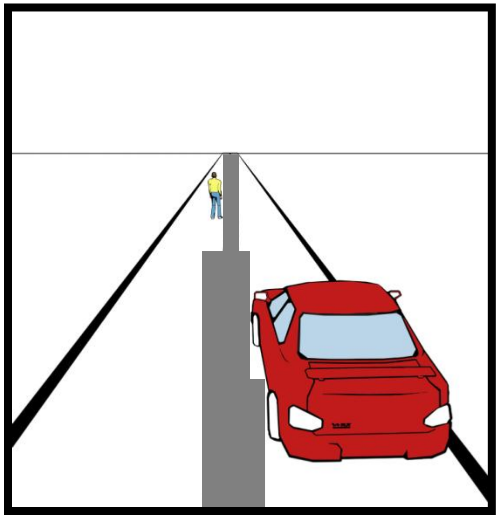
Finding occlusions
(Hoiem et al, ICCV'07)



Occlusion Reasoning as Classification



Object Size / Camera Viewpoint



Image

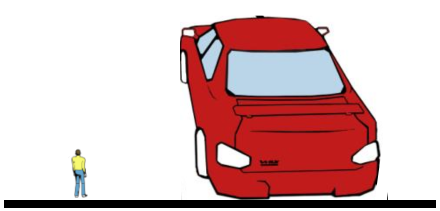
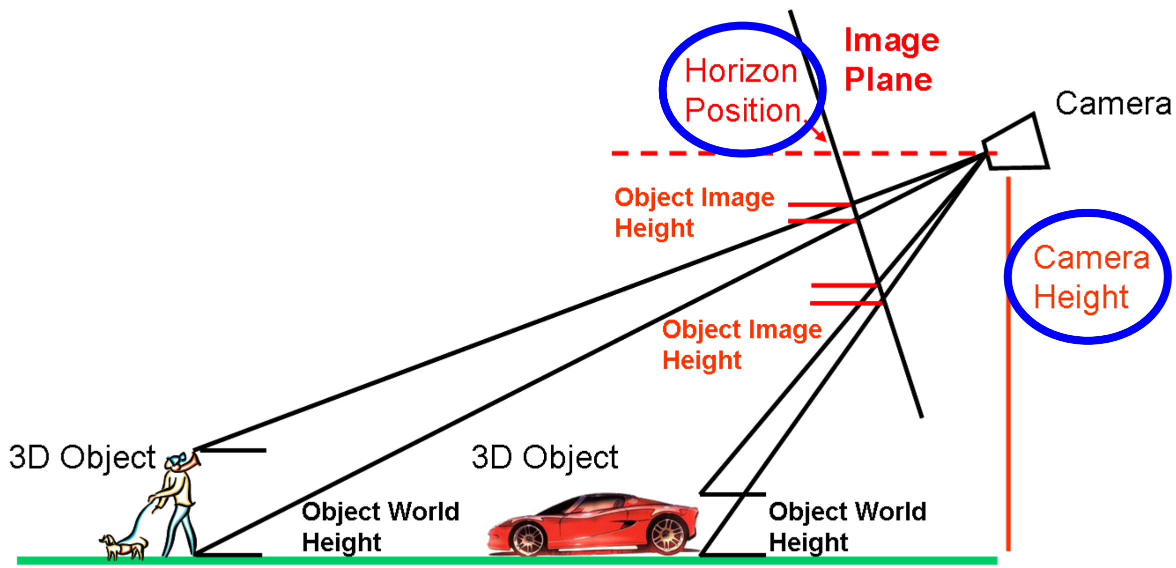


Image Coordinates

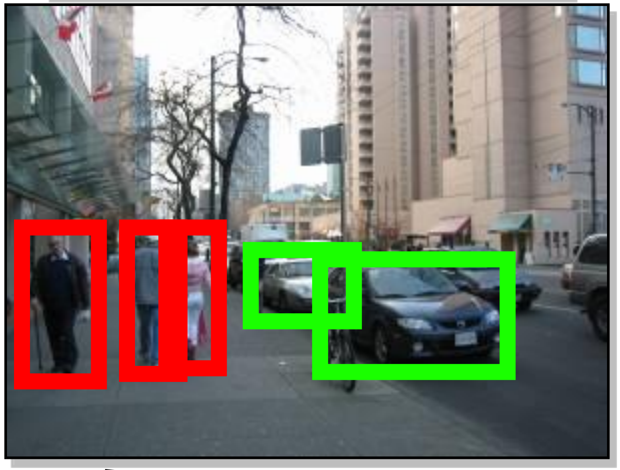


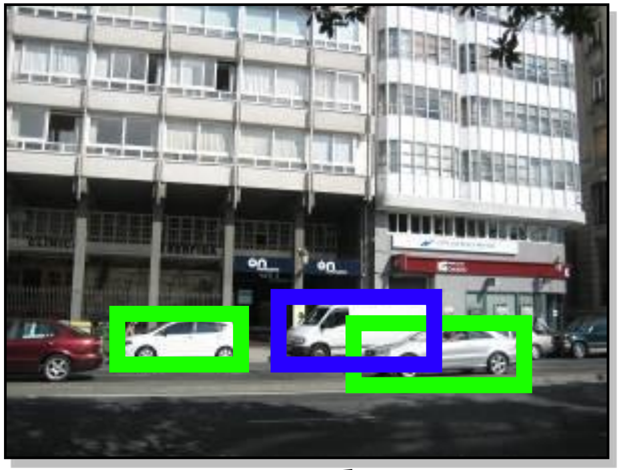
World Coordinates

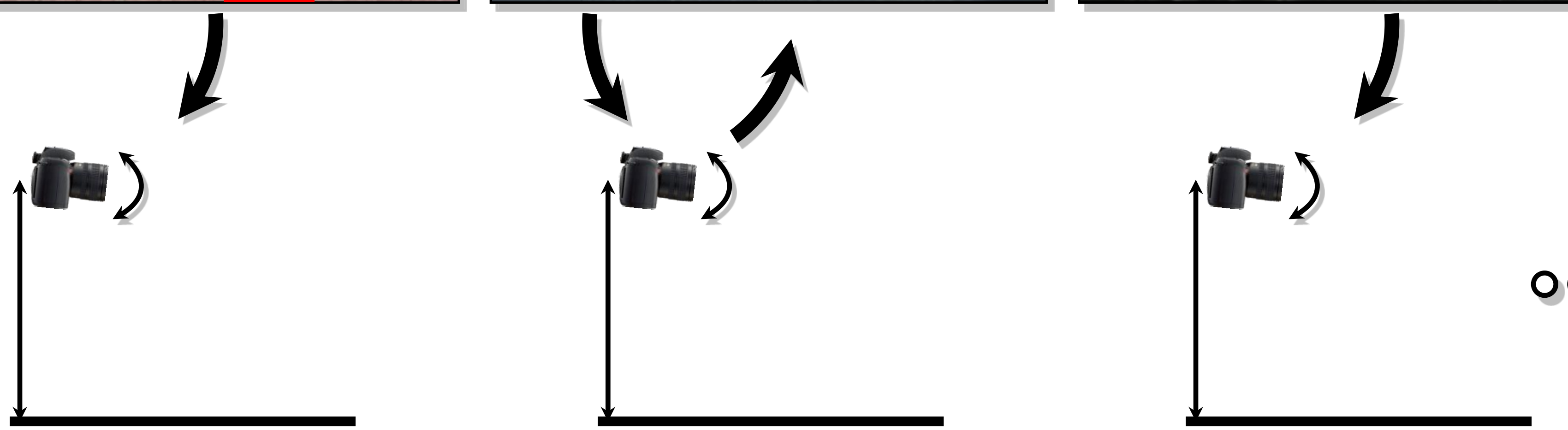
Camera viewpoint for LabelMe

Human height distribution
1.7 +/- 0.085 m
(National Center for Health Statistics)

Car height distribution
1.5 +/- 0.19 m
(automatically learned)

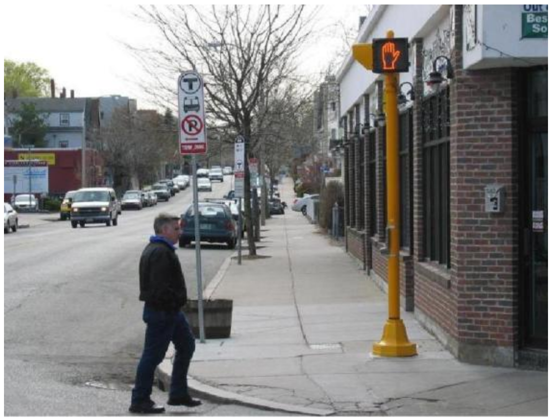






…

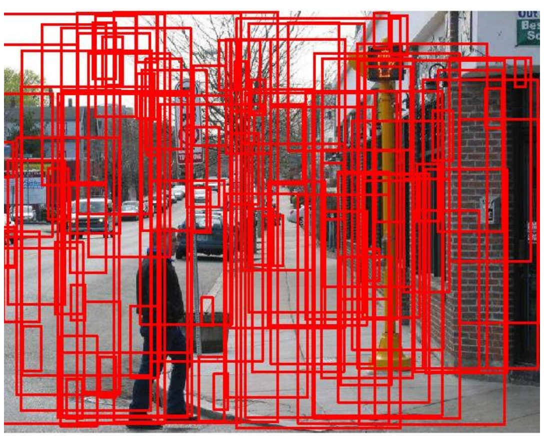
Helping Object Detection



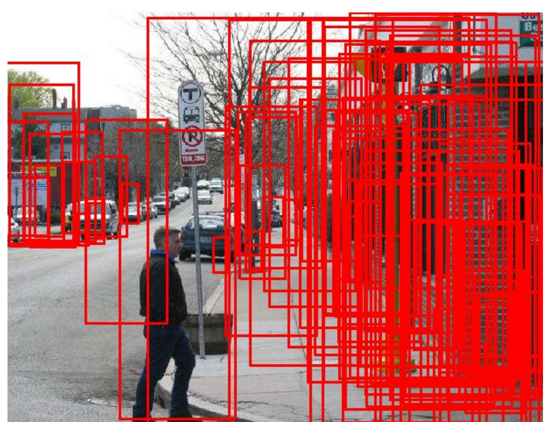
Image

P(surfaces)

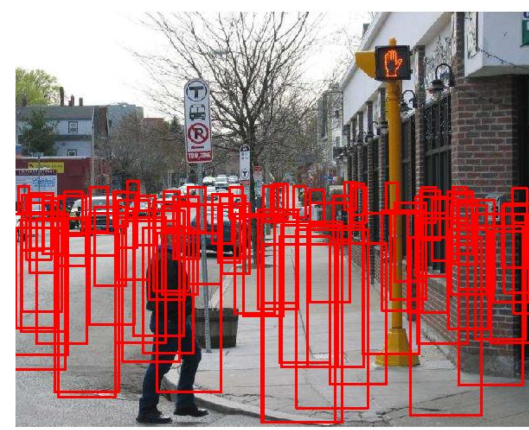
P(viewpoint)



P(object)

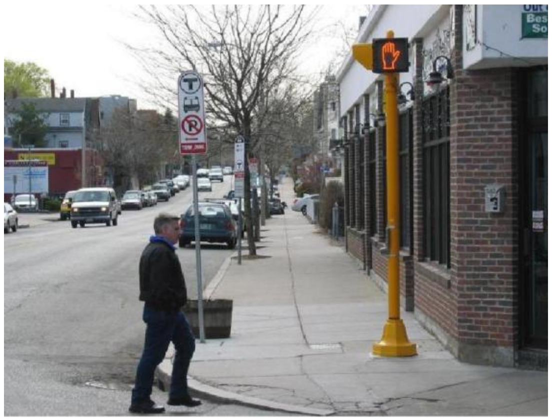


P(object | surfaces)



P(object | viewpoint)

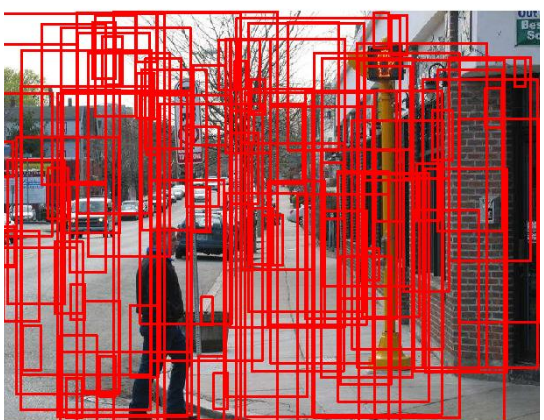
Helping Object Detection



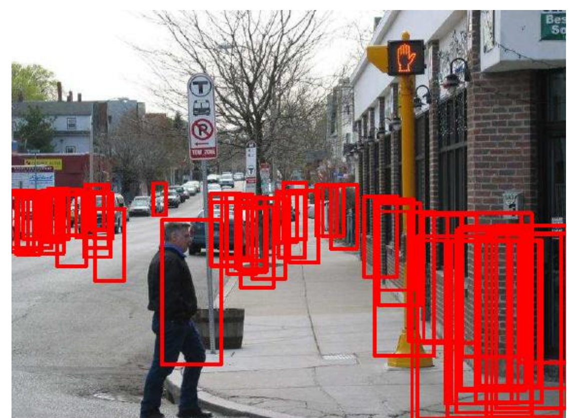
Image

P(surfaces)

P(viewpoint)

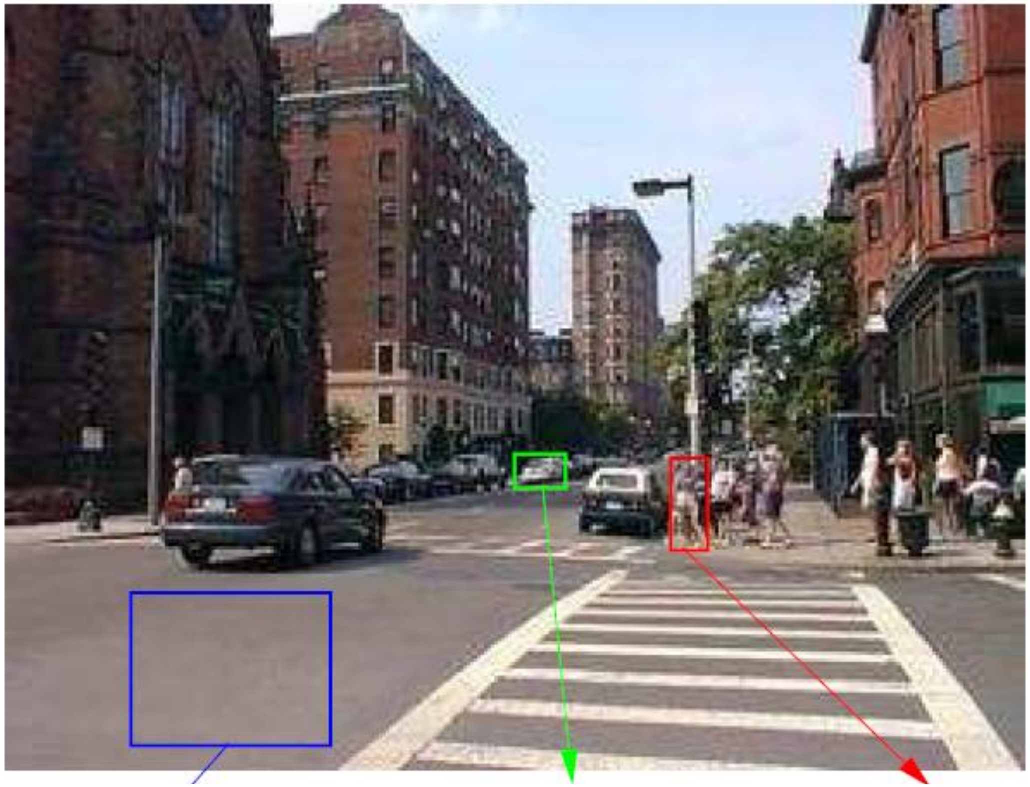


P(object)



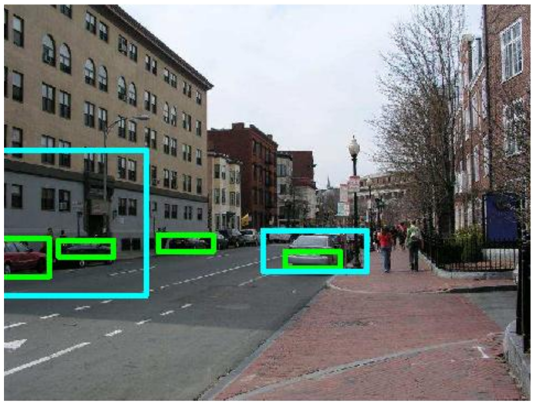
P(object | surfaces, viewpoint)

More Chickens, More Eggs…

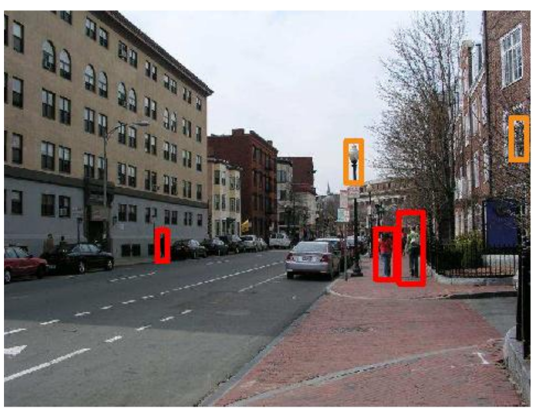


Best Guesses

Object Detection



Local Car Detector

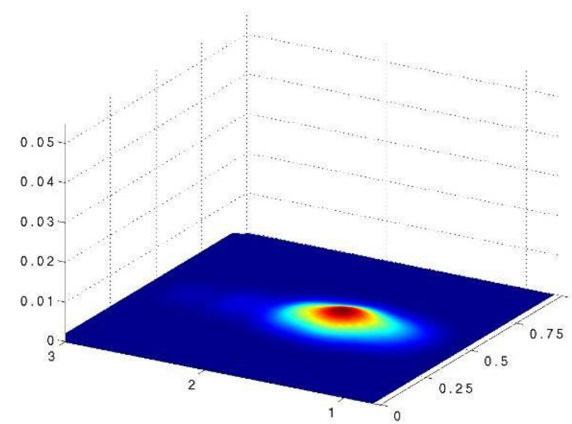


Local Ped Detector

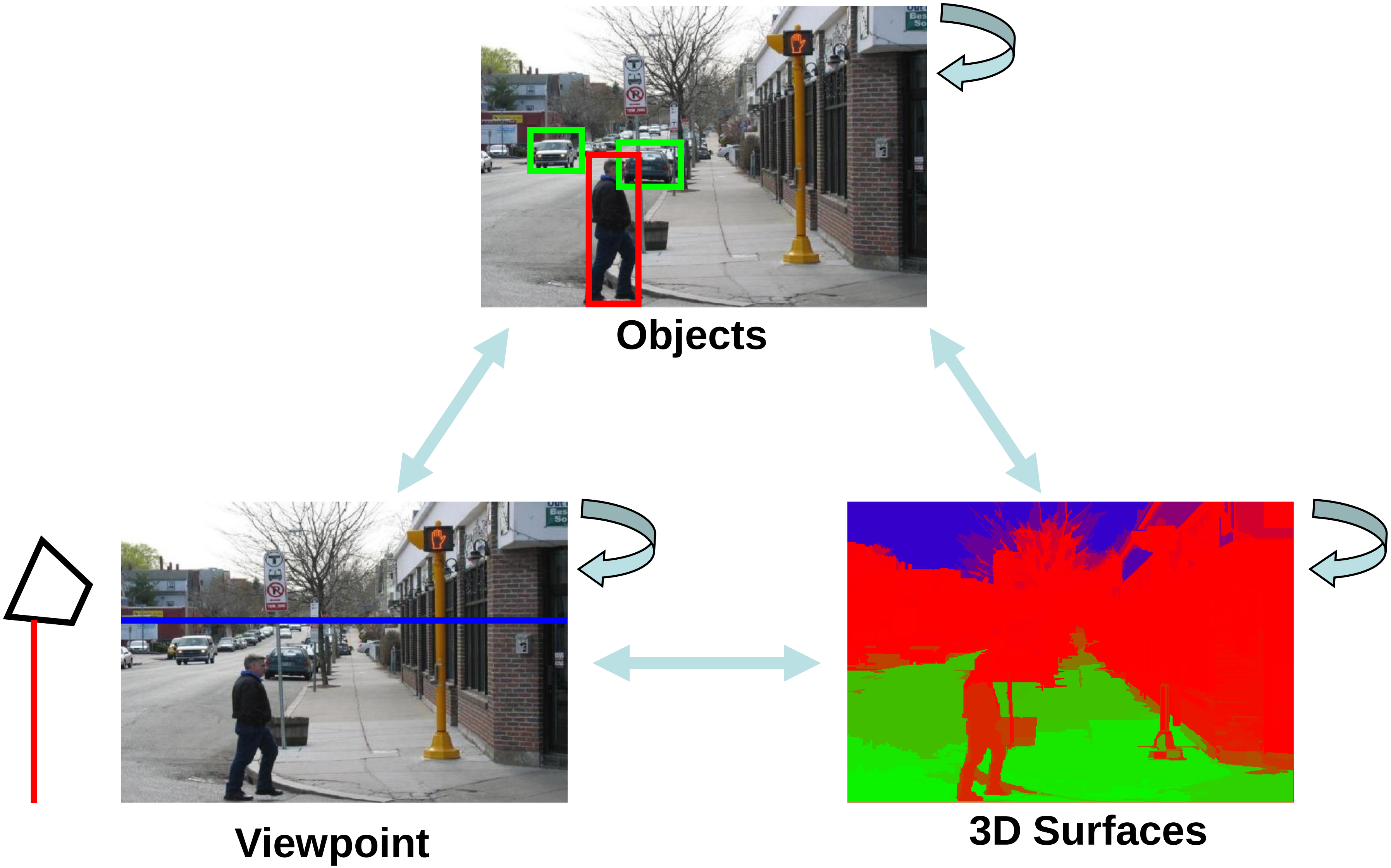
Surface Estimates

Surfaces

Viewpoint Prior

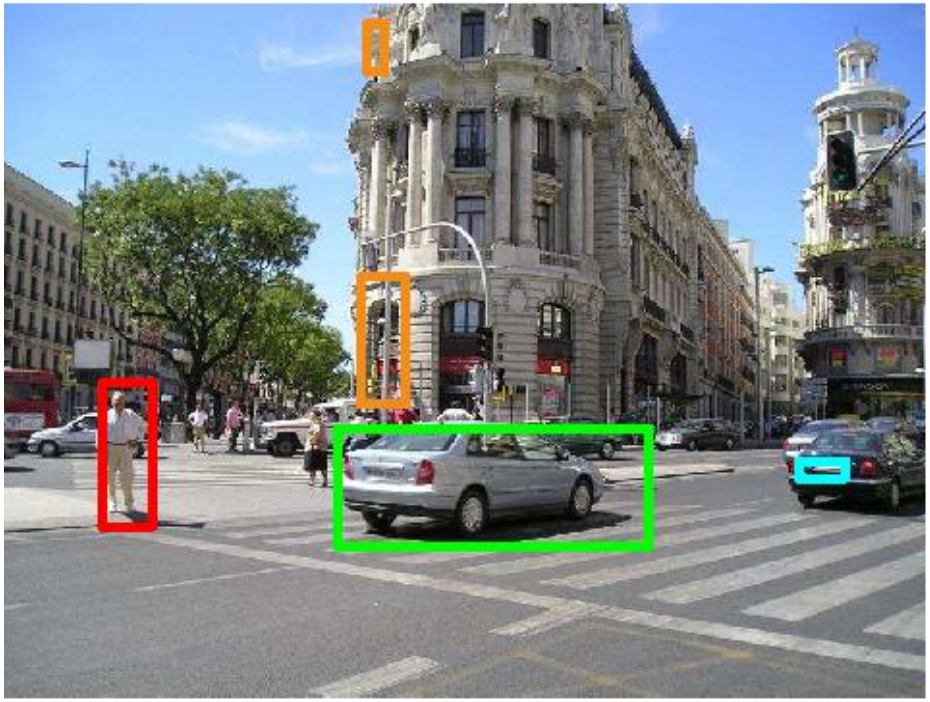


Putting it all together

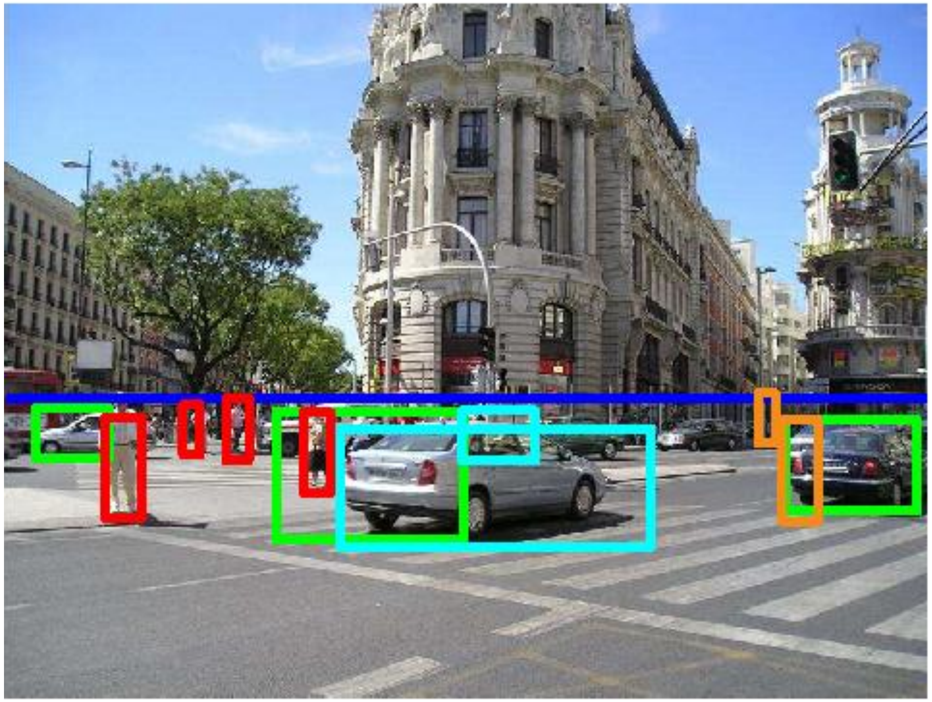


Some Results

Car: TP / FP Ped: TP / FP



Initial: 2 TP / 3 FP

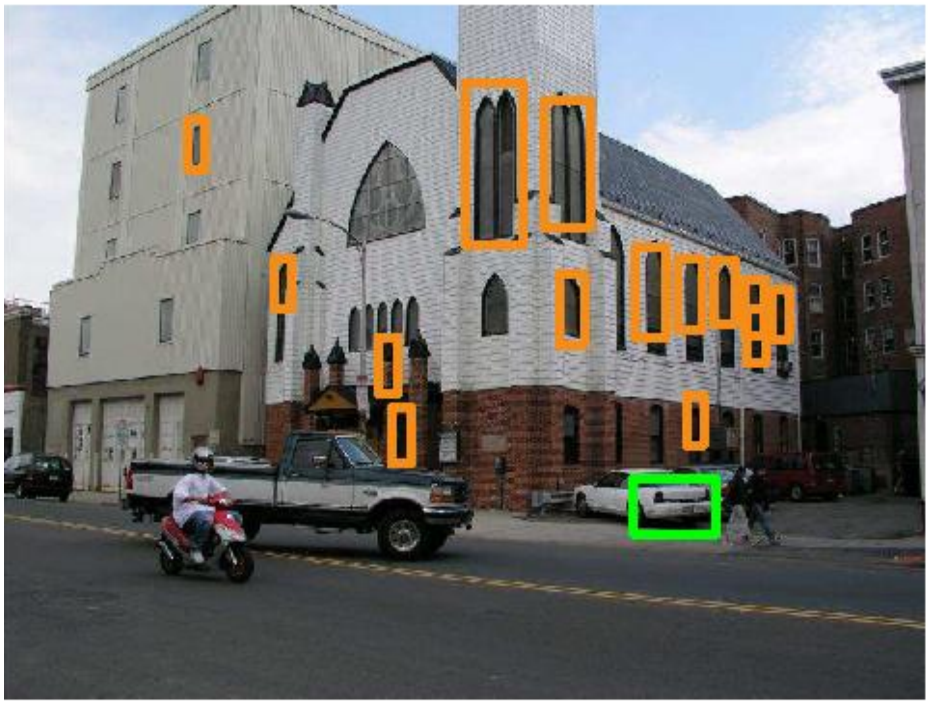


Final: 7 TP / 4 FP

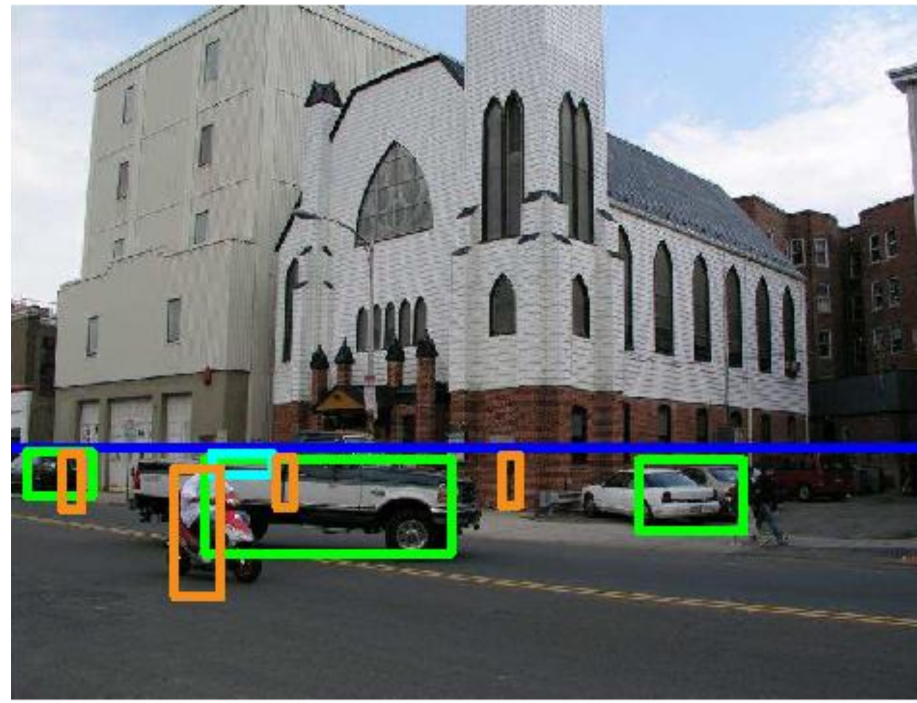
Local Detector from [Murphy-Torralba-Freeman 2003]

Some Results

Car: TP / FP Ped: TP / FP



Initial: 1 TP / 14 FP



Final: 3 TP / 5 FP

Local Detector from [Murphy-Torralba-Freeman 2003]

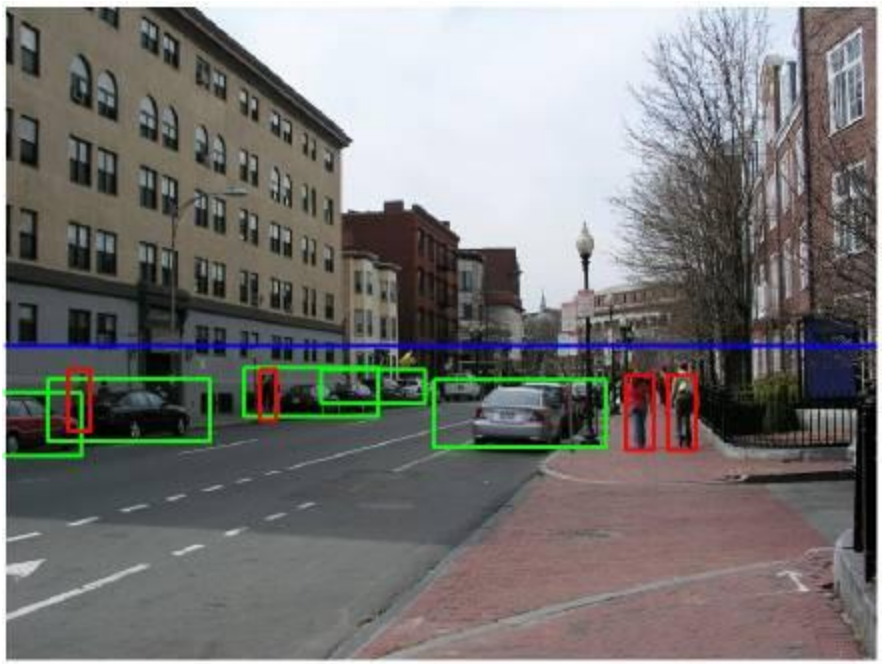
More Results

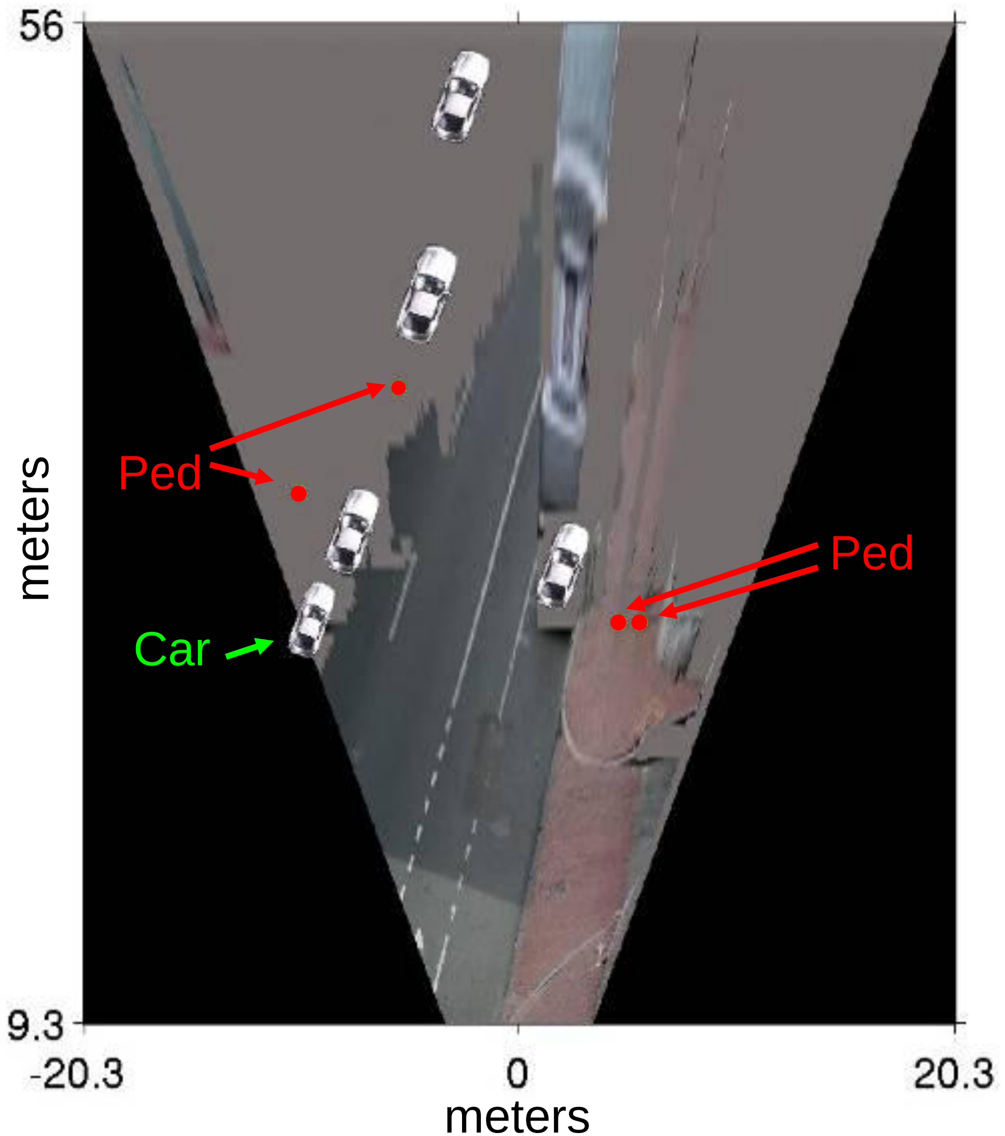
Car: TP / FP Ped: TP / FP

Initial: 1 TP / 5 FP

Final: 5 TP / 2 FP

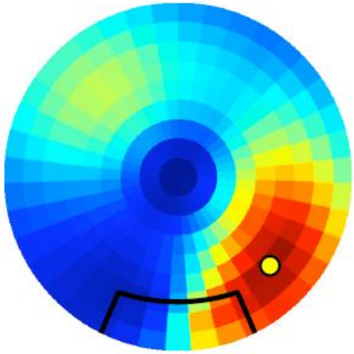
Putting Objects in Perspective





Illumination from a Single Image

Illumination from a Single Image



Lalonde. Efros, Narshimhan
ICCV'09

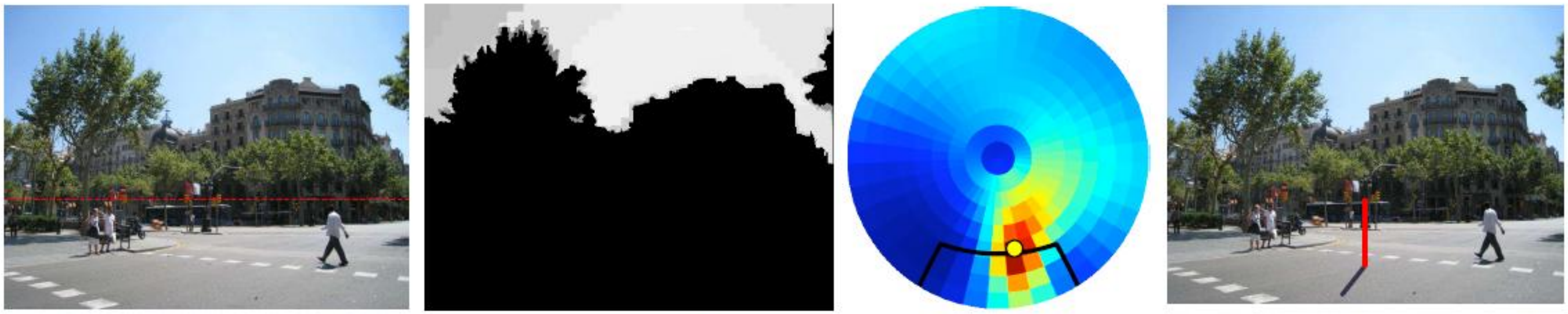
Illumination from a Single Image

Synthetic Object Insertion

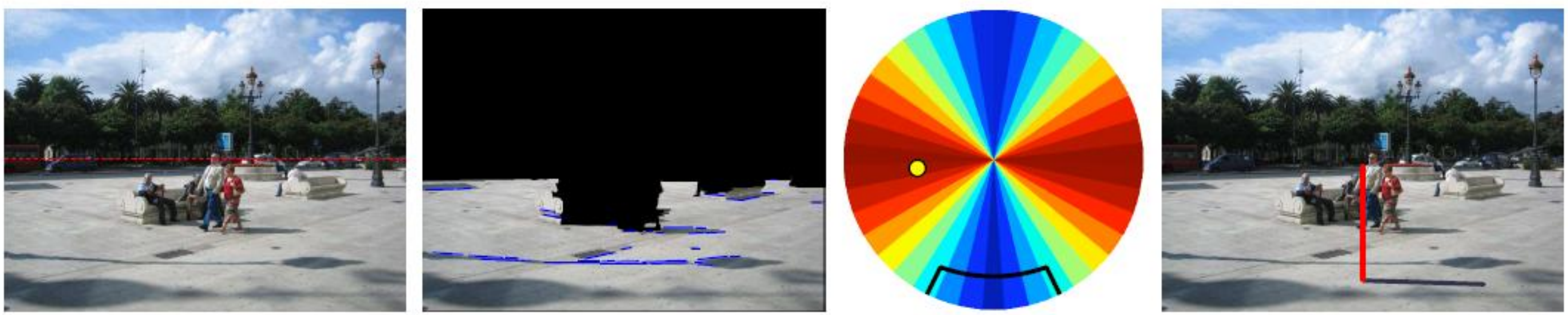
Algorithm

- Step 1: use weak cues considering 1) Sky, 2) Shadows, 3) Shading
- Step 2: Integrated them with a data-driven prior (6 million Geo-tagged images)
- Step 3: Hope for the best!!

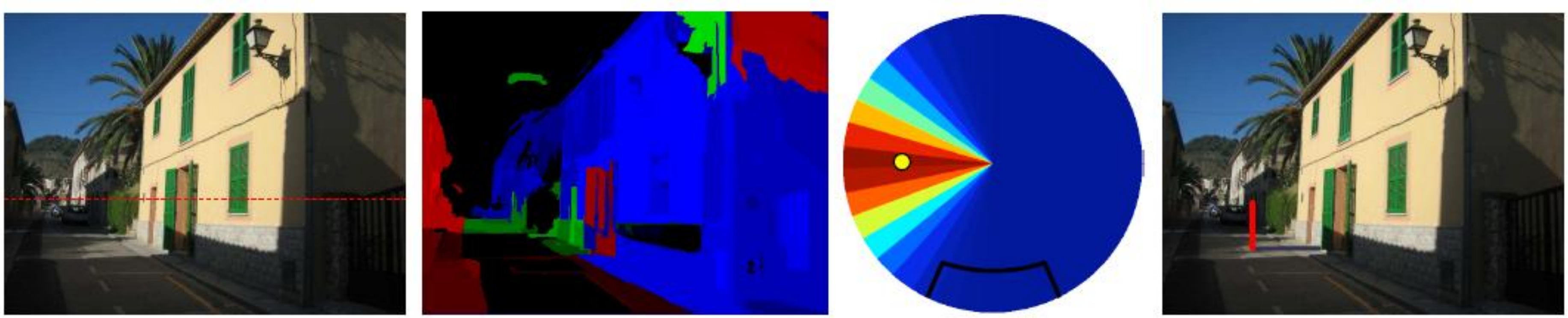
Weak cues



(a) Image and estimated horizon (b) Sky mask [9] (c) $P(\theta_s, \Delta\phi_s|\mathcal{S})$ (d) Inserted sun dial

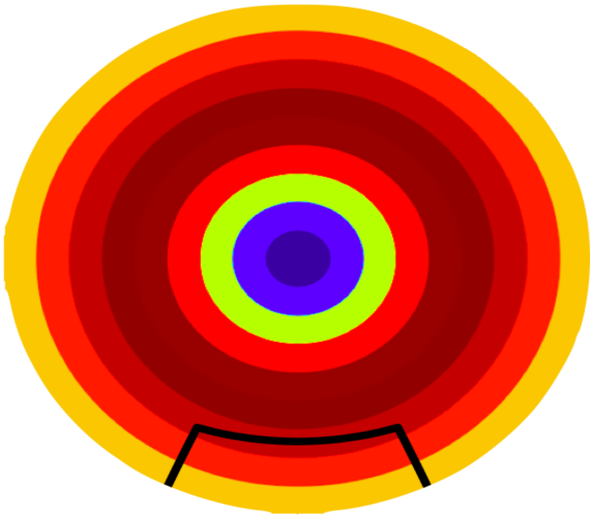


(a) Image and estimated horizon (b) Ground mask [9] and shadow lines (c) $P(\theta_s, \Delta\phi_s|\mathcal{G})$ (d) Inserted sun dial



(a) Image and estimated horizon (b) Vertical mask [9] (c) $P(\theta_s, \Delta\phi_s|\mathcal{V})$ (d) Inserted sun dial

Data-driven Sun Elevation Prior

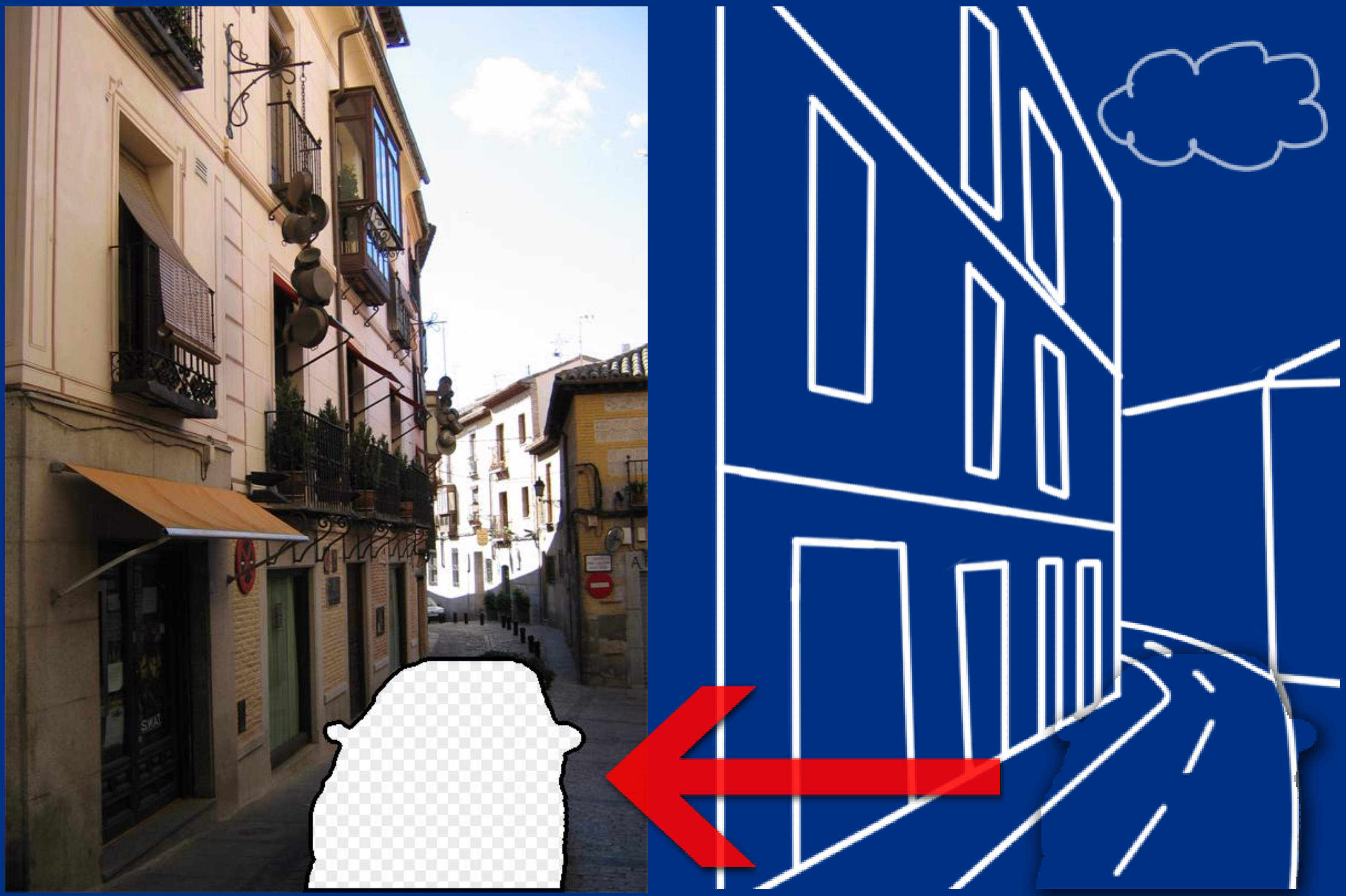


Scene Semantics:
Understanding the Entire Scene

Hays & Efros, SIGGRAPH'07

Where does the knowledge come from?

Scene Semantics!



Google

alley

Search Images | Search the Web

Advanced Image Search
Preferences

Strict SafeSearch is on

Images Showing: All image sizes

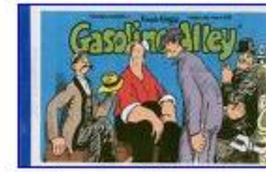
Results **1** - **20** of about **908,000** for **alley** [definition] with **Safesearch on**. (**0.07** seconds)

Change **Alley** Aerial Plaza with its ...
300 x 400 - 21k
en.wikipedia.org

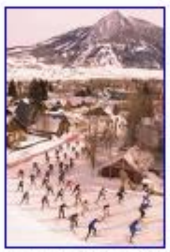
The Printer's **Alley** sign looking ...
679 x 450 - 469k - jpg
franklin.thefuntimesguide.com

Looking west past Printers **Alley**.
679 x 450 - 464k - jpg
franklin.thefuntimesguide.com

More Bubble Gum **Alley** photos can be ...
764 x 591 - 33k - gif
www.locallinks.com



Gasoline **Alley** gang
692 x 430 - 177k - jpg
newcritics.com



2007 **Alley** Loop Sponsors
300 x 453 - 51k - jpg
www.cbnordic.org

Change **Alley** : interior
550 x 413 - 98k
infopedia.nlb.gov.sg

Earl G. **Alley ...**
321 x 383 - 19k - jpg
www.msstate.edu

Gun **Alley** 8.5x11 Full Color Ink Wash **...**
390 x 301 - 14k - jpg
www.rorschachentertainment.com

Grace Court **Alley**
732 x 549 - 98k - jpg
www.bridgeandtunnelclub.com

Grace Court **Alley**
732 x 549 - 80k - jpg
www.bridgeandtunnelclub.com

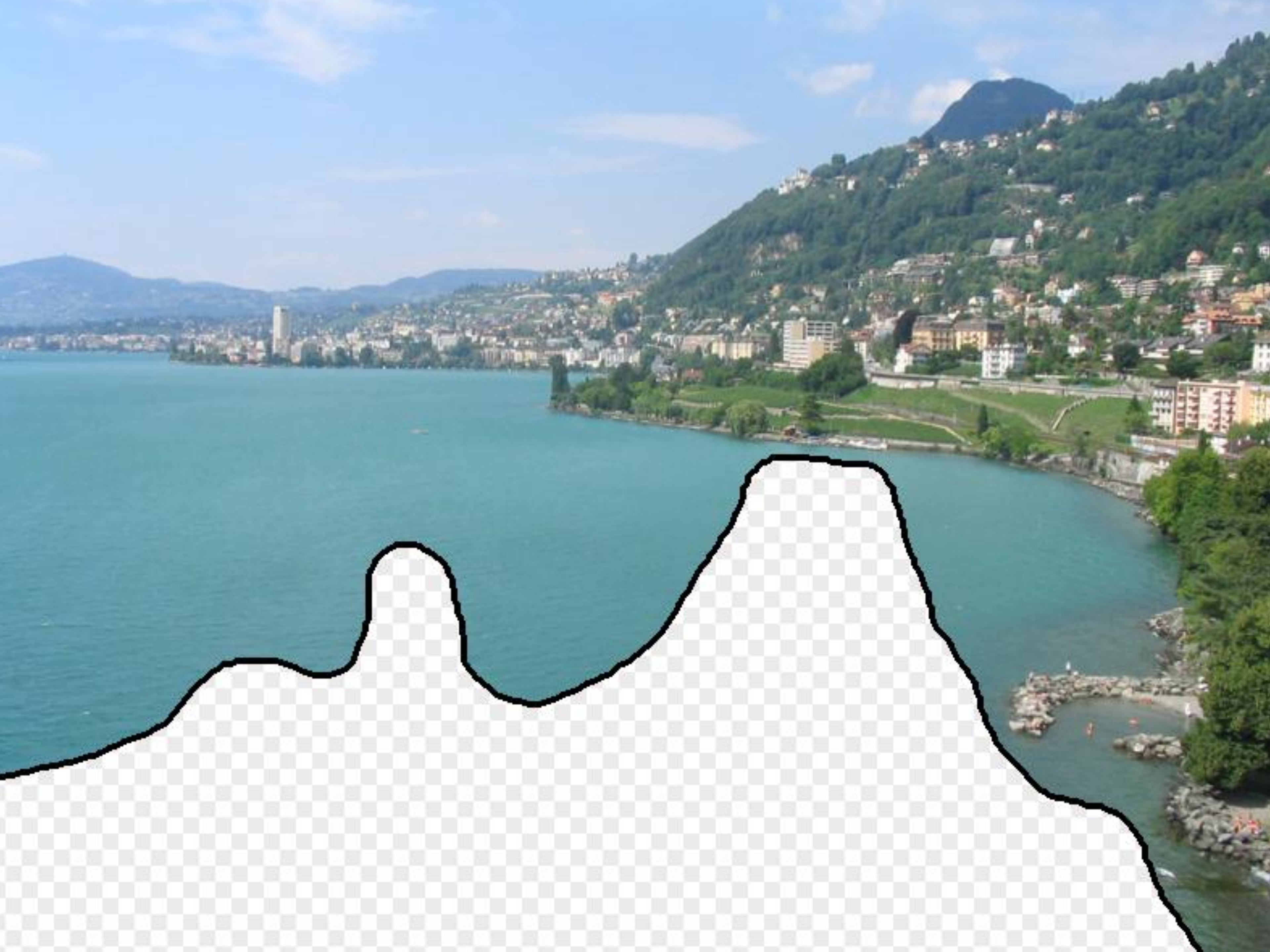
panoramic photo of Alligator **Alley**
4902 x 460 - 1048k - jpg
sflwww.er.usgs.gov

Richard B. **Alley**
450 x 361 - 29k - gif
www.ncdc.noaa.gov

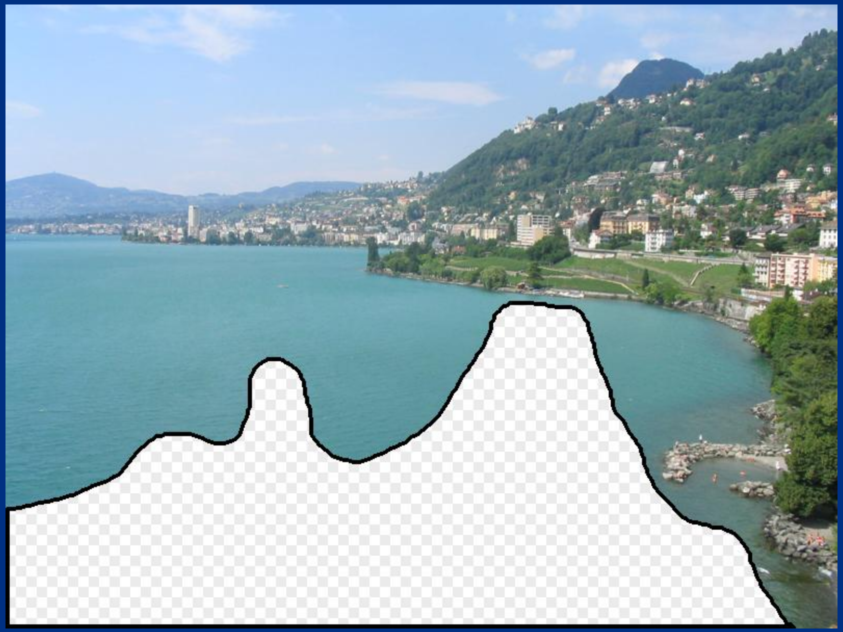
Also, Chicken **Alley** is reported to **...**
450 x 337 - 82k
phidoux.typepad.com

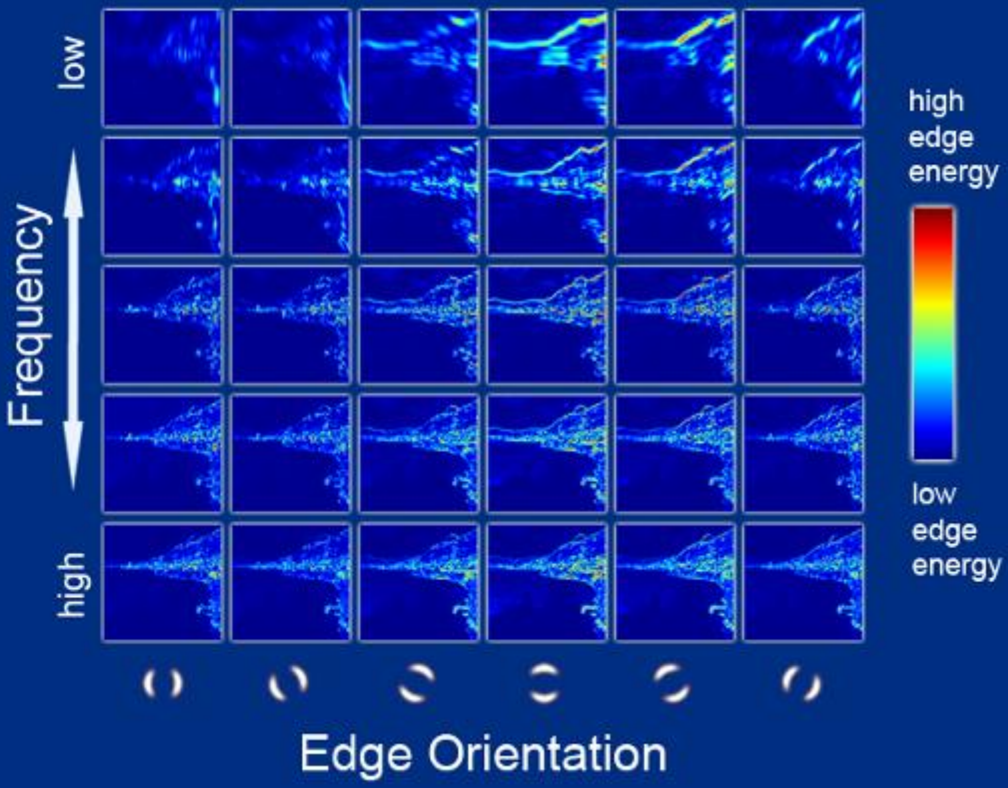
Ego **Alley**
500 x 375 - 48k - jpg
dc.about.com

Scene Completion Result

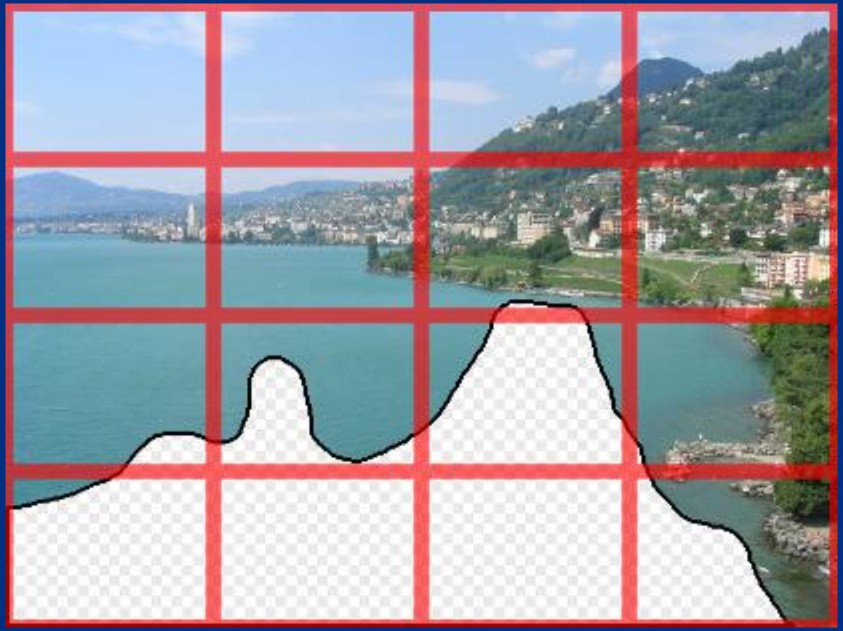


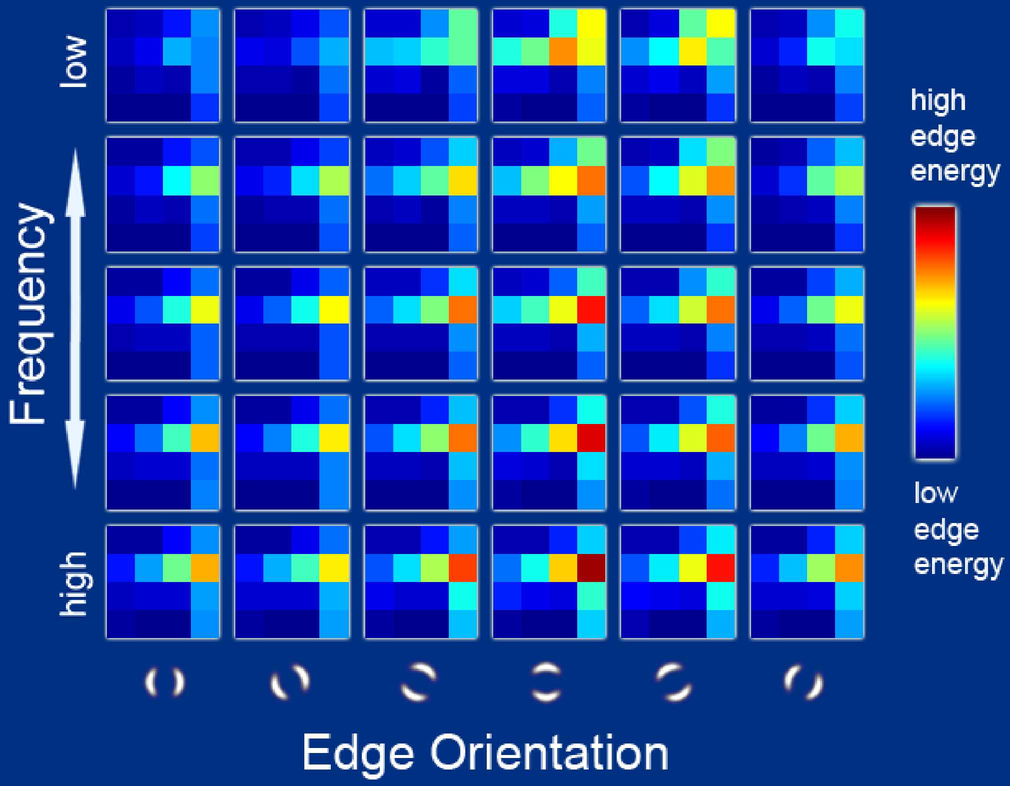
Scene Descriptor





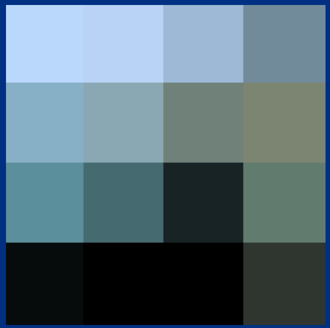
Scene Descriptor



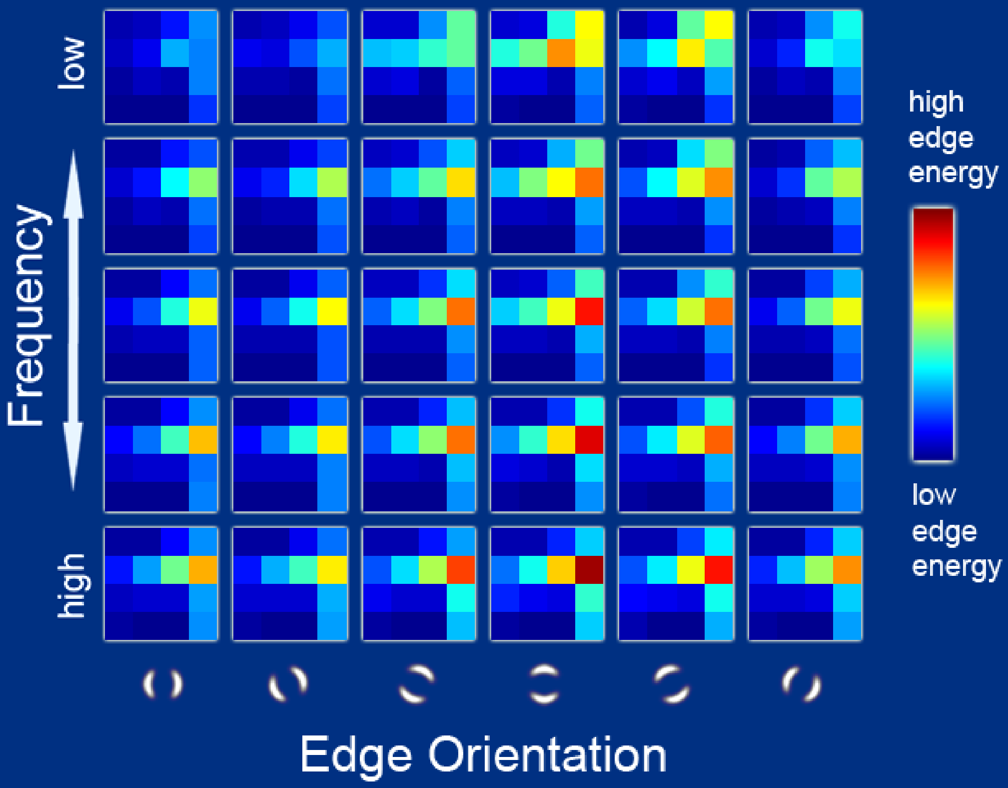


Gist scene descriptor
(Oliva and Torralba 2001)

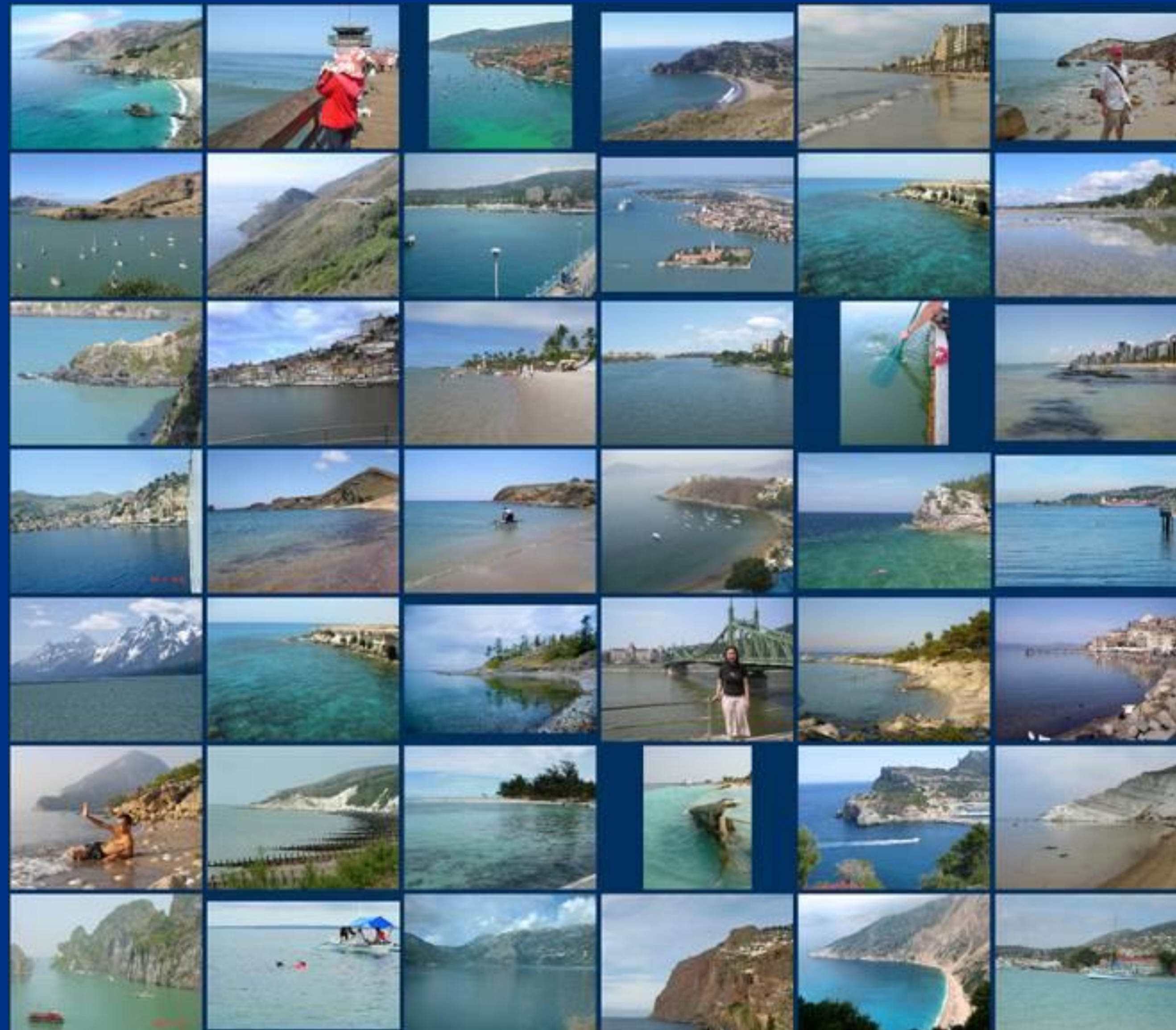
Scene Descriptor



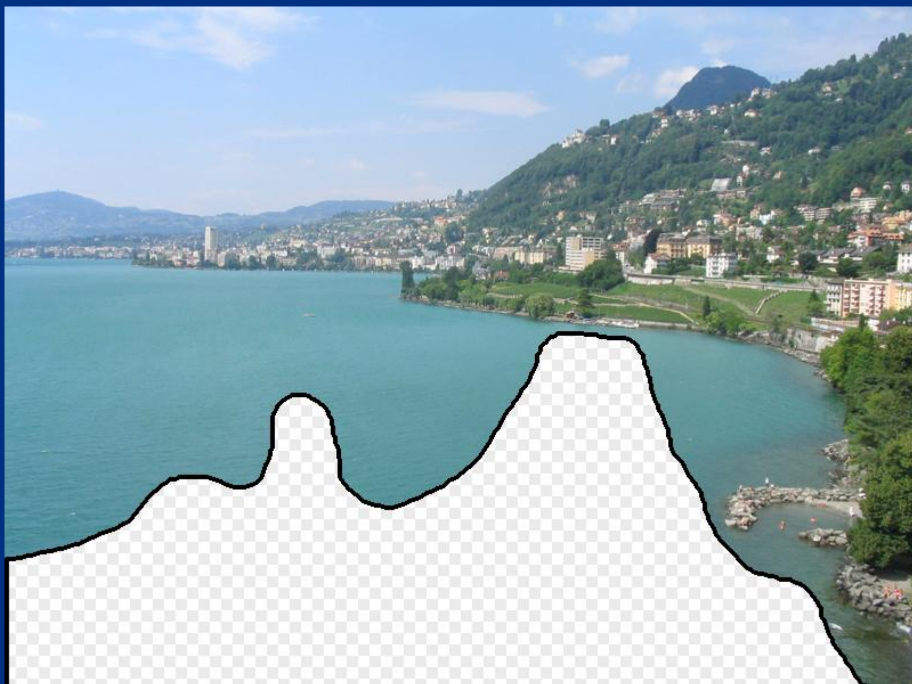
+



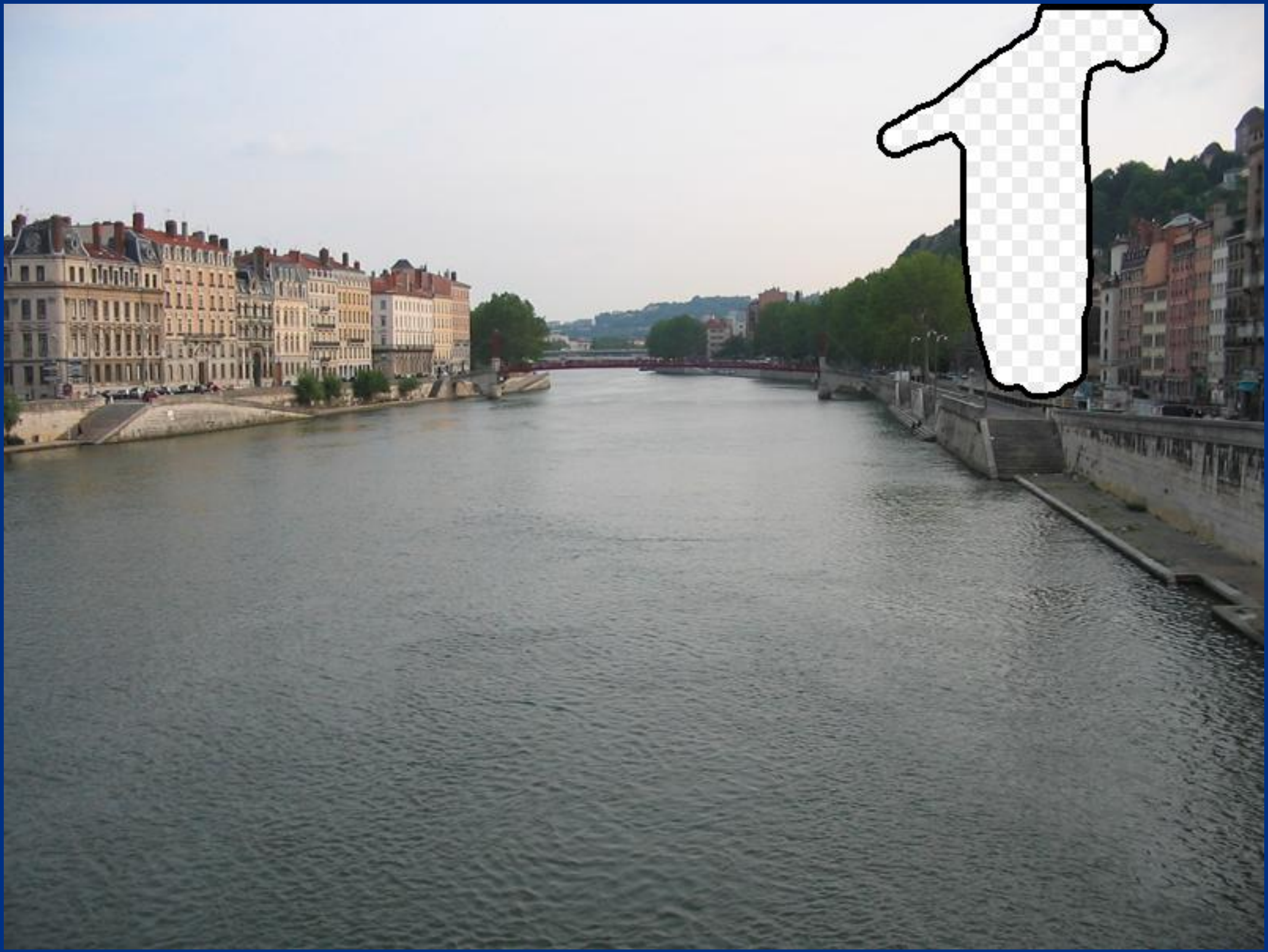
Gist scene descriptor
(Oliva and Torralba 2001)

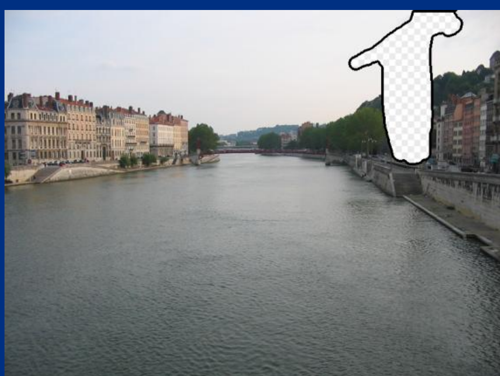


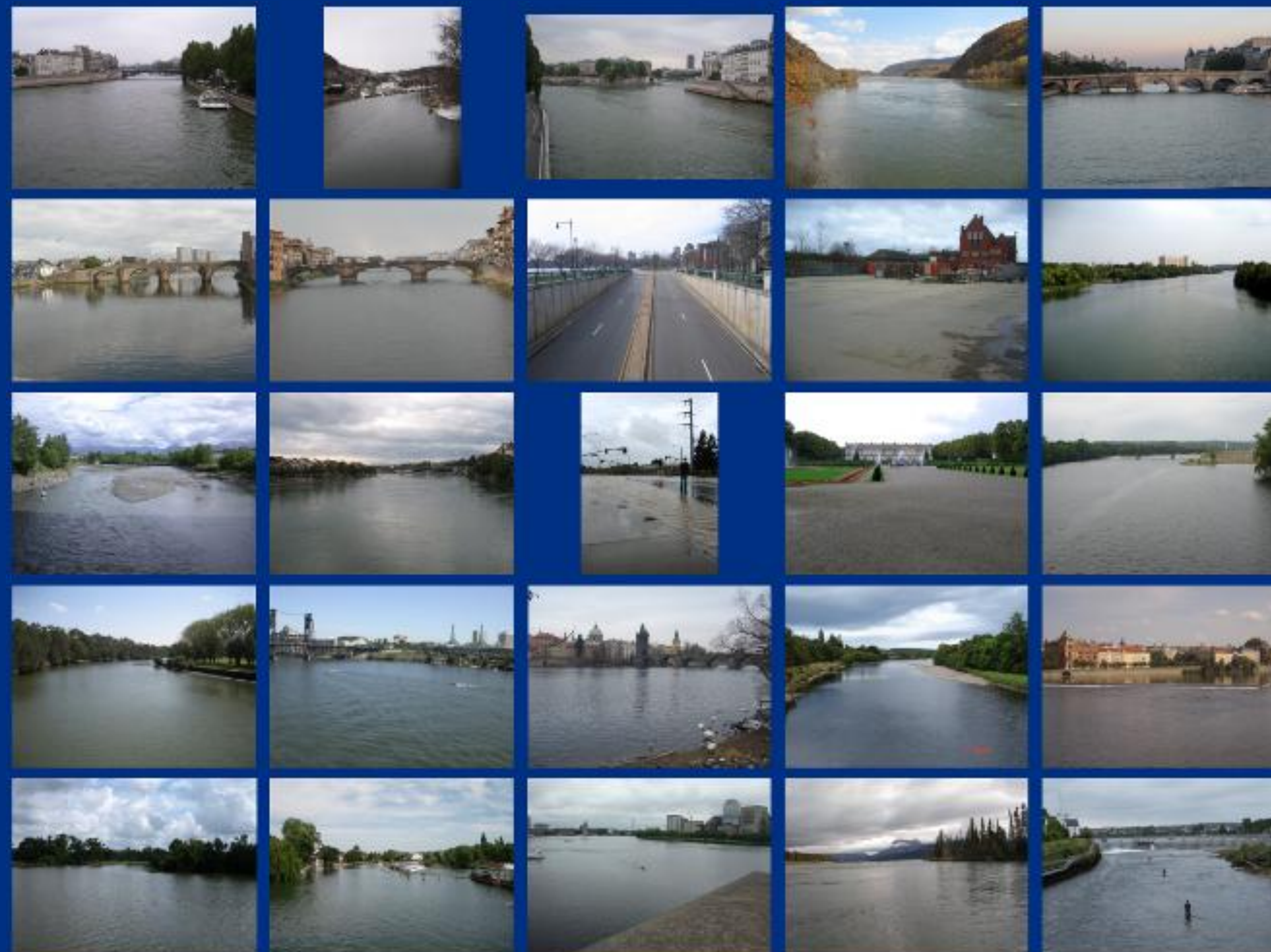
… 200 total



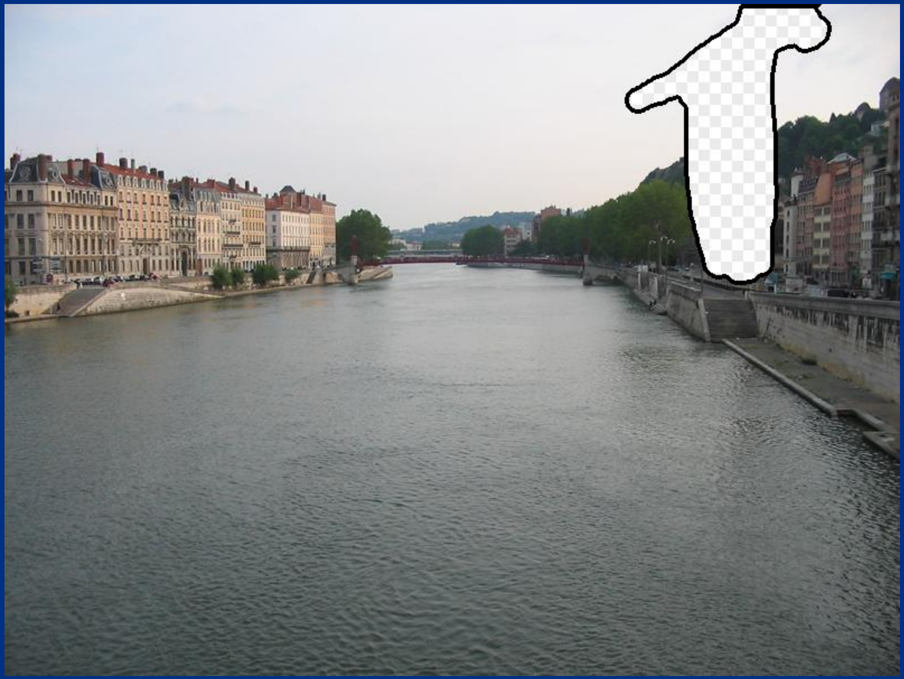
Graph cut + Poisson blending

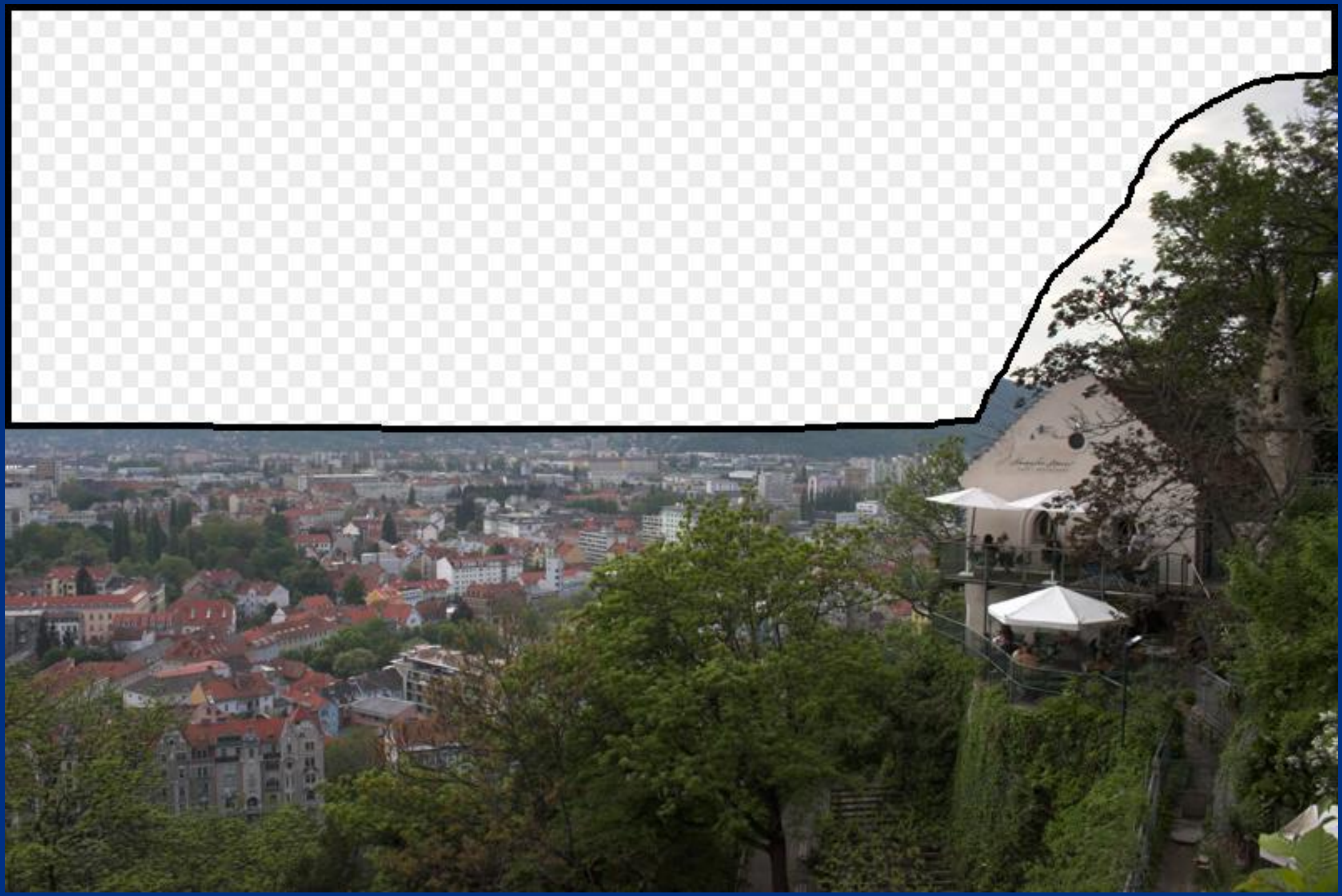


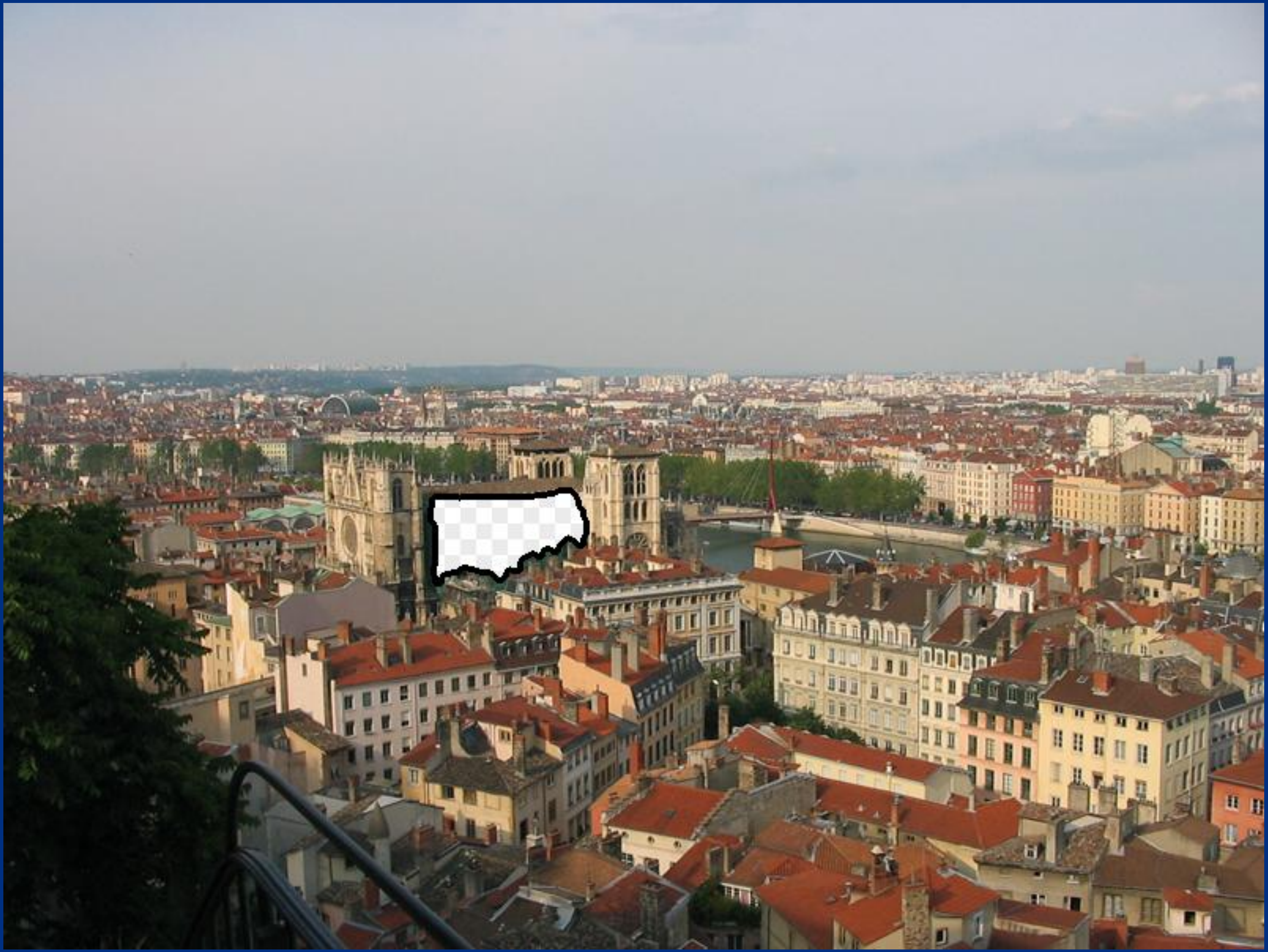


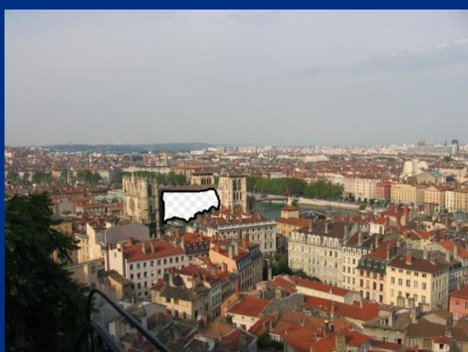


… 200 scene matches

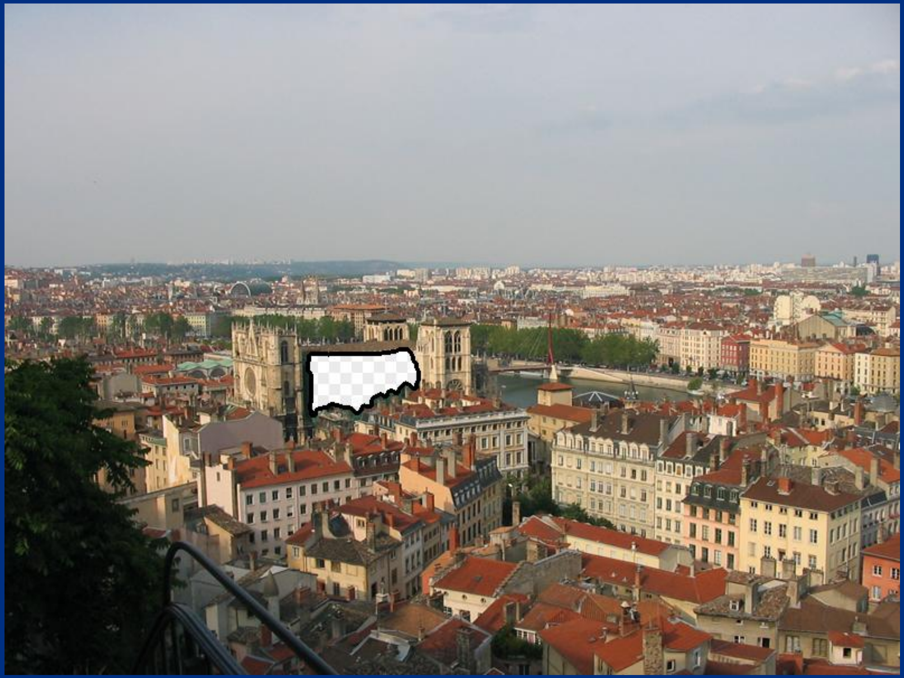


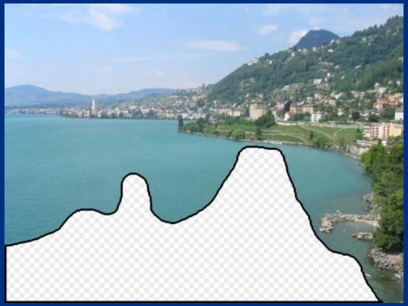




… 200 scene matches

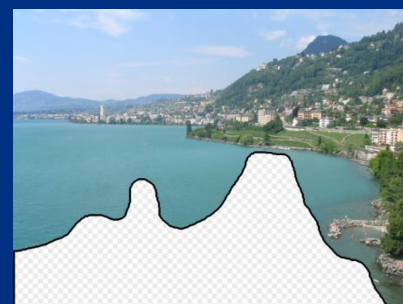


Why does it work?





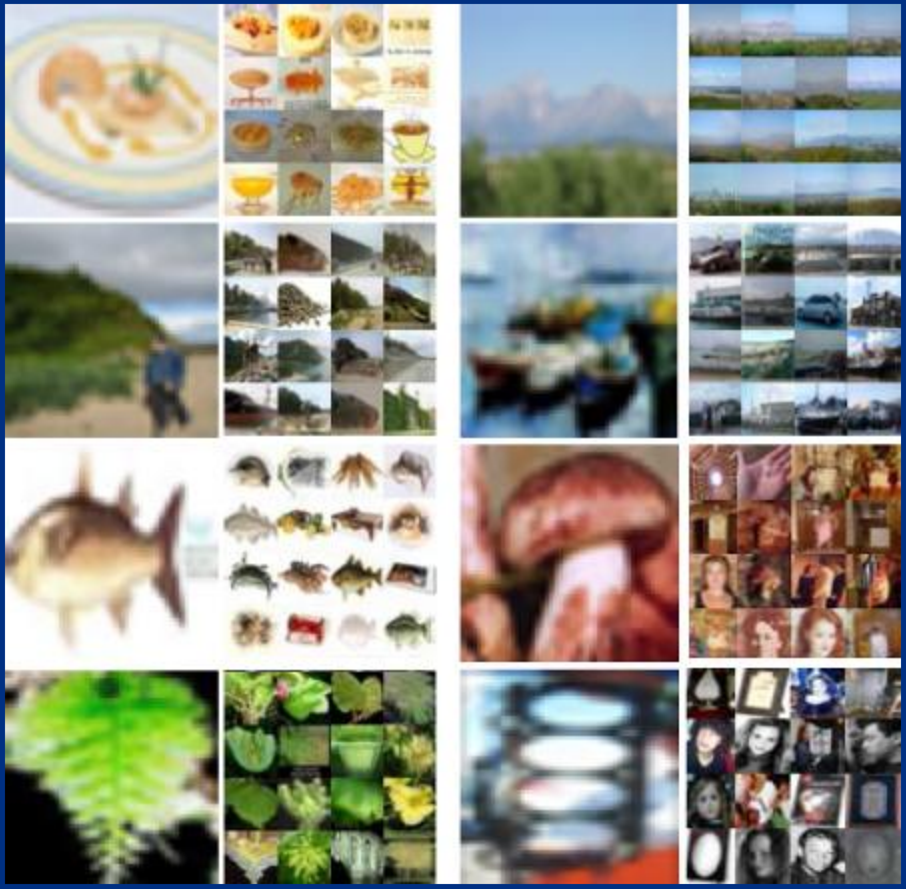





10 nearest neighbors from a collection of 20,000 images

10 nearest neighbors from a
collection of 2 million images



Database of 70 Million 32x32 images

Torralba, Fergus, and Freeman. Tiny Images. MIT-CSAIL-TR-2007-024. 2007.

The Big Picture

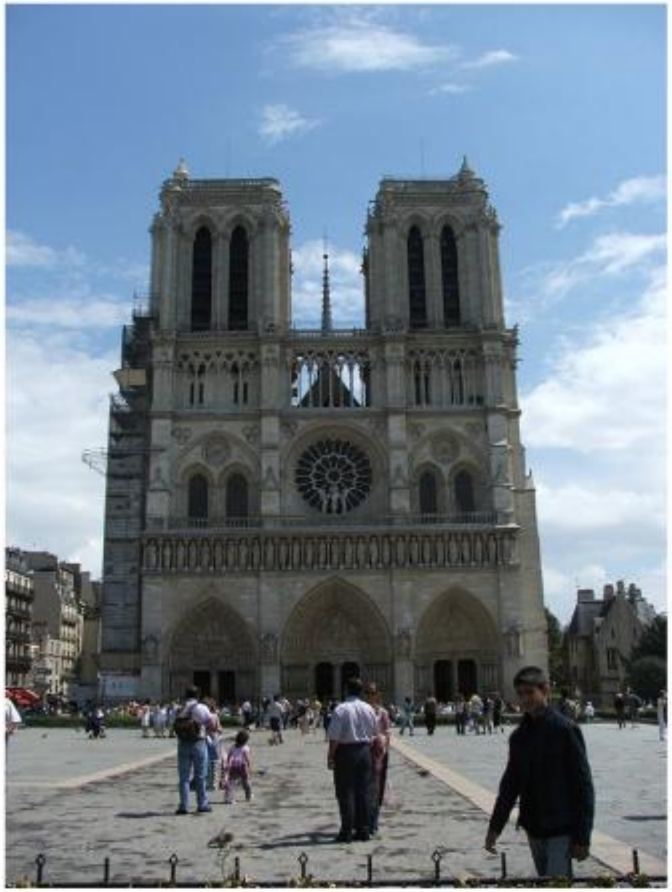
Sky, Water, Hills, Beach,
Sunny, mid-day

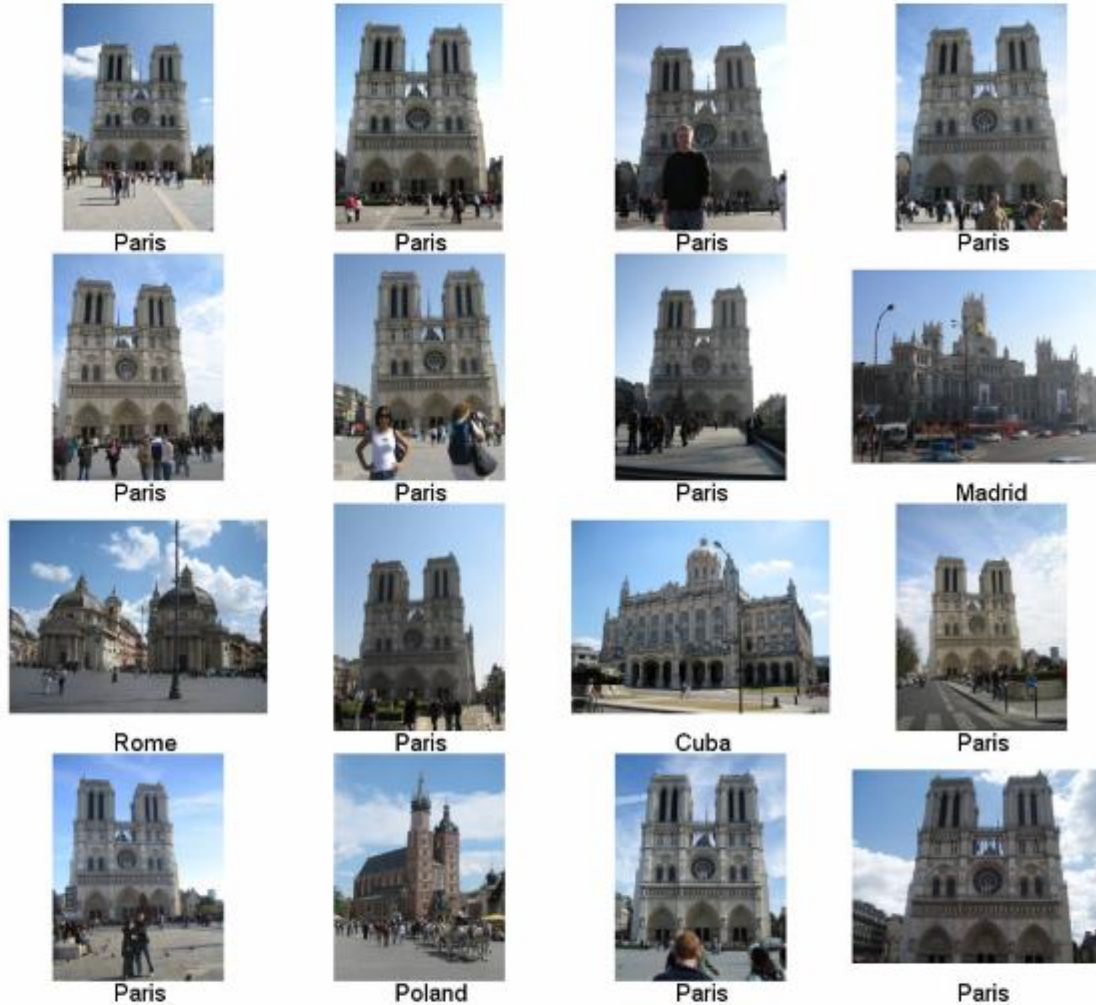
Brute-force Image Understanding

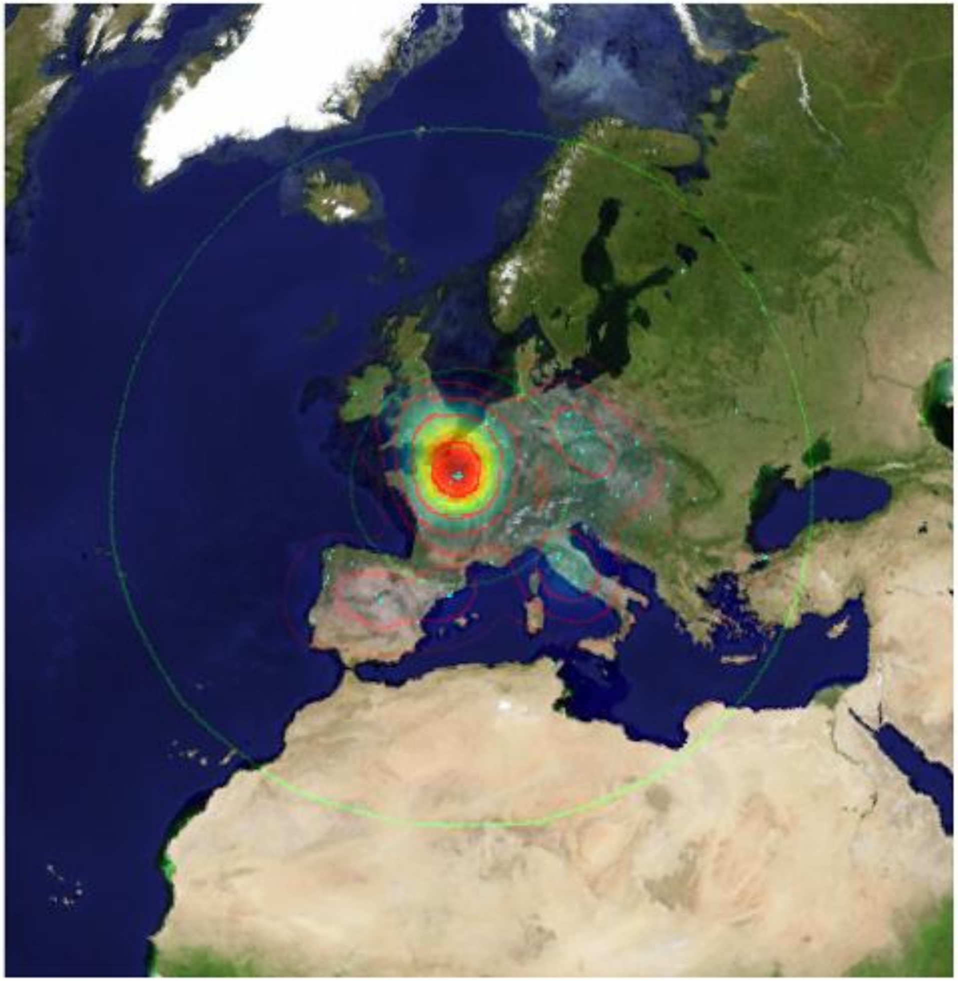
im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

How much can an image tell about its geographic location?



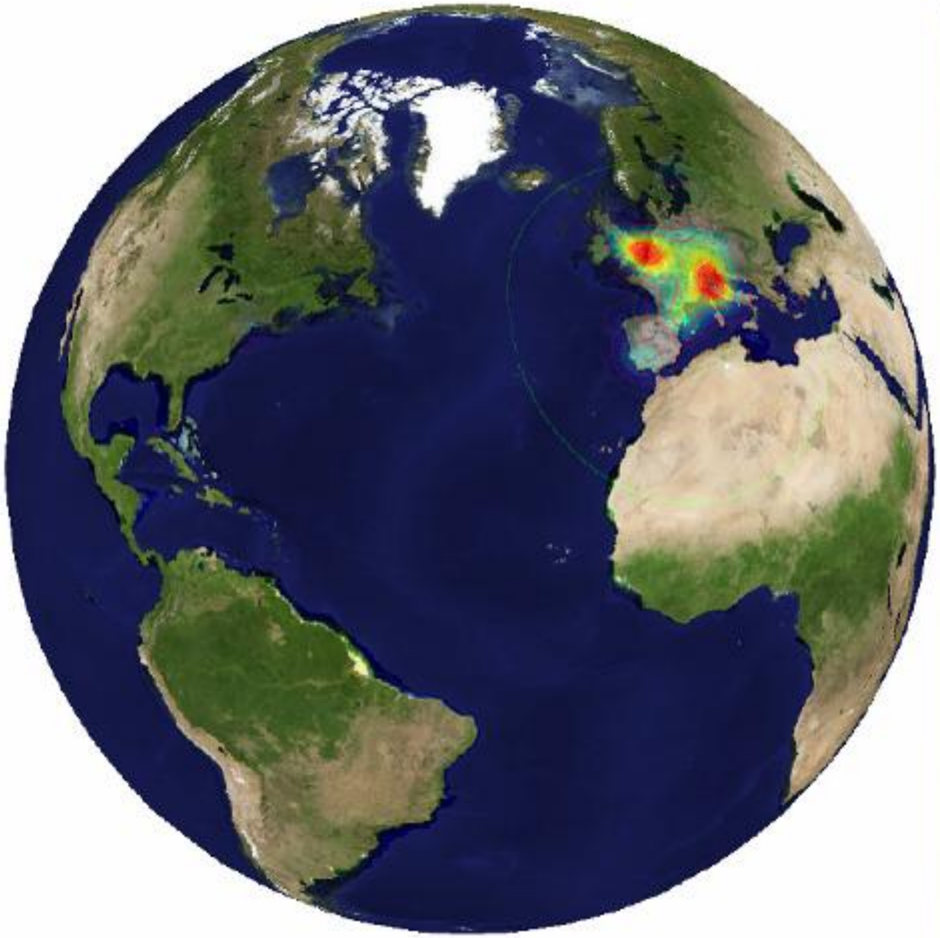

Paris
Paris
Paris
Paris
Paris
Paris
Paris
Madrid
Rome
Paris
Cuba
Paris
Paris
Poland
Paris
Paris

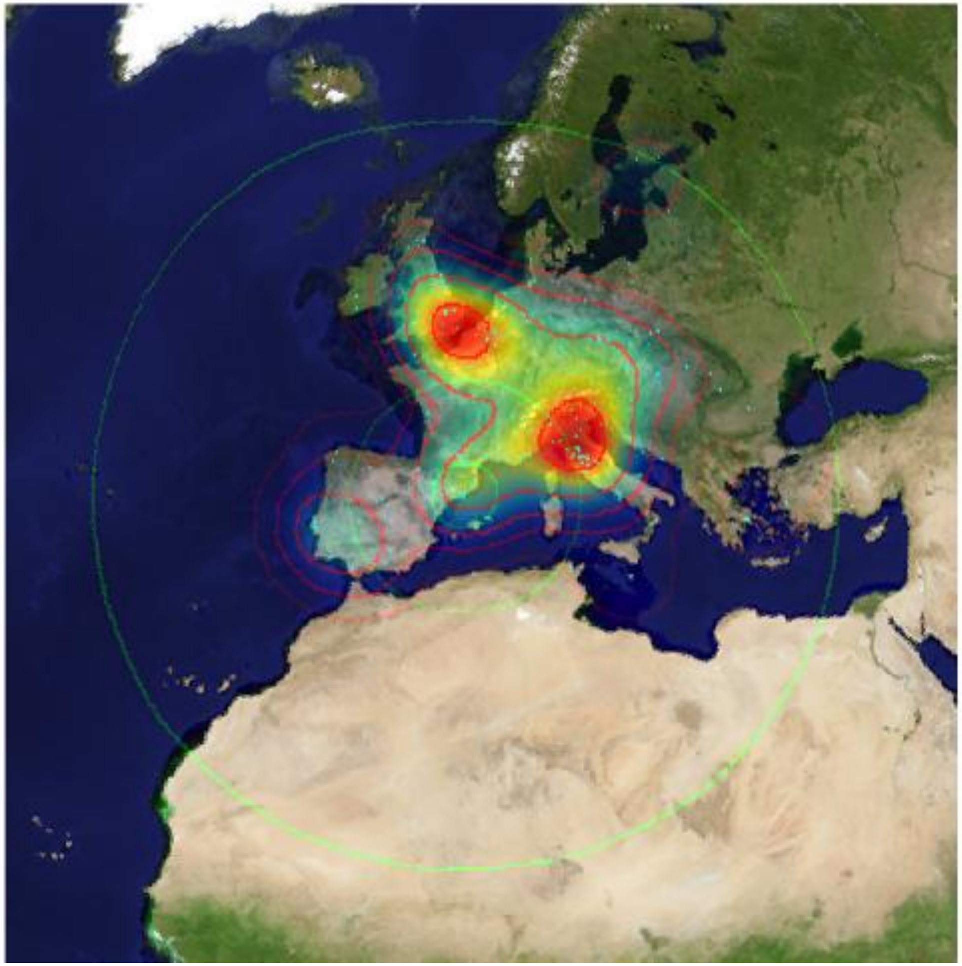



Im2gps

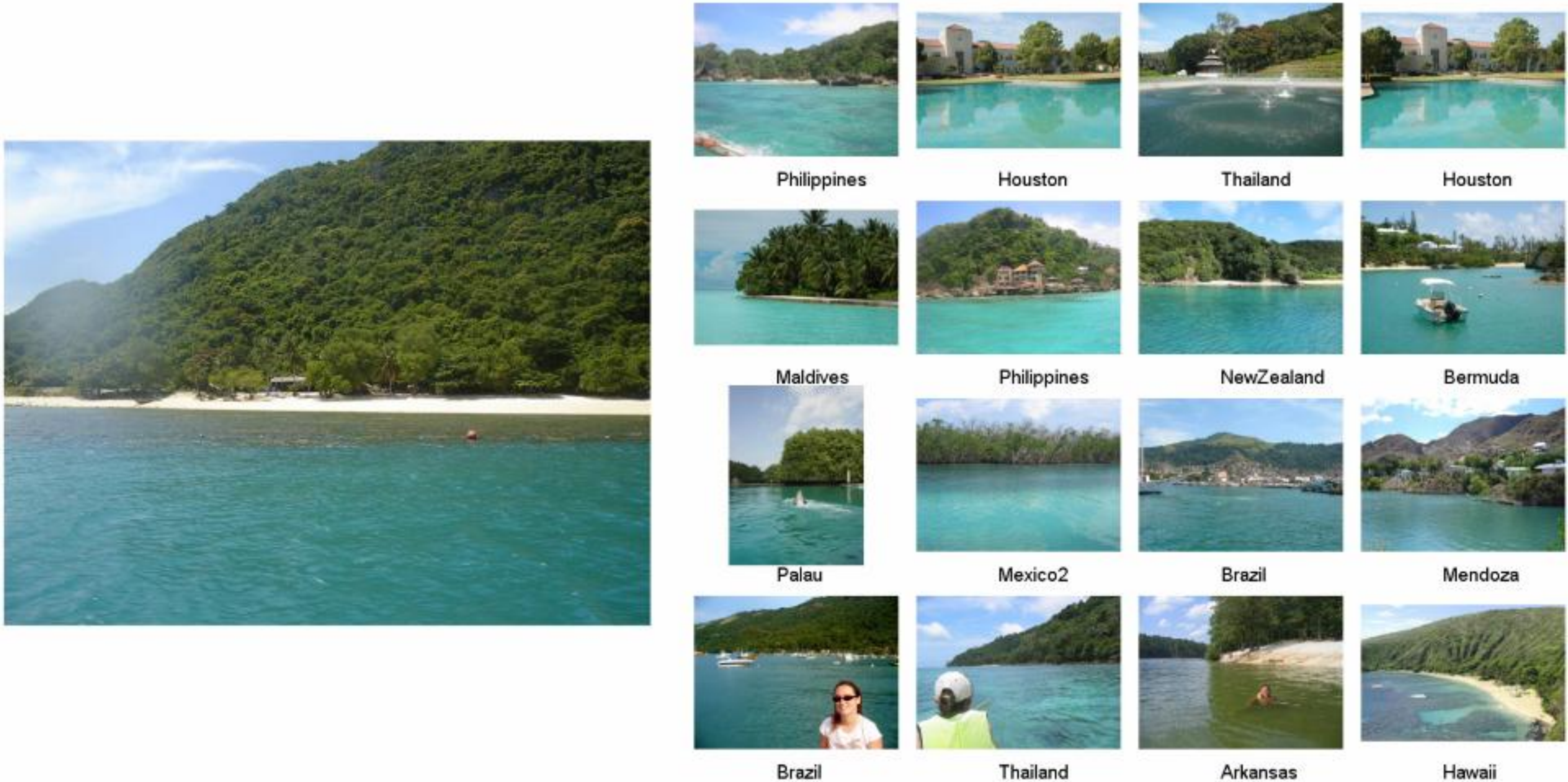
Example Scene Matches

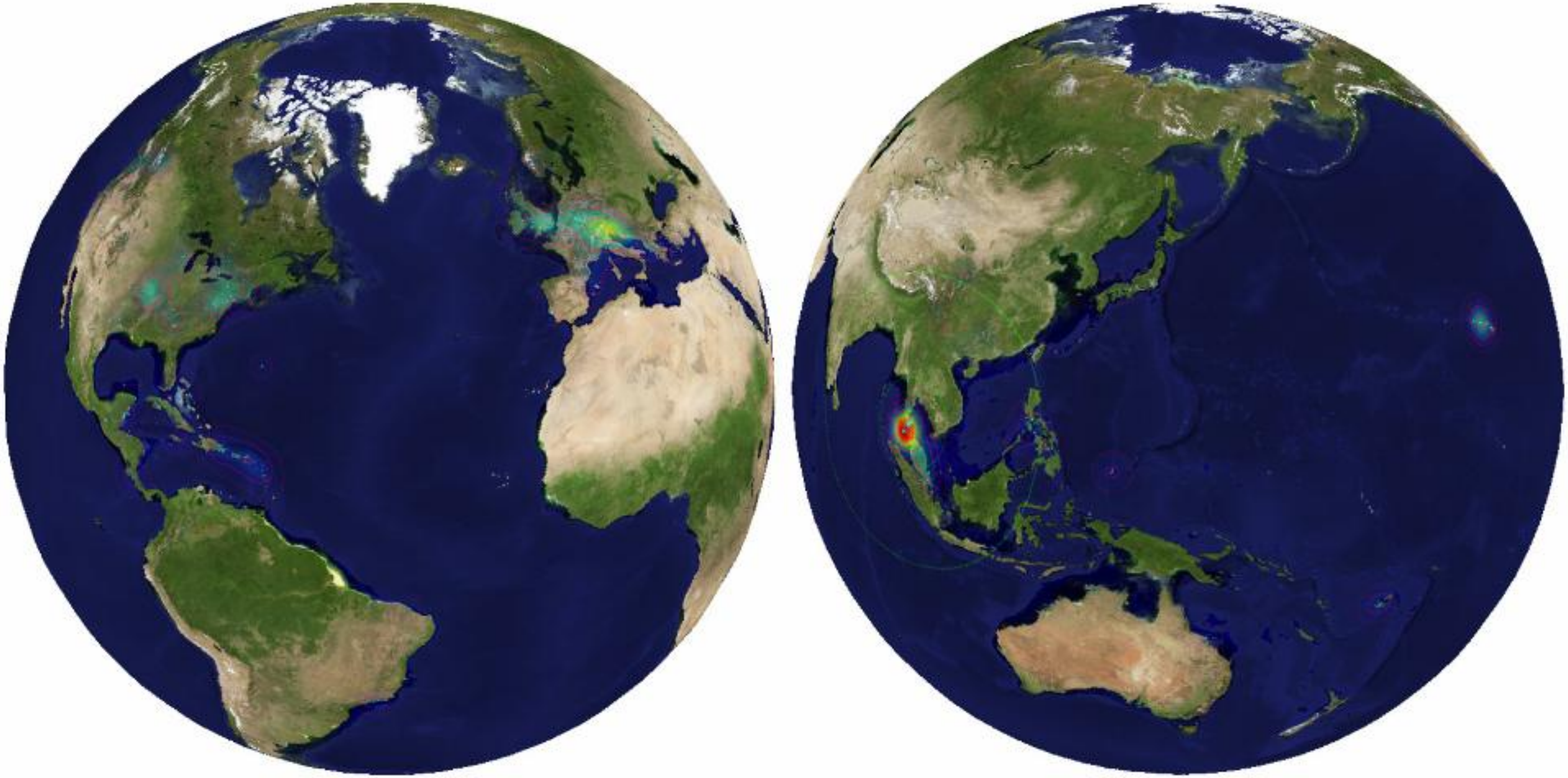
Voting Scheme

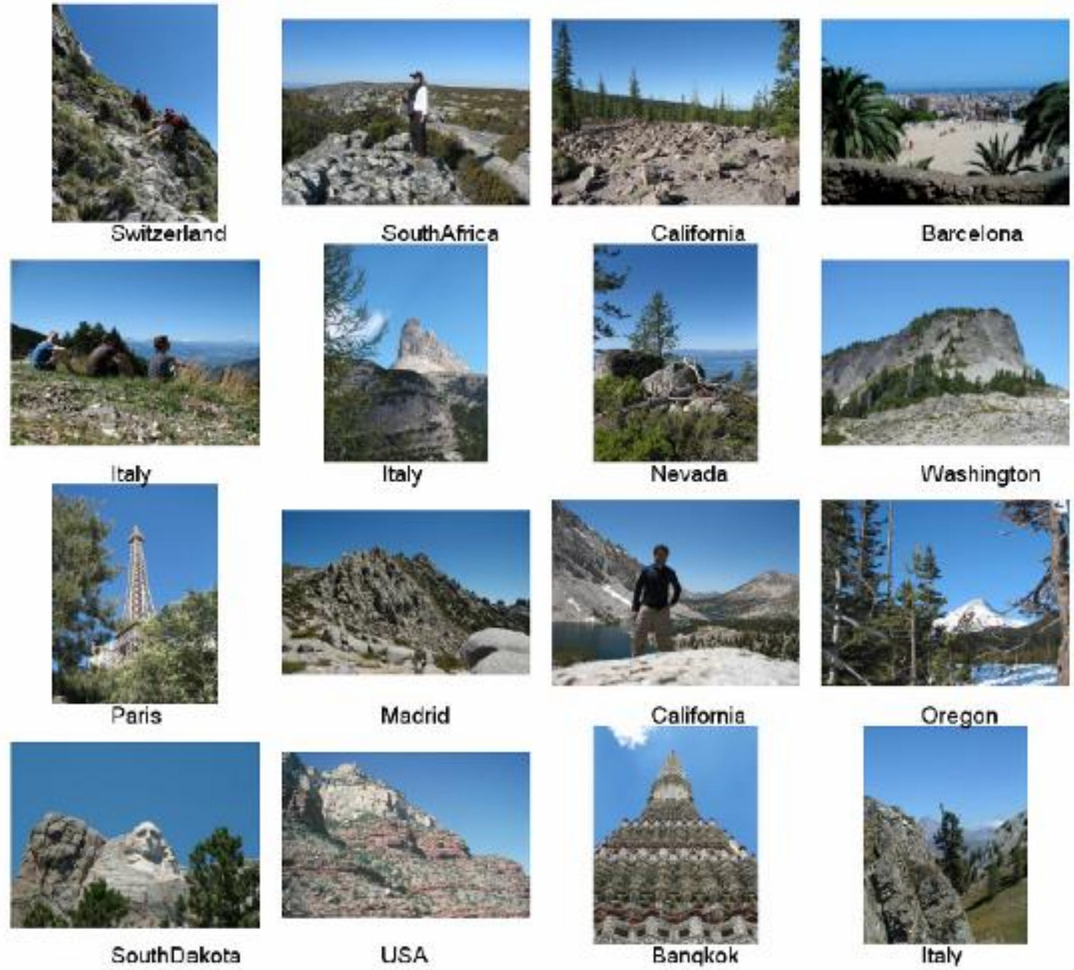


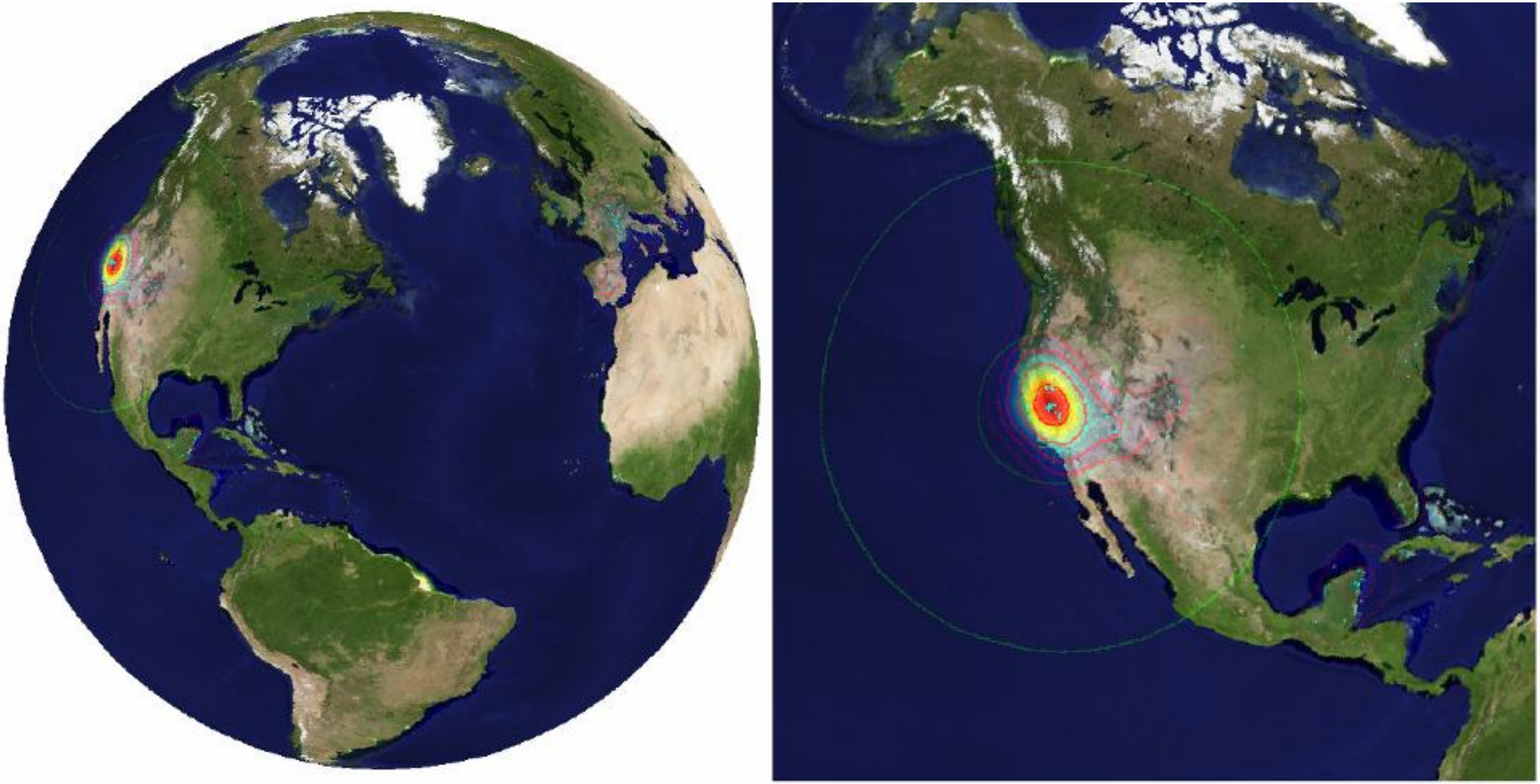


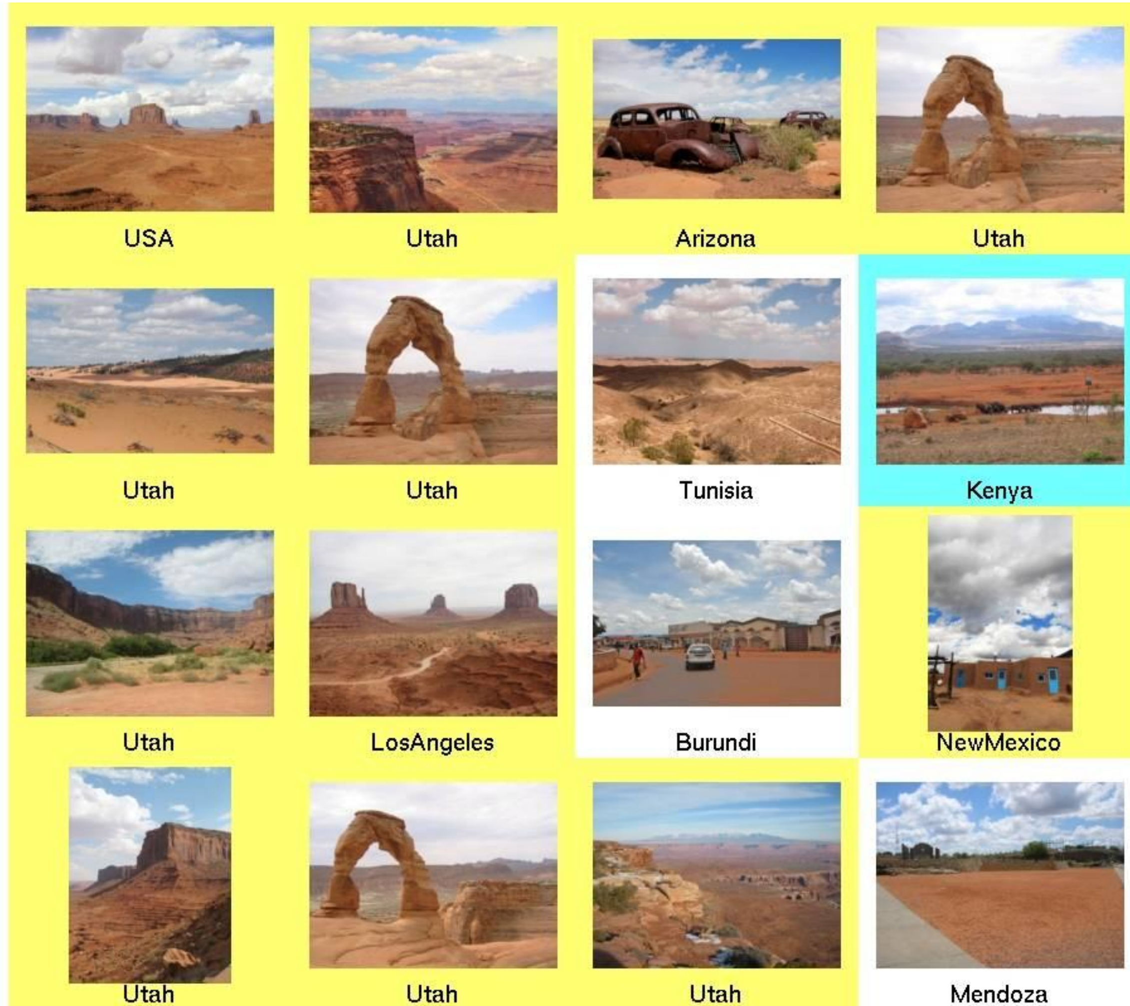
im2gps

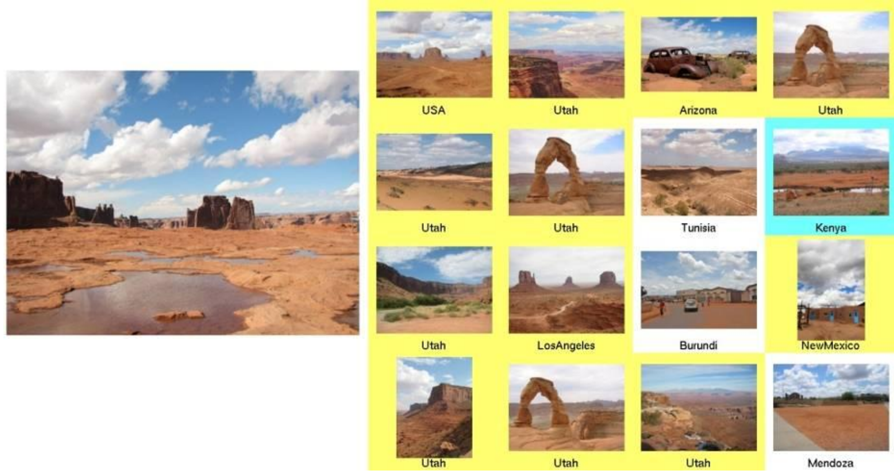

Philippines
Houston
Thailand
Houston
Maldives
Philippines
NewZealand
Bermuda
Palau
Mexico2
Brazil
Mendoza
Brazil
Thailand
Arkansas
Hawaii



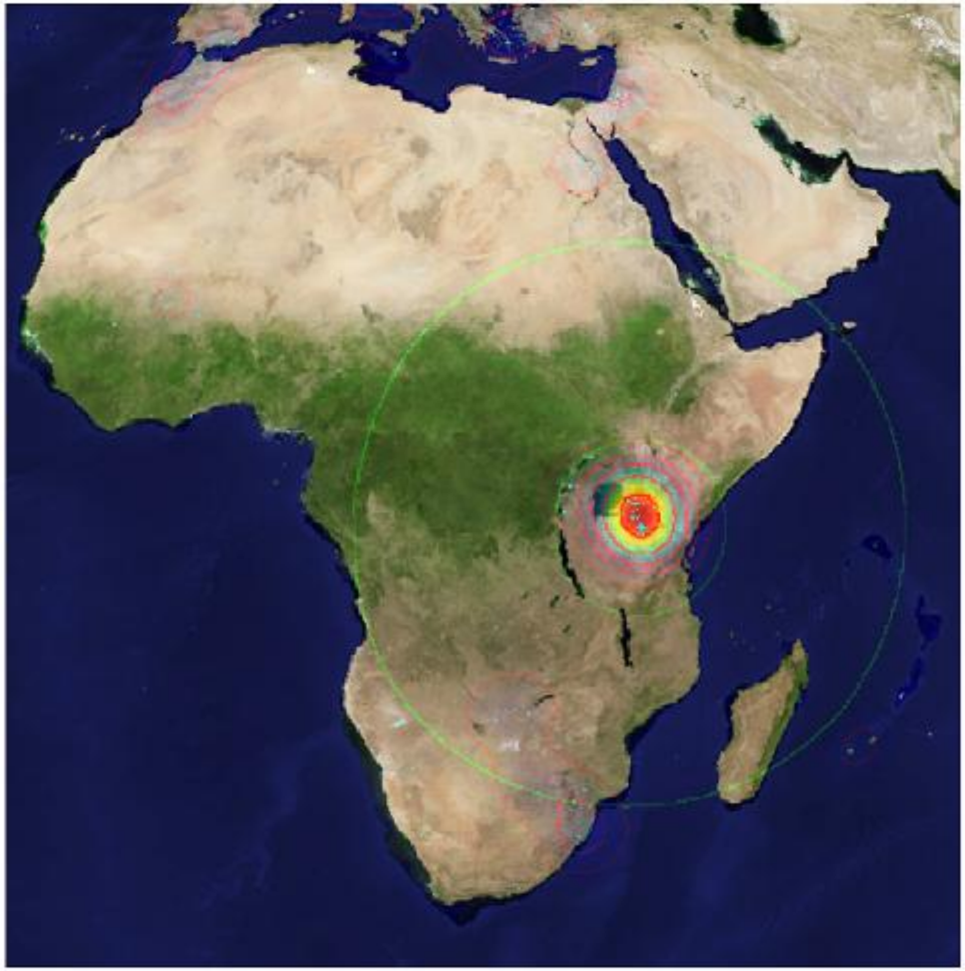

Switzerland
SouthAfrica
California
Barcelona
Italy
Italy
Nevada
Washington
Paris
Madrid
California
Oregon
SouthDakota
USA
Bangkok
Italy



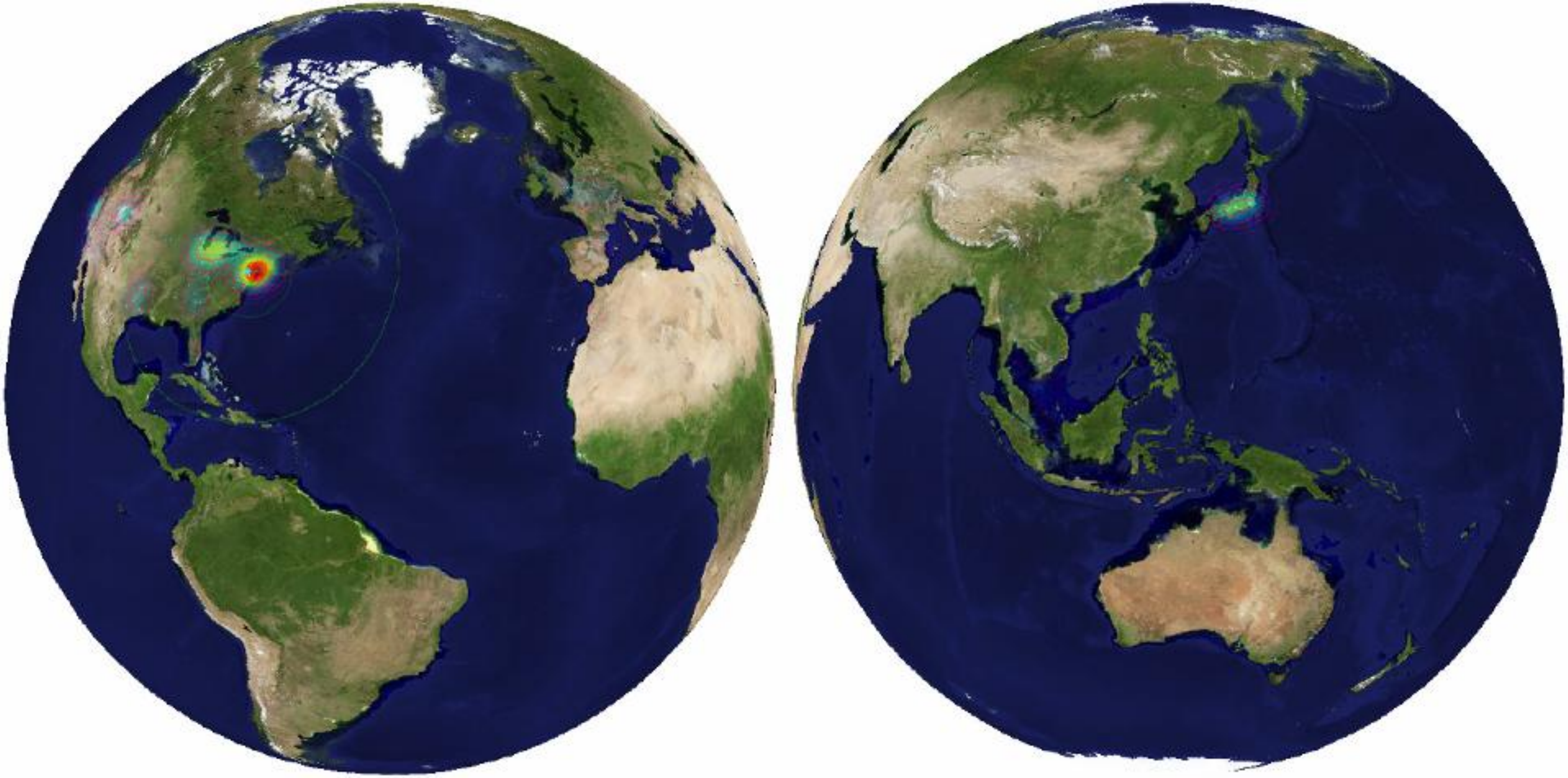

USA
Utah
Arizona
Utah
Utah
Utah
Tunisia
Kenya
Utah
LosAngeles
Burundi
NewMexico
Utah
Utah
Utah
Mendoza


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

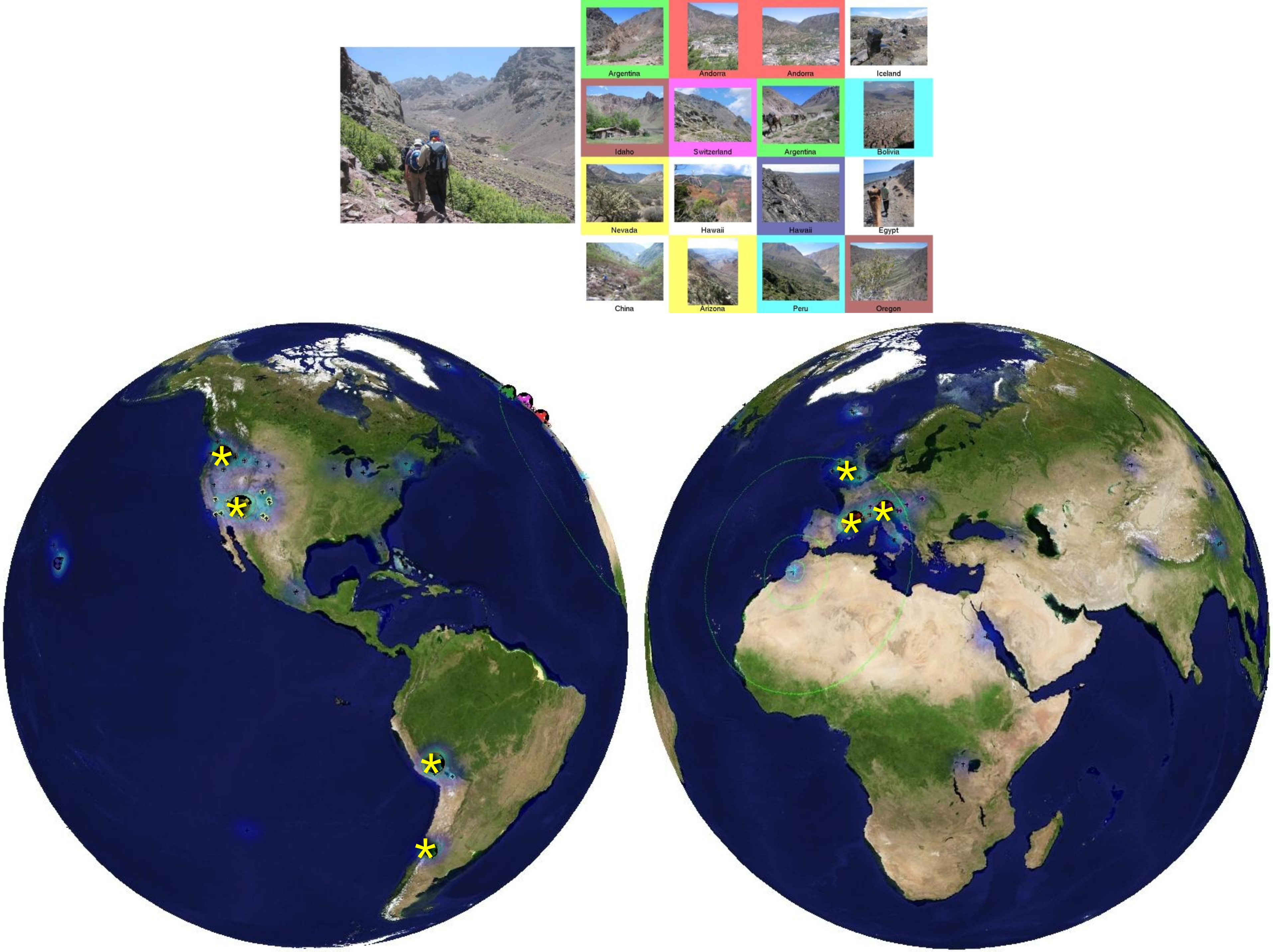

California
Oklahoma
SouthAfrica
Zambia
Kenya
Hyderabad
Mongolia
SouthAfrica
Kenya
Kenya
Zambia
Ethiopia
Nevada
africa
Morocco
Tennessee


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

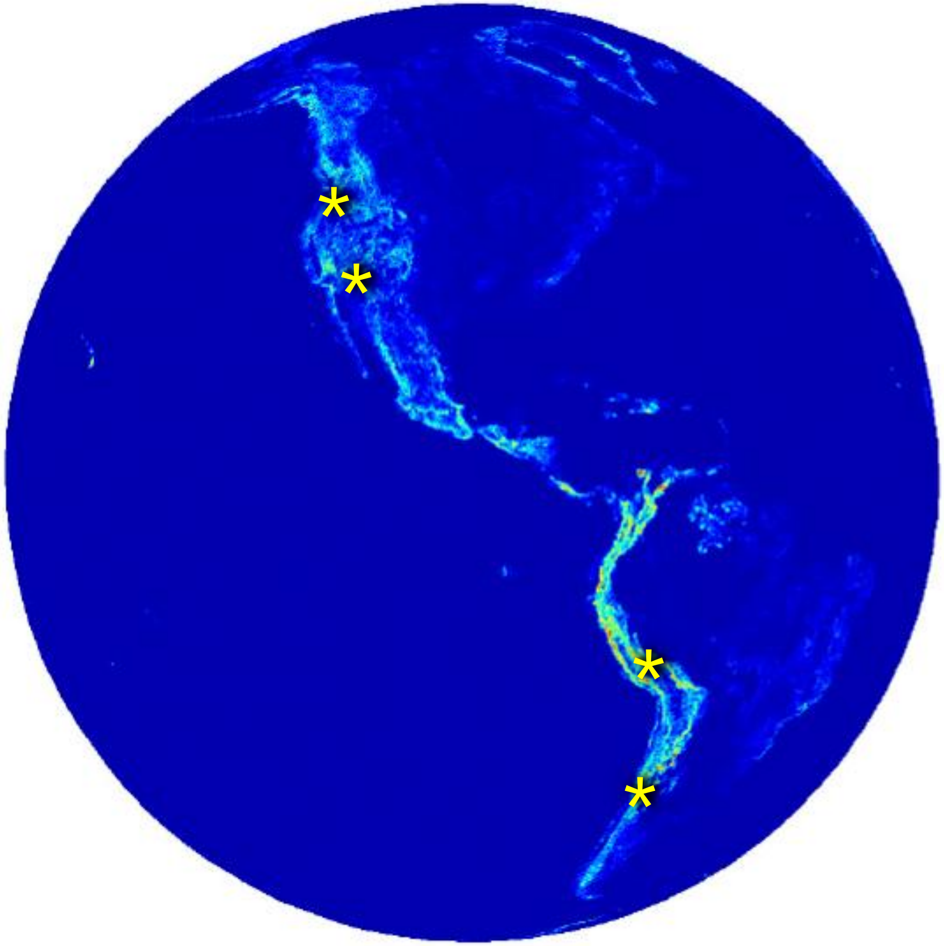


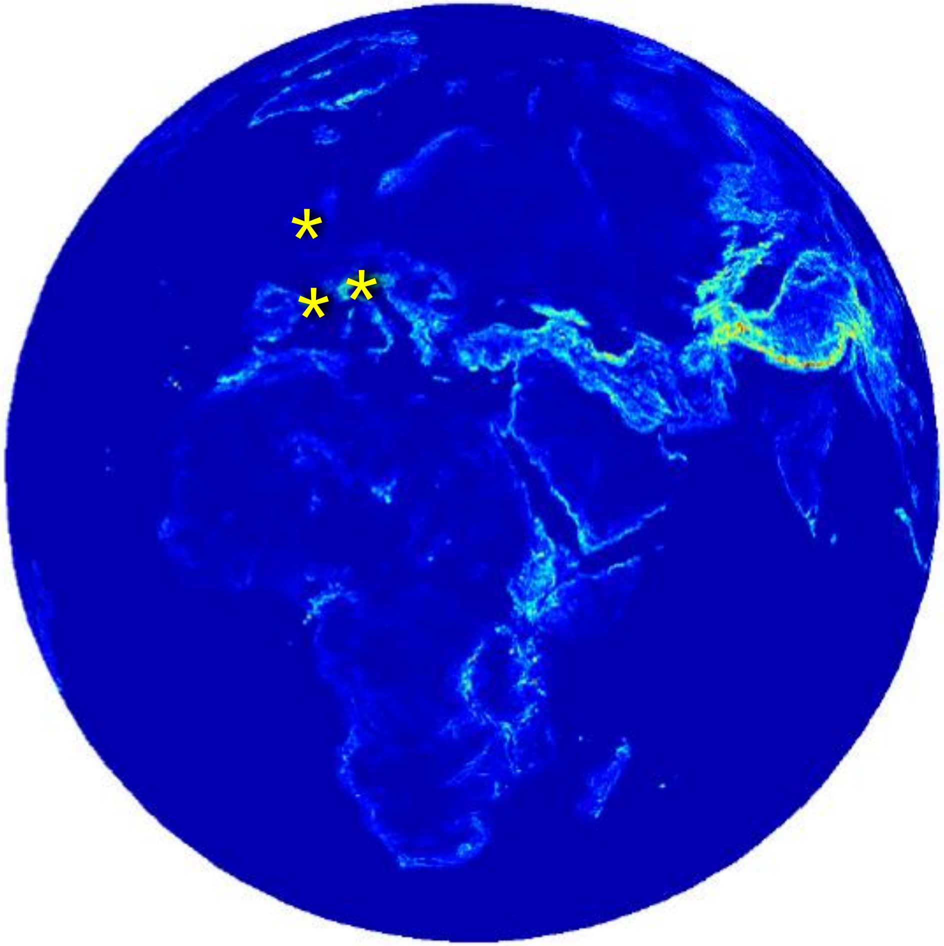
Data-driven categories



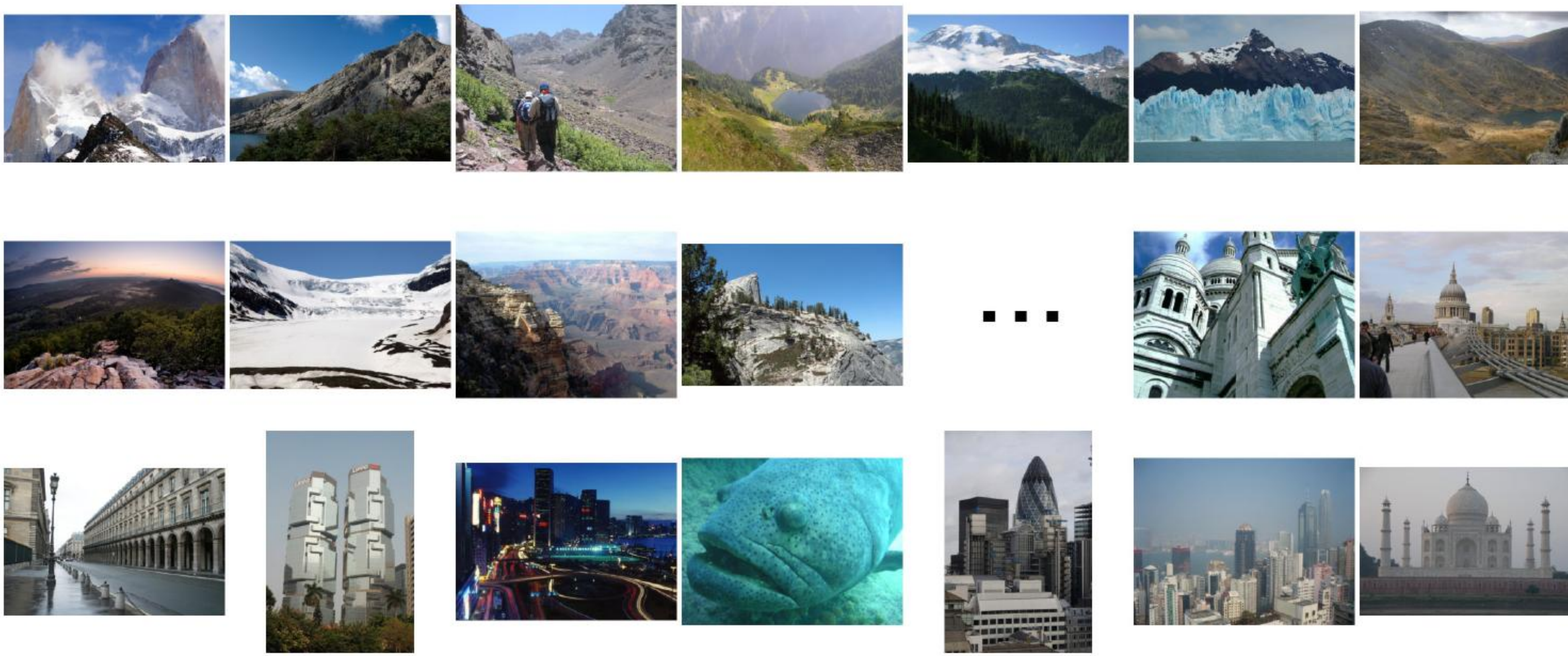
Argentina
Andorra
Andorra
Iceland
Idaho
Switzerland
Argentina
Bolivia
Nevada
Hawaii
Hawaii
Egypt
China
Arizona
Peru
Oregon

Elevation gradient =
112 m / km





Elevation gradient magnitude ranking



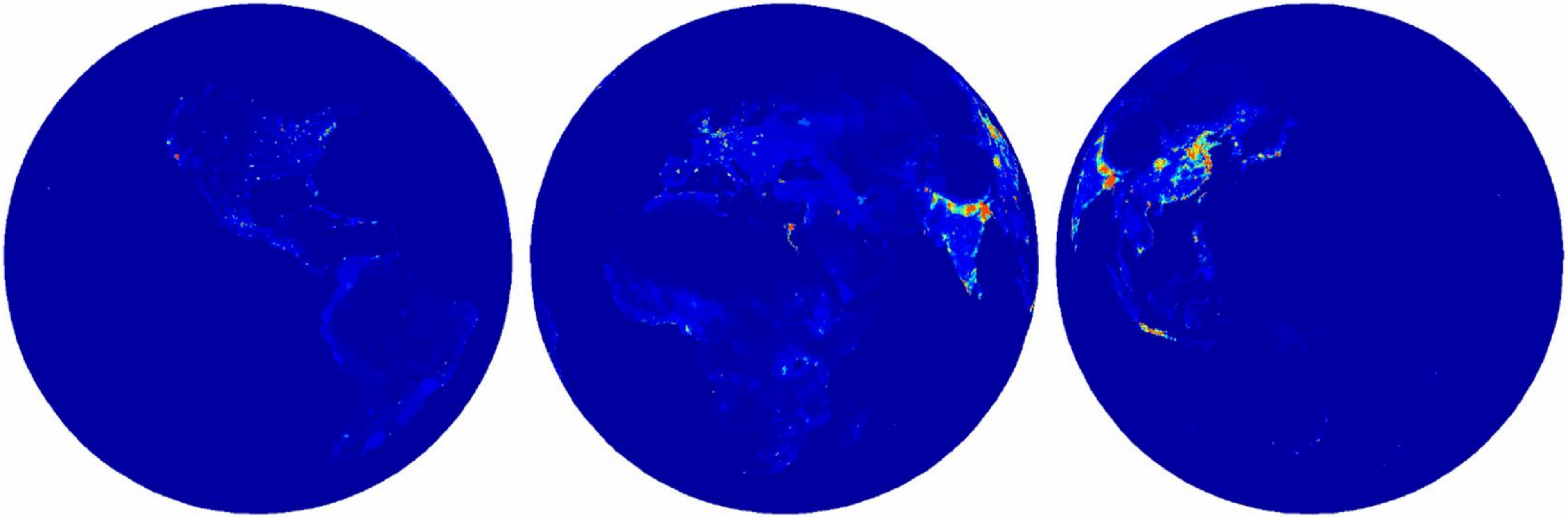
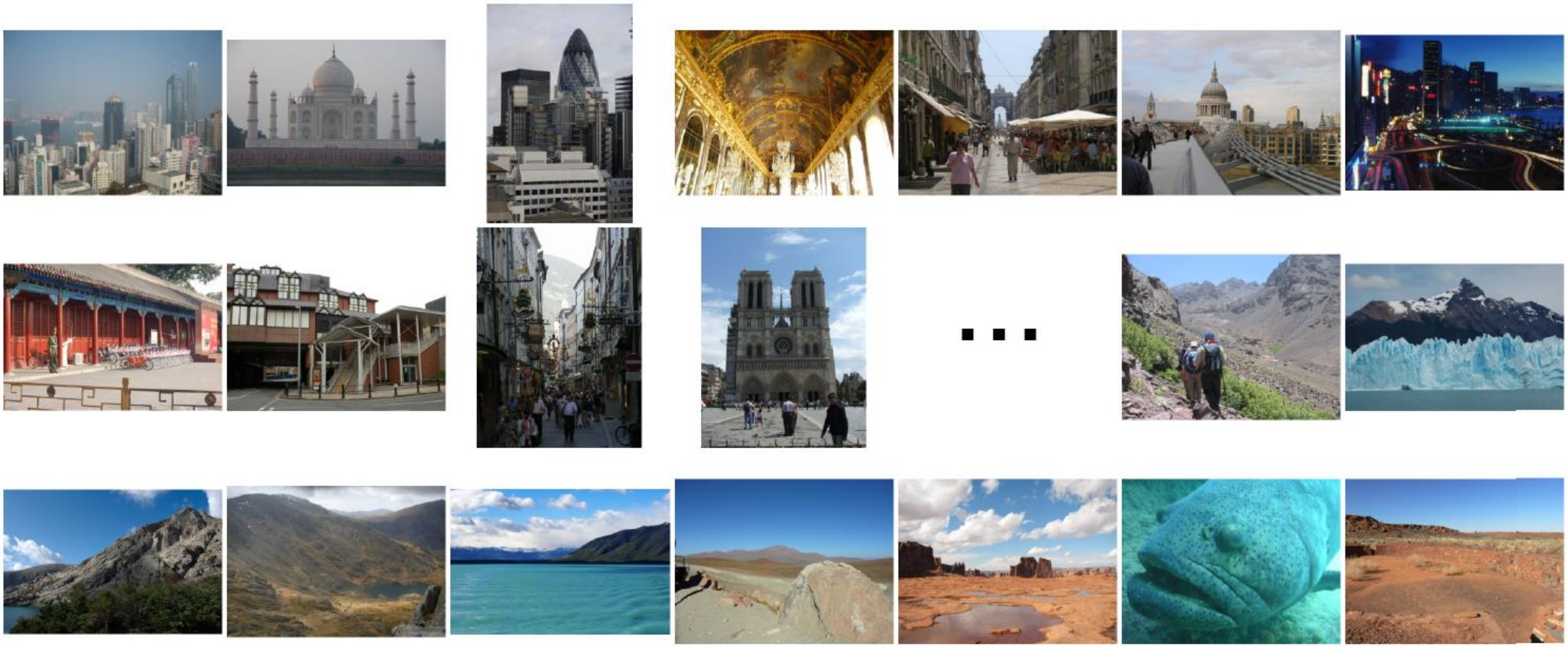


Figure 2. Global population density map.

Population density ranking



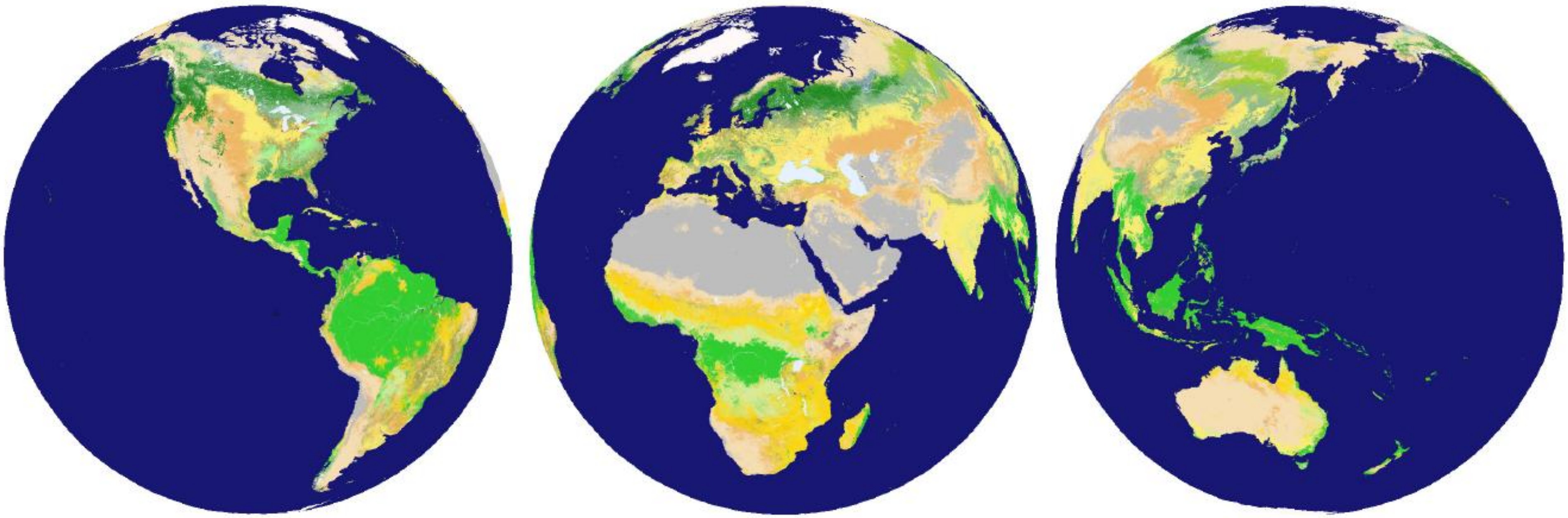
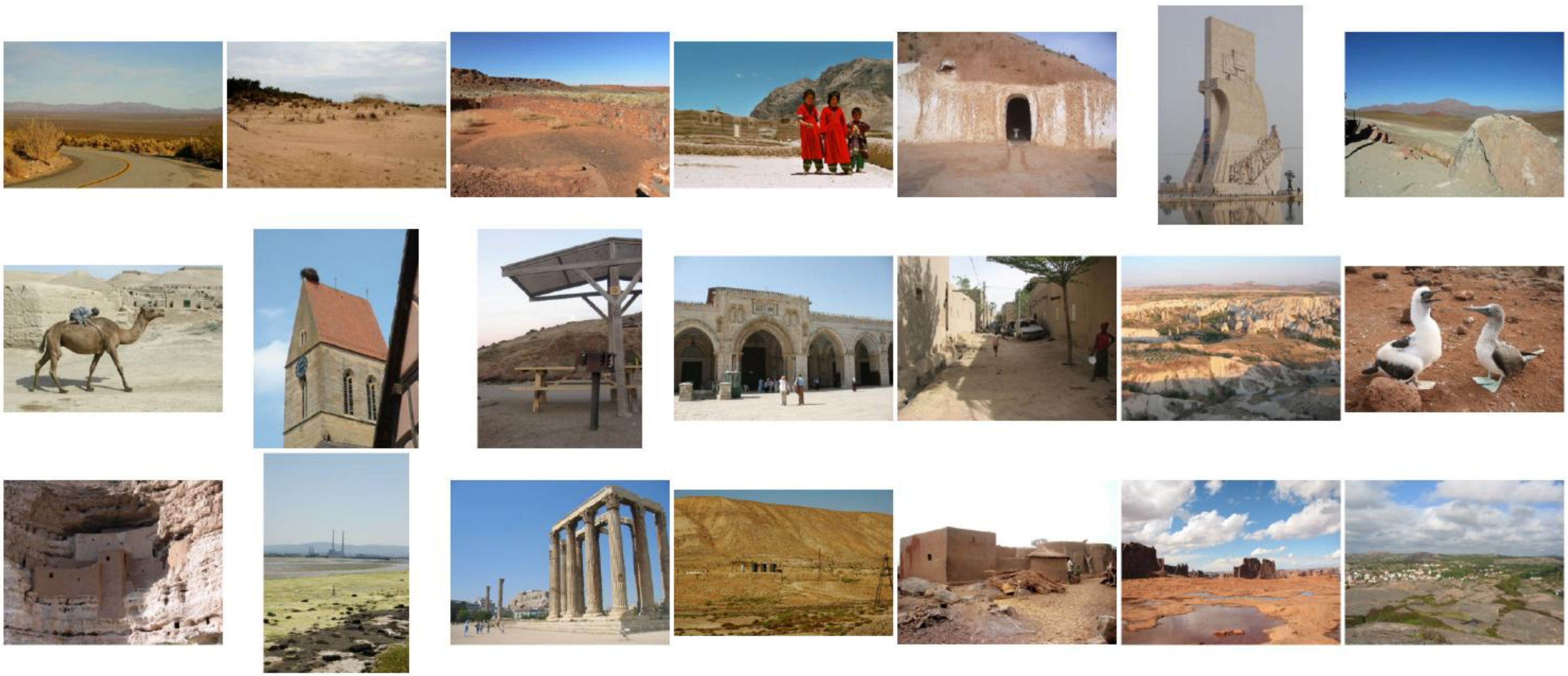
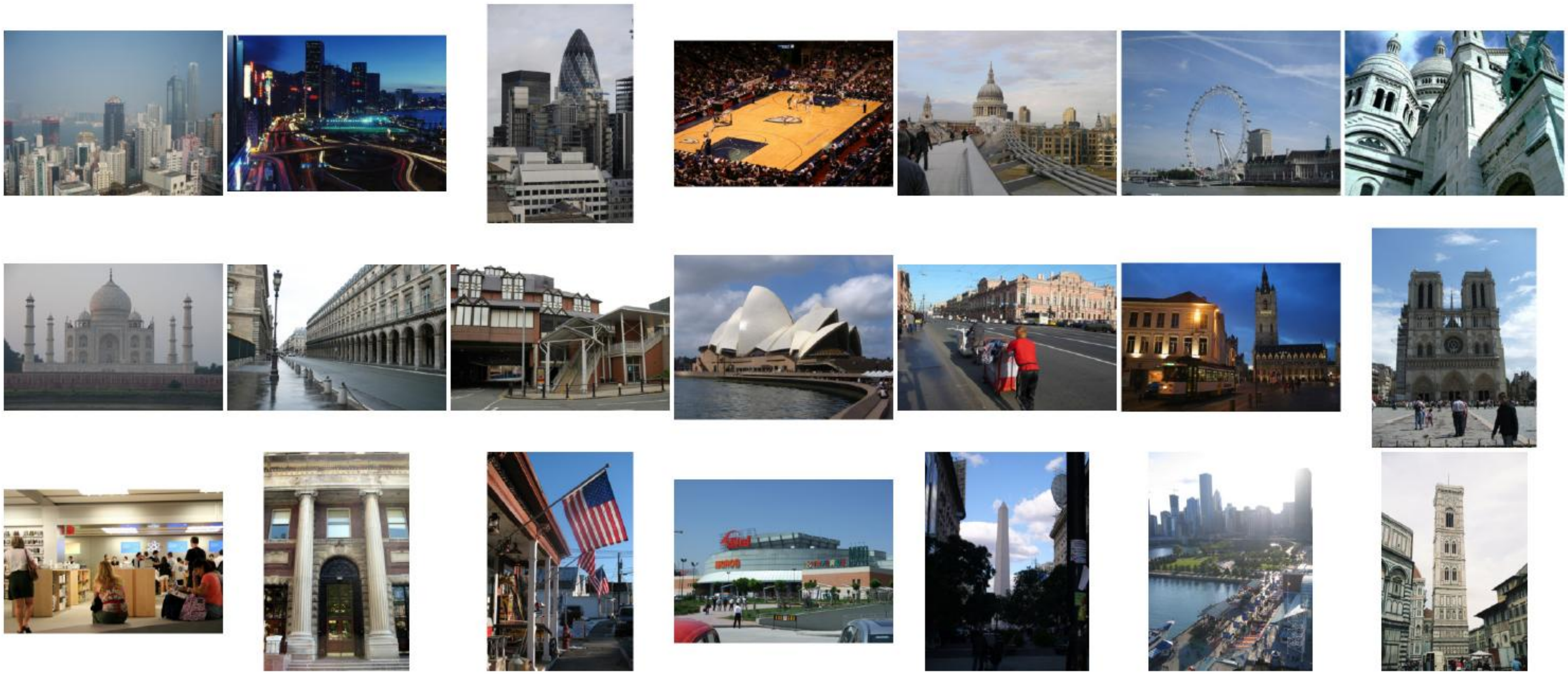


Figure 4. Global land cover classification map.

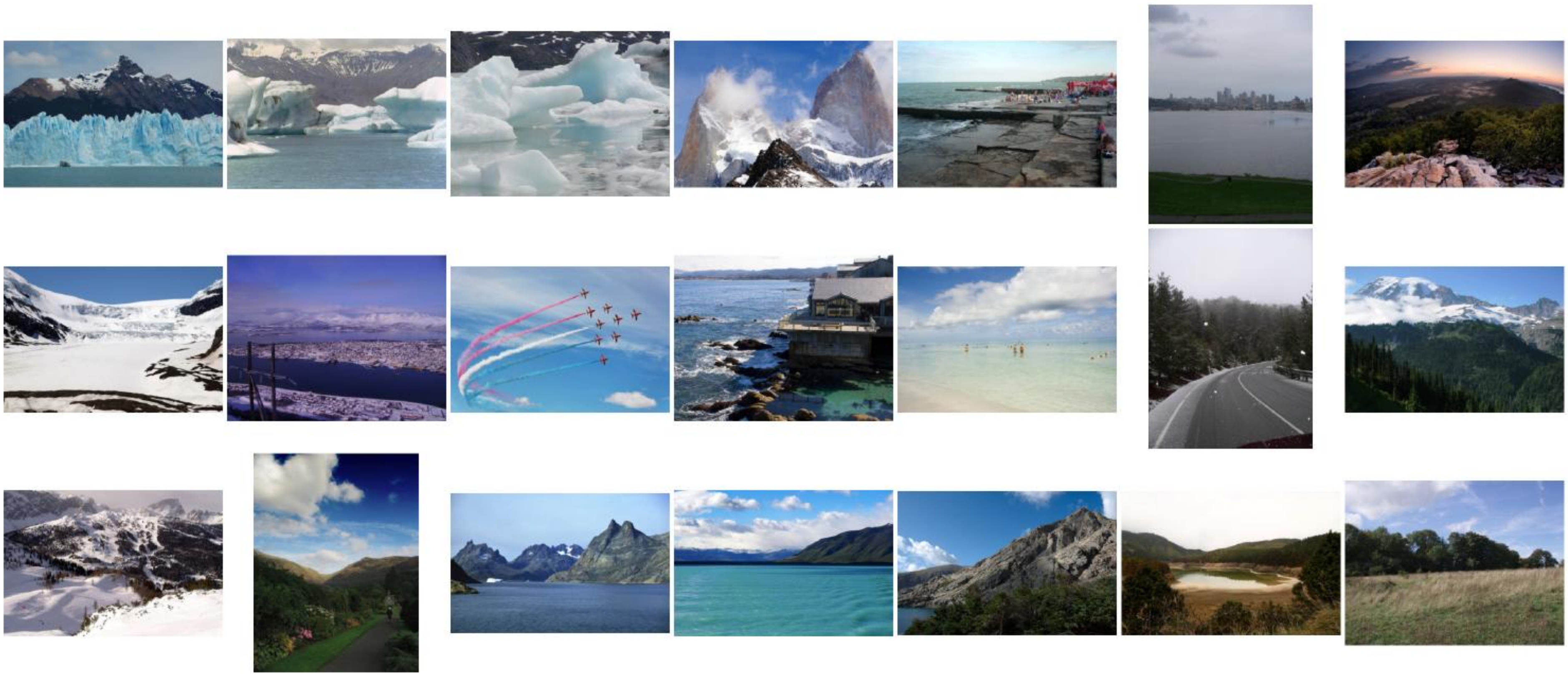
Barren or sparsely populated



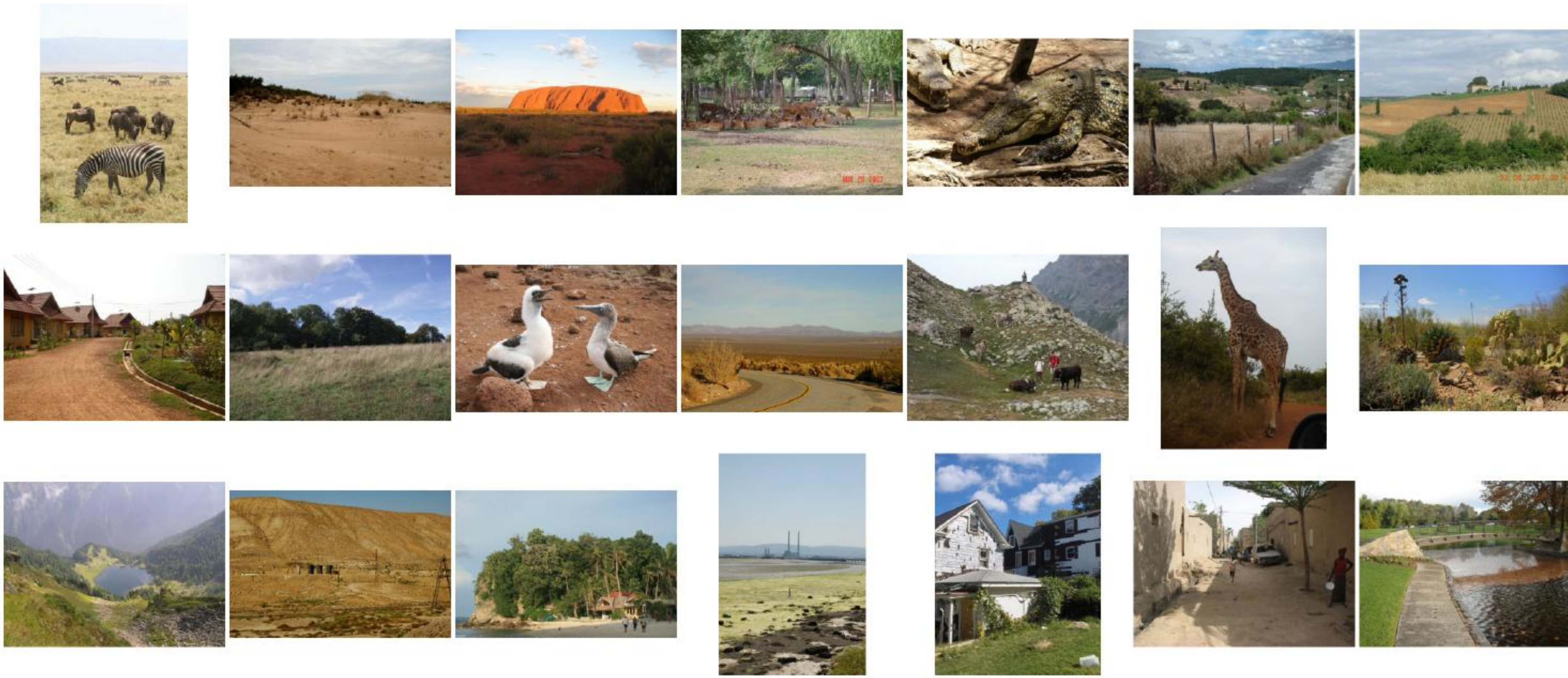
Urban and built up



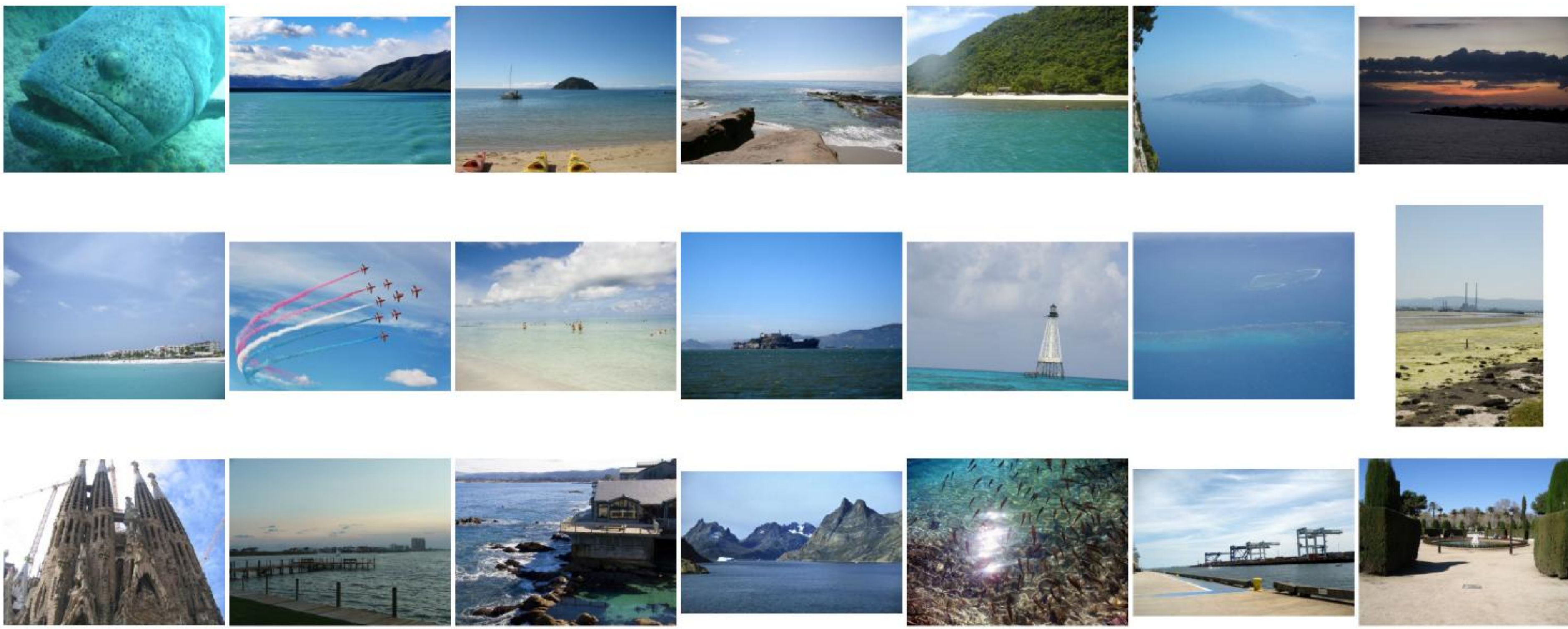
Snow and Ice



Savannah



Water

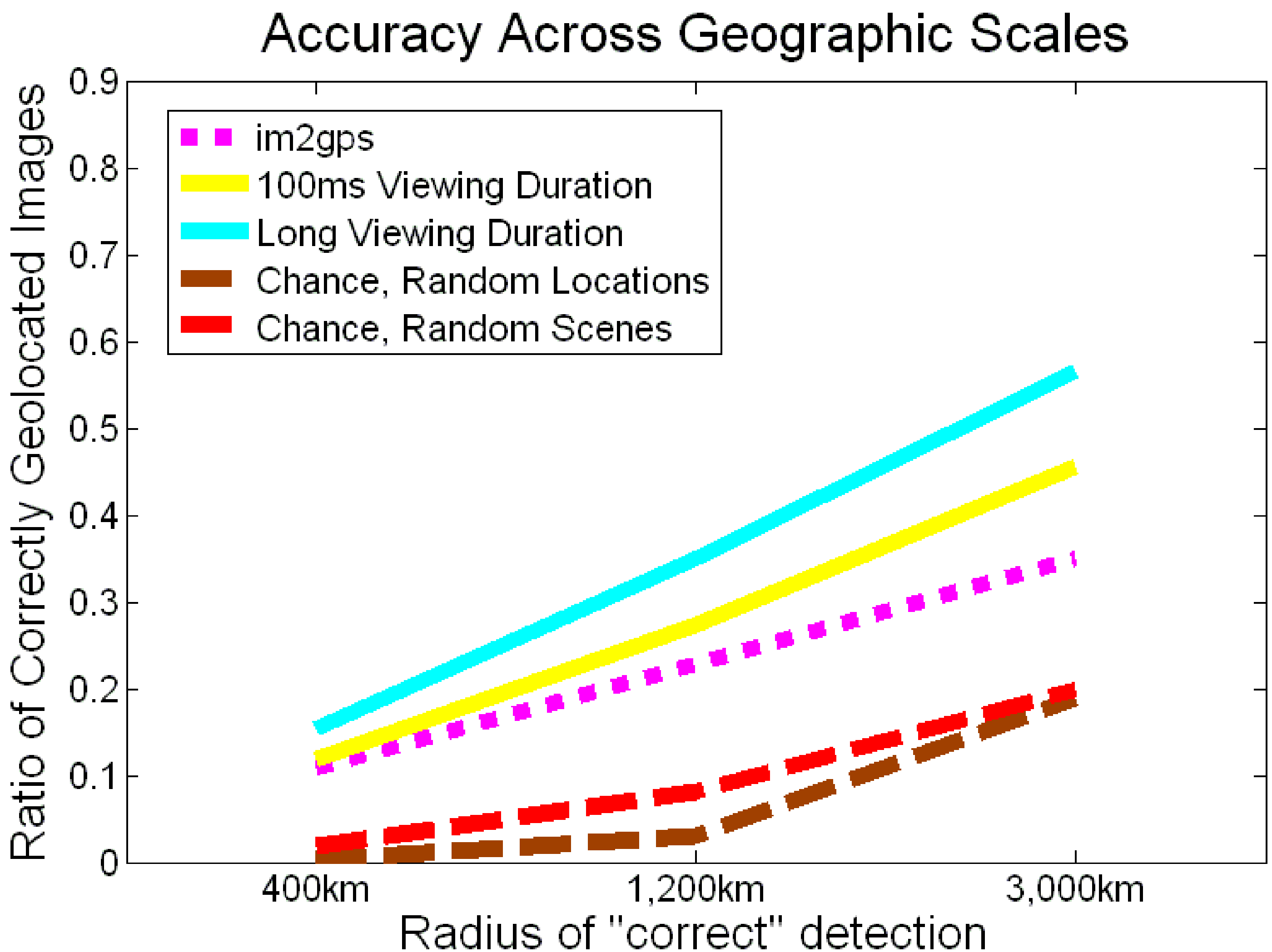


Conclusions

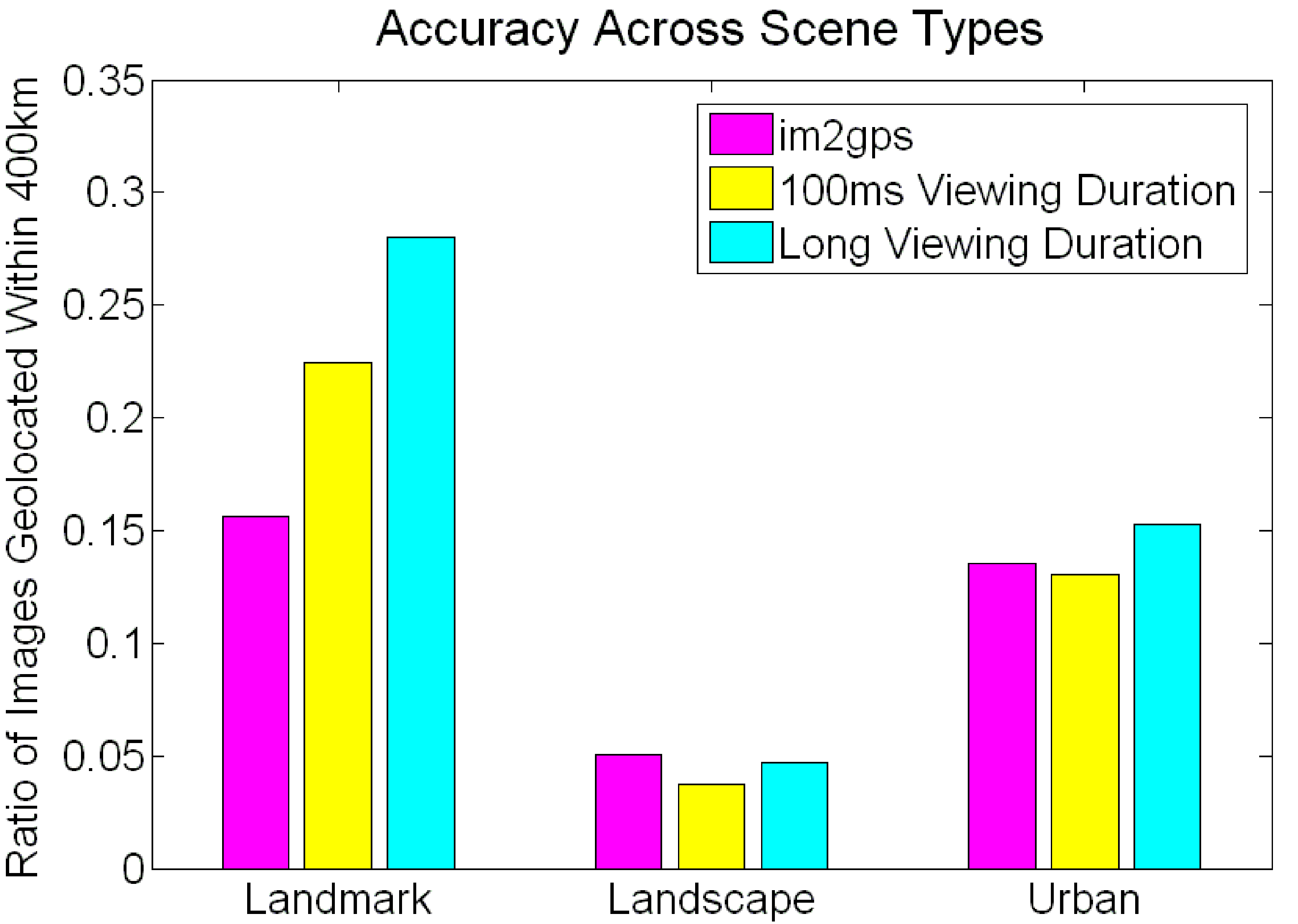
- There is plenty of useful information in a single image!
- …but we must use the rest of the visual world to understand it

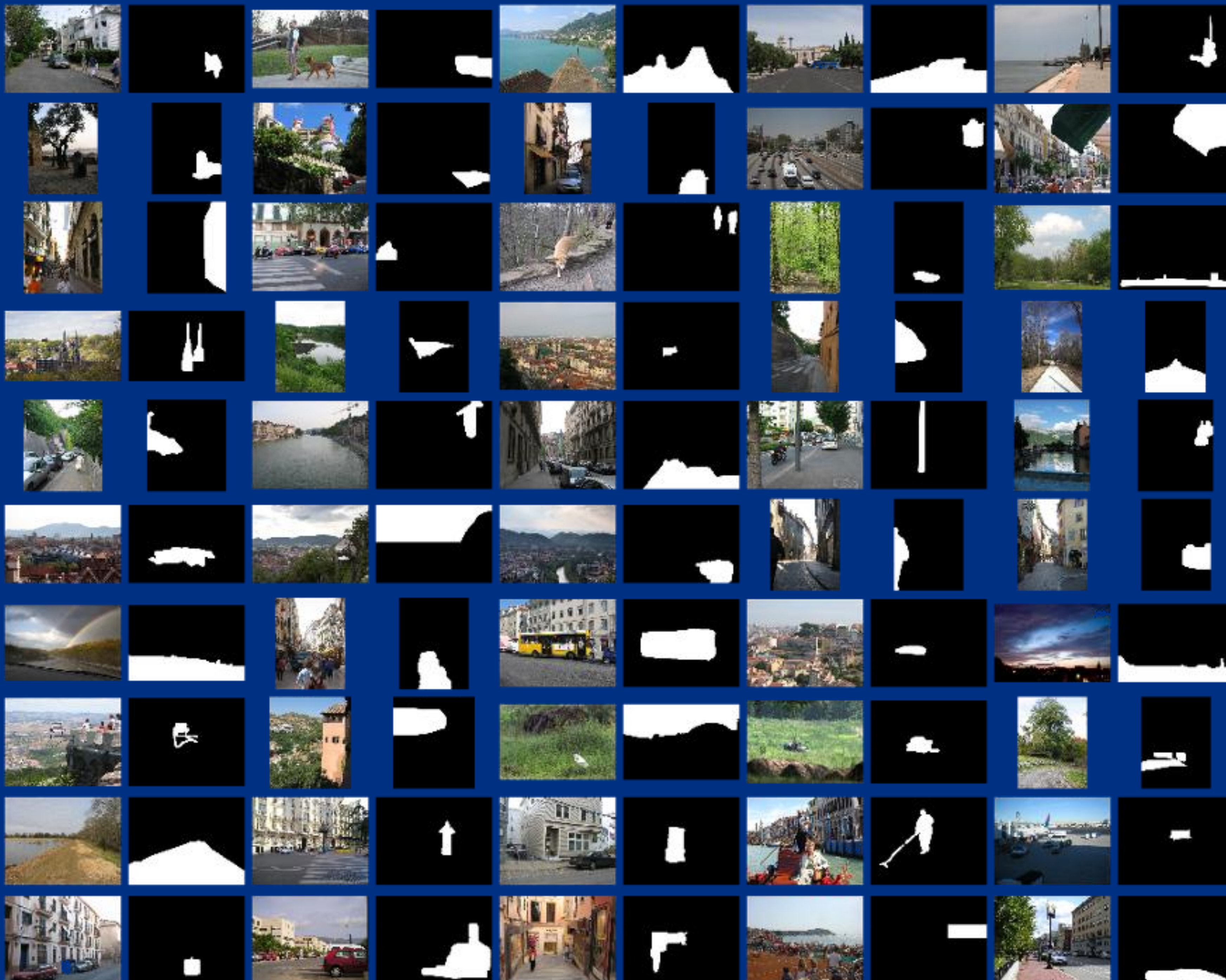
Quantitative Evaluation
(first time ever!)

Human vs. Machine

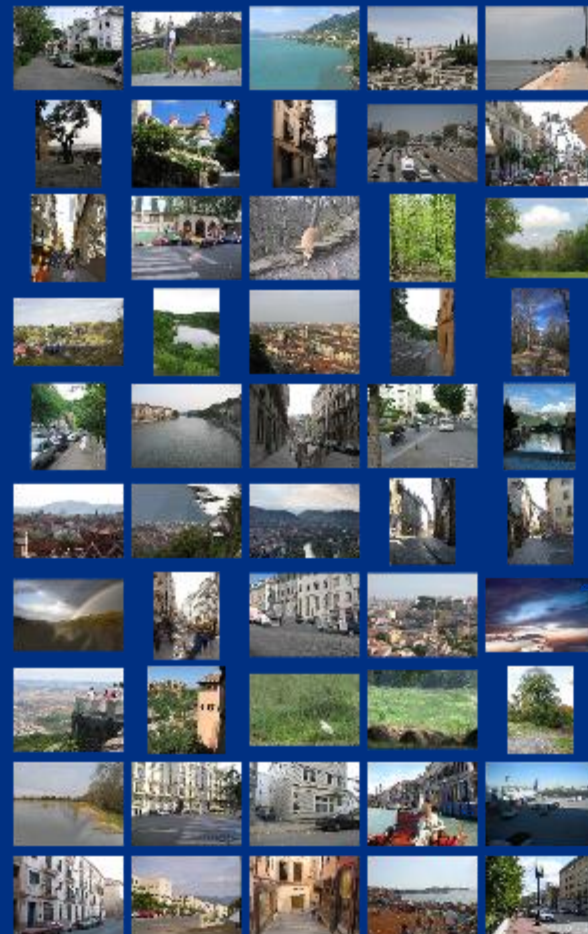


Human vs. Machine

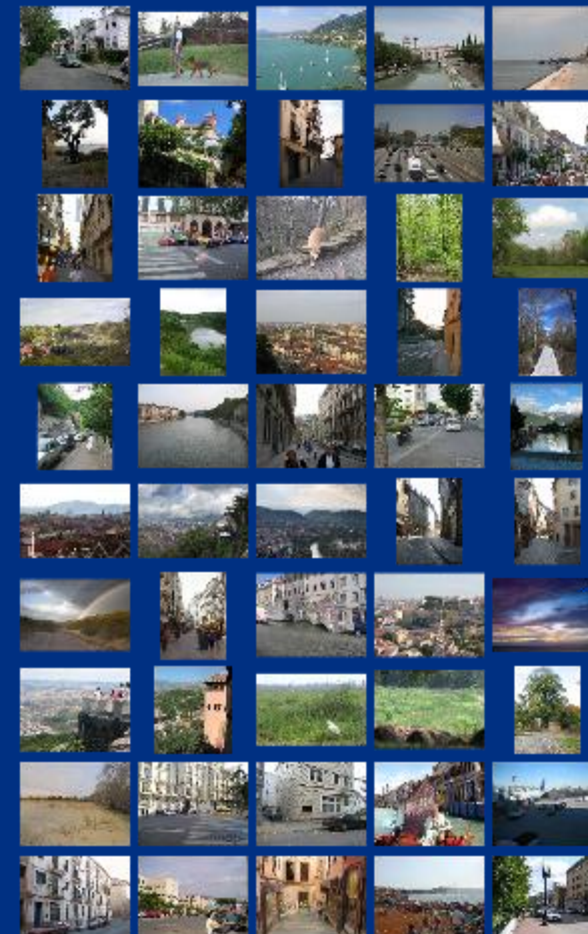




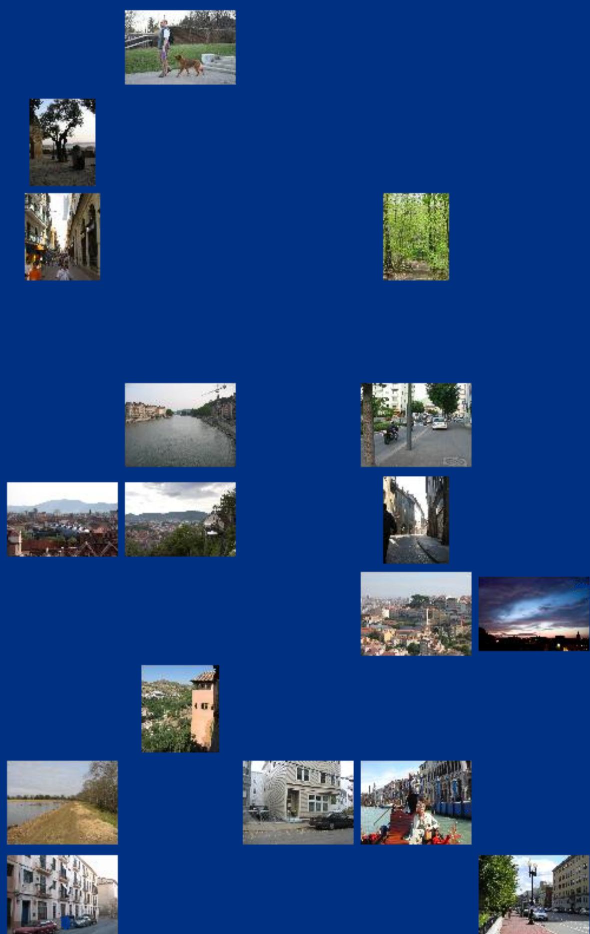
Original Images



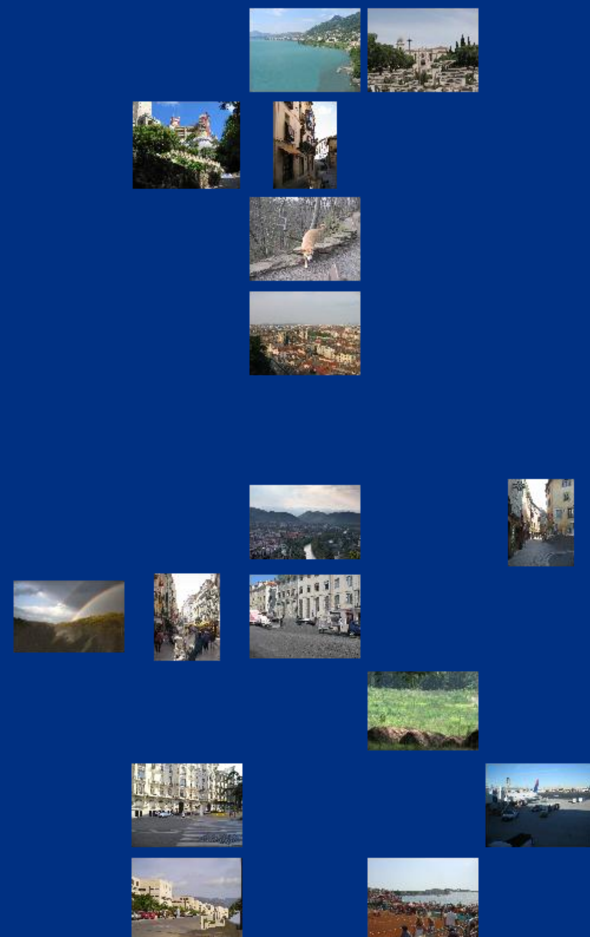
Criminisi et al.



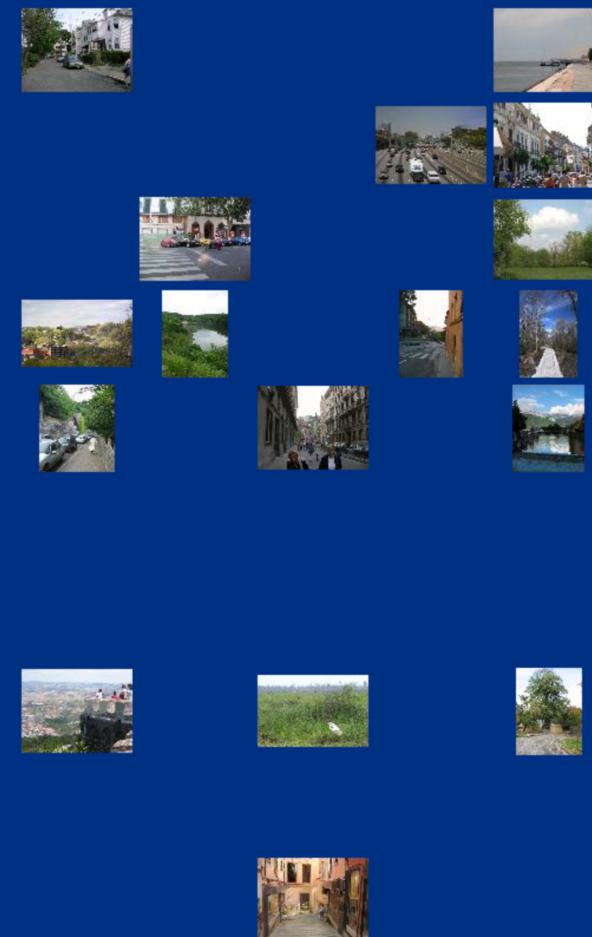
Scene Completion



Original Images



Criminisi et al.



Scene Completion

Real Image. This image has not been manipulated

or

Fake Image. This image has been manipulated


User Study Results