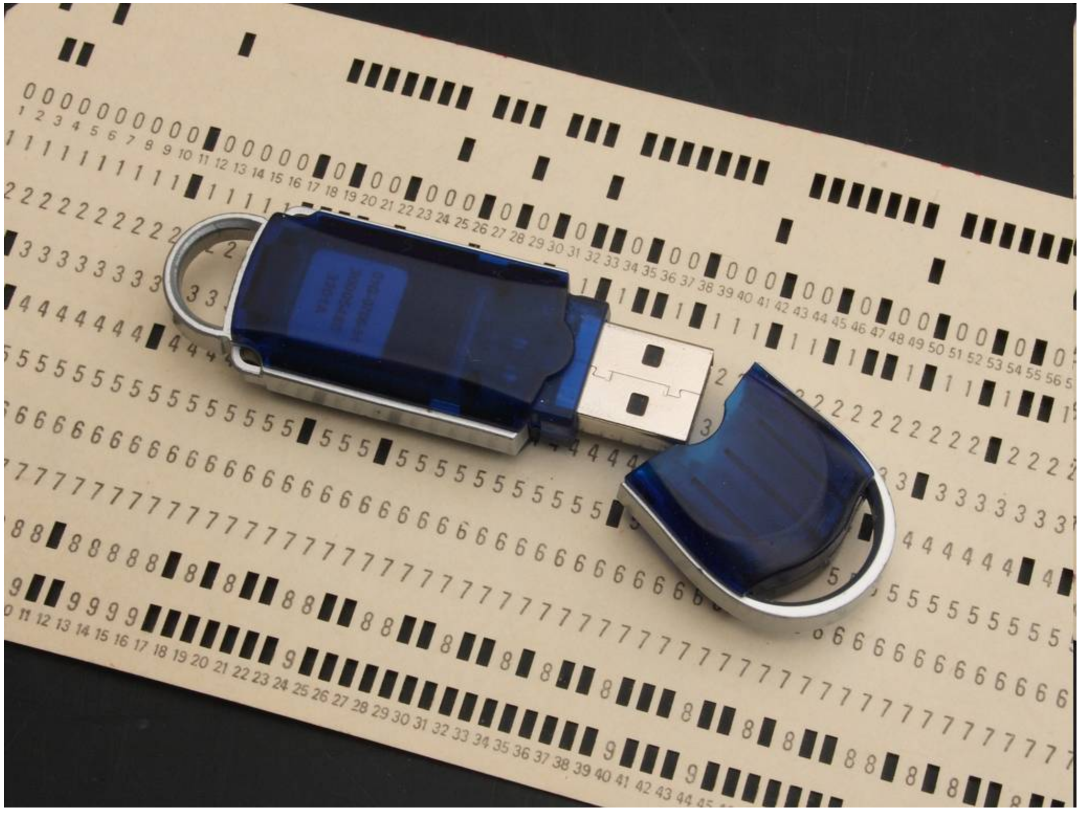
The Promise and Peril of

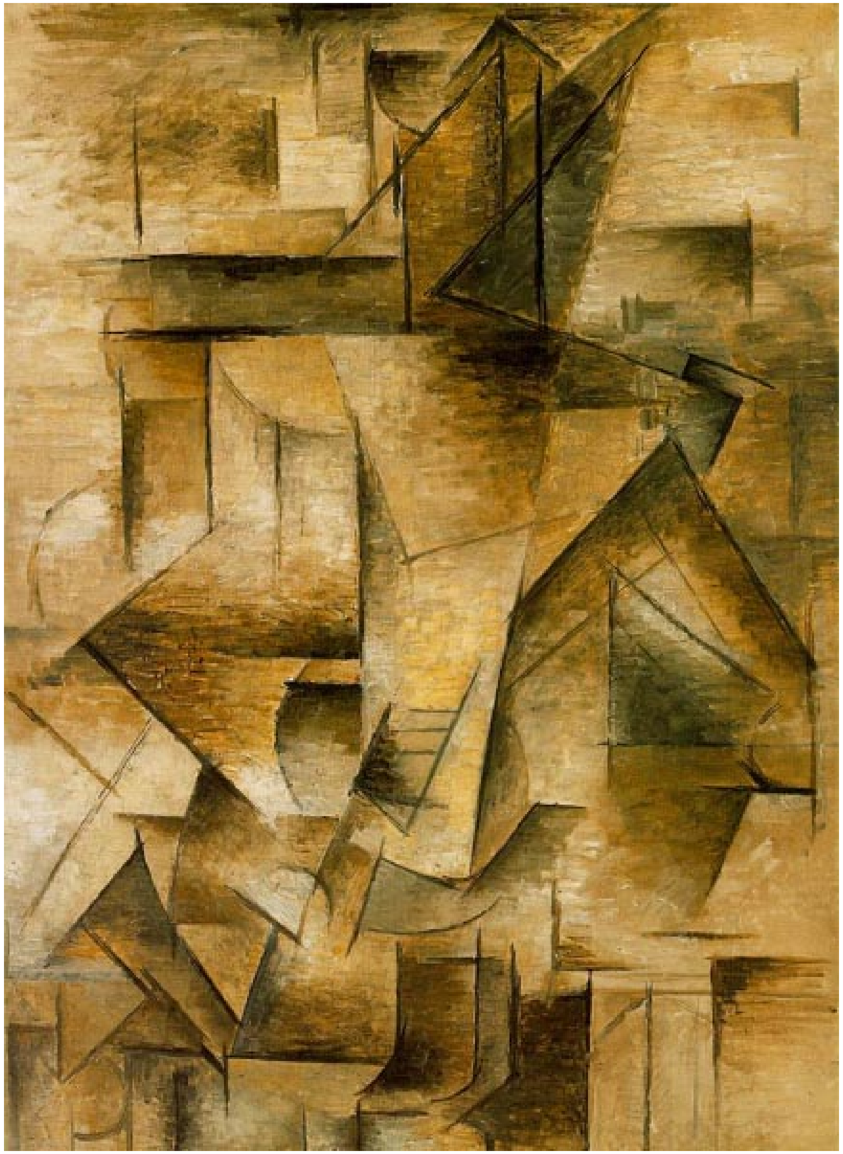
Big Data



Alexei (Alyosha) Efros
CMU (school-year), INRIA (summer)

It’s hard to be a computer…

How the computer sees the world:



The Guitar Player
Pablo Picasso (1911)

Some early work…

“Data, Data, Data… Watson, I need Data!”

[Sherlock Holmes, 1886]

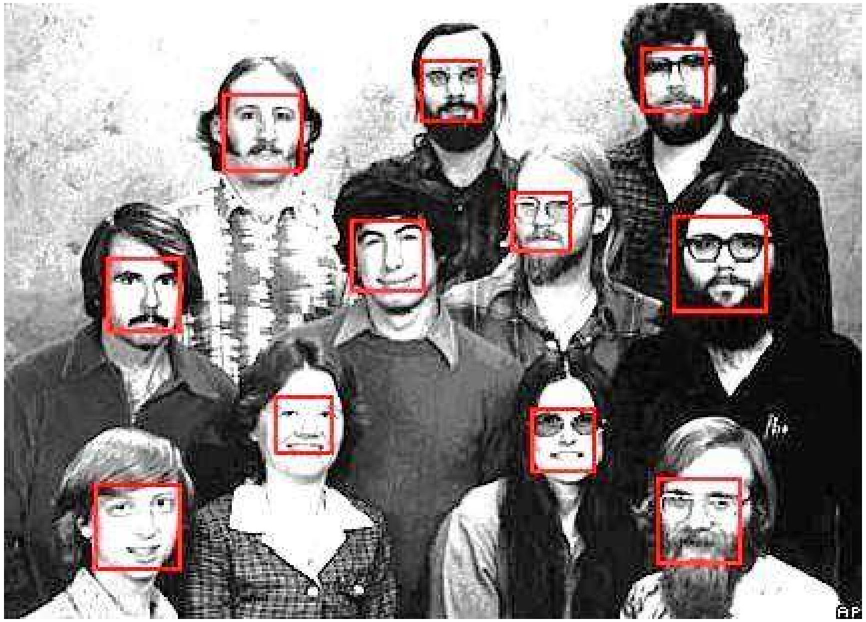
But If you want to publish a NIPS paper…

Data

Features

Learning Algorithm

Face Detection: Big Success Story



- Rowley, Baluja, and Kanade, 1998

- Schniderman & Kanade, 1999
- Viola & Jones, 2001

Modern Recognition is largely Data-Driven

- In non-linear SVMs:
  - In ML, people report ~10% of data are support vectors
  - In recognition, up to 2/3 of data are support vectors!!!
- In linear SVMs:
  - Typical setup: 4000 dim. HOG, only 300 “chair” examples

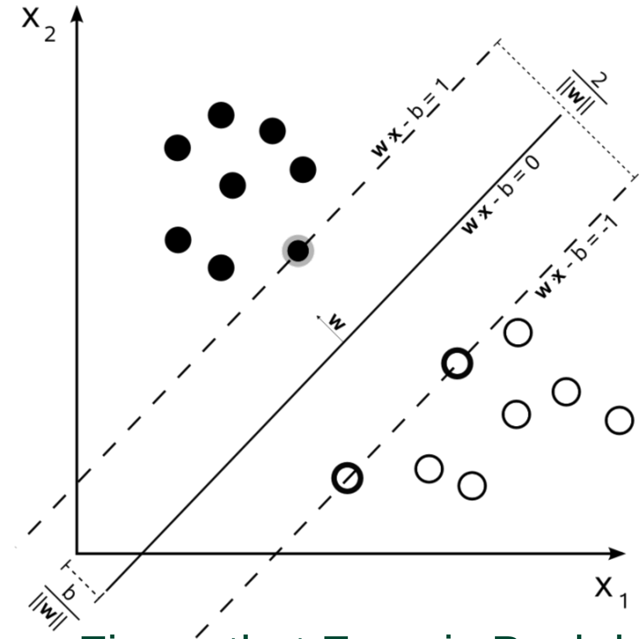


Figure that Francis Bach hates

Recognition Learning Spectrum

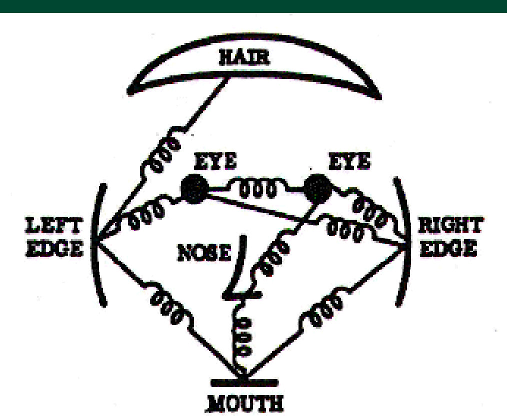
Extrapolation problem
Generalization

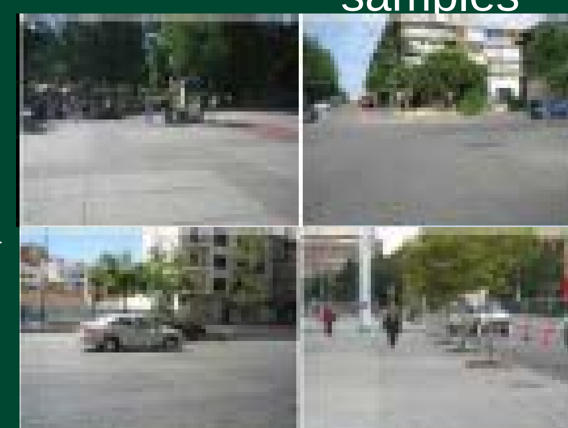
Interpolation problem
Correspondence

1 — 10 — 10^2 — 10^3 — 10^4 — 10^5 — 10^6 — ∞

Number of training samples

Traditional datasets





Slide by Antonio Torralba

Everything else being equal…

… the visual world is just much **richer**!

- MNIST Digits
  - 10 digits *
  - ~1,000 variations = 10,000
- English words
  - ~100,000 words *
  - ~5 variations = 500,000
- Visual world
  - ~100,000 objects *
  - ~10,000 variations (pose, scale, lighting, intra-category)
  - **= 1,000,000,000 (1 billion!)**

MNIST:
60,000 examples

Topic Models (Blei):
11M examples

Yet, we train on 15 examples?!

Caltech 101

If you want to start a company…

Data

Features

Algorithm

Amnon Shashua

To make research progress…

Data / Features / Algorithms

Big Message…

Keep the data --
you never know when you will need it!

Texture

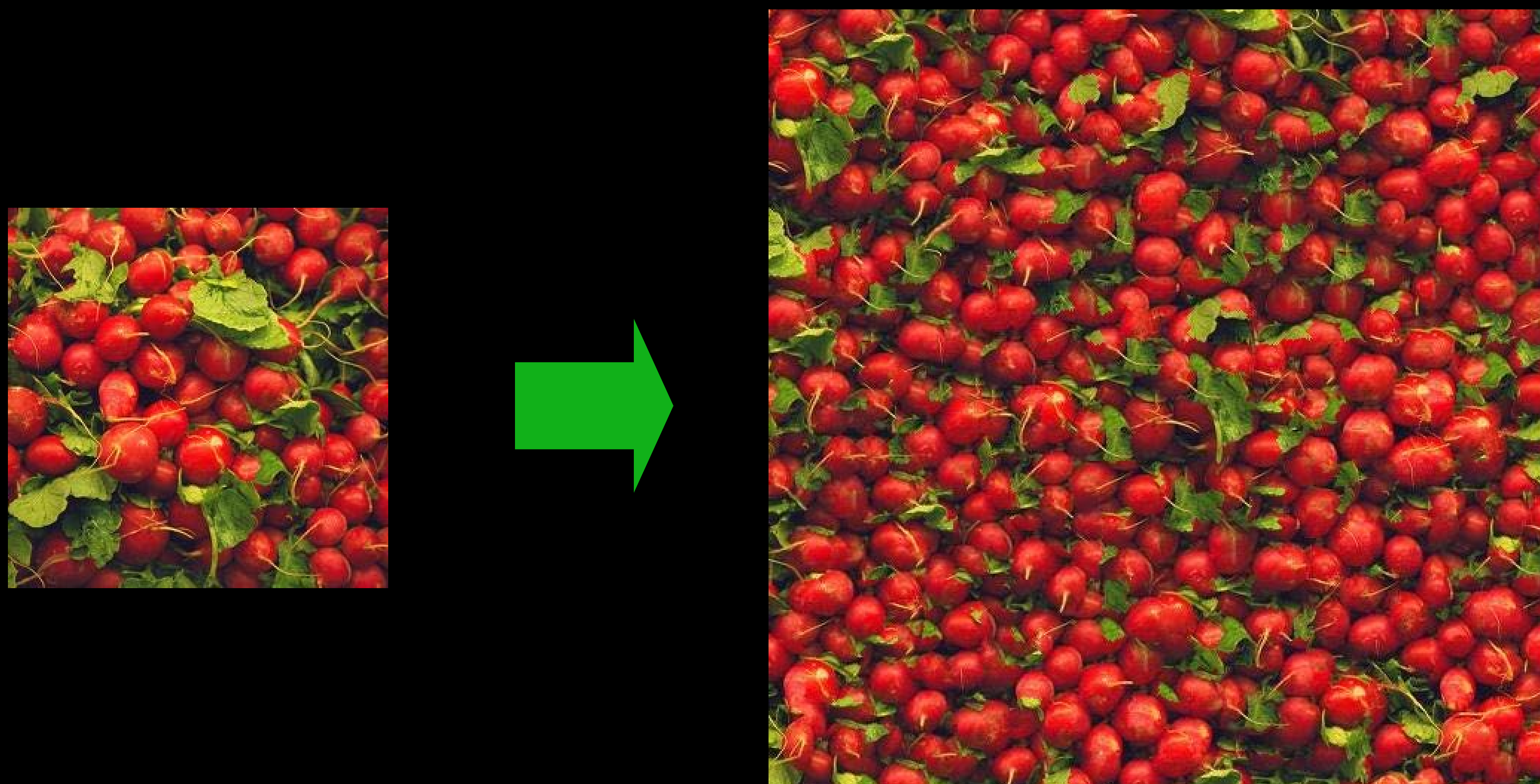
- Texture depicts spatially repeating patterns
- Many natural phenomena are textures

radishes

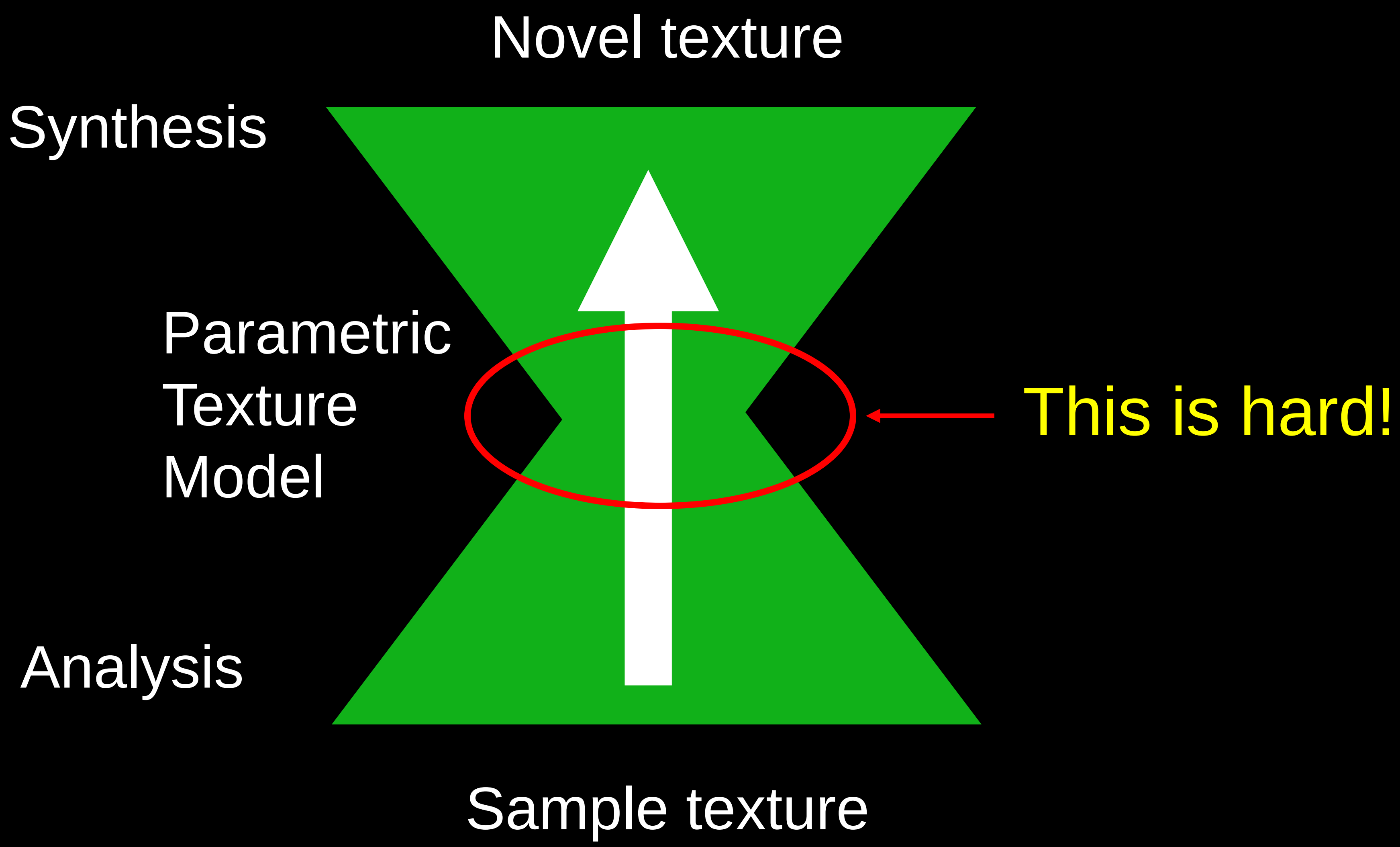
rocks

yogurt

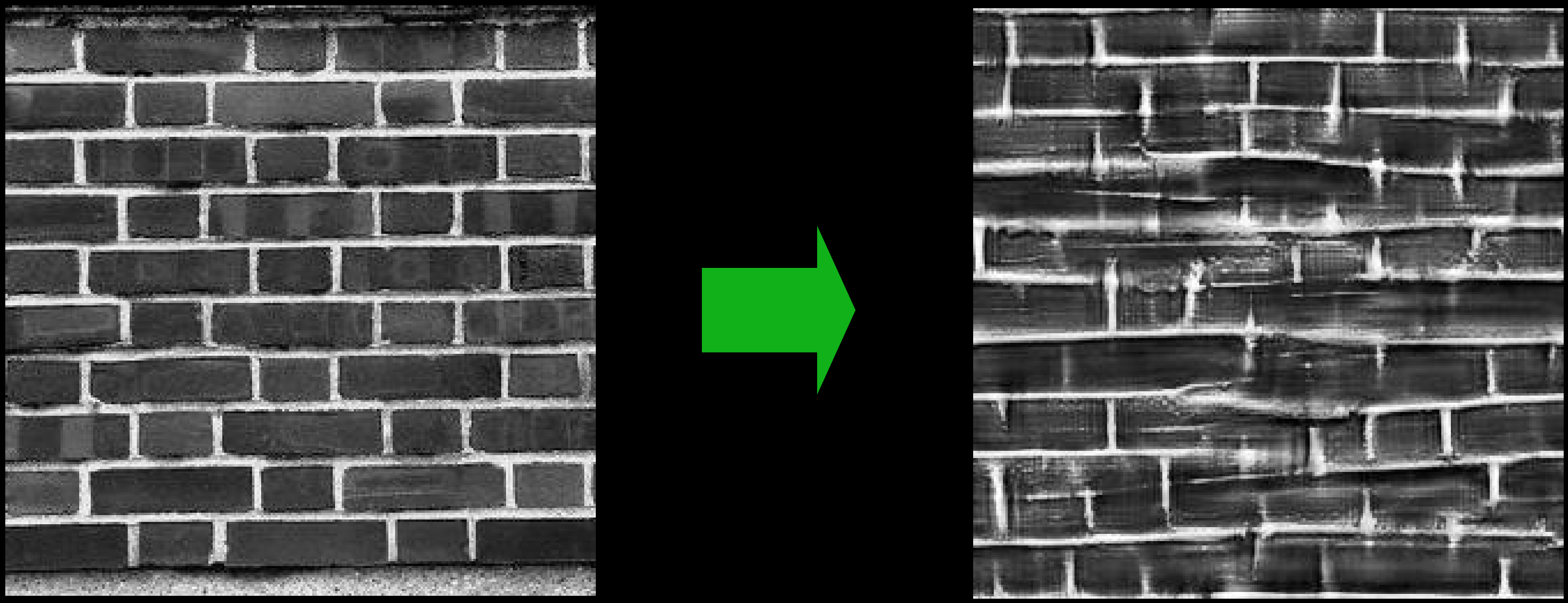
Texture Synthesis



Classical Texture Synthesis



Throwing away too much too soon?

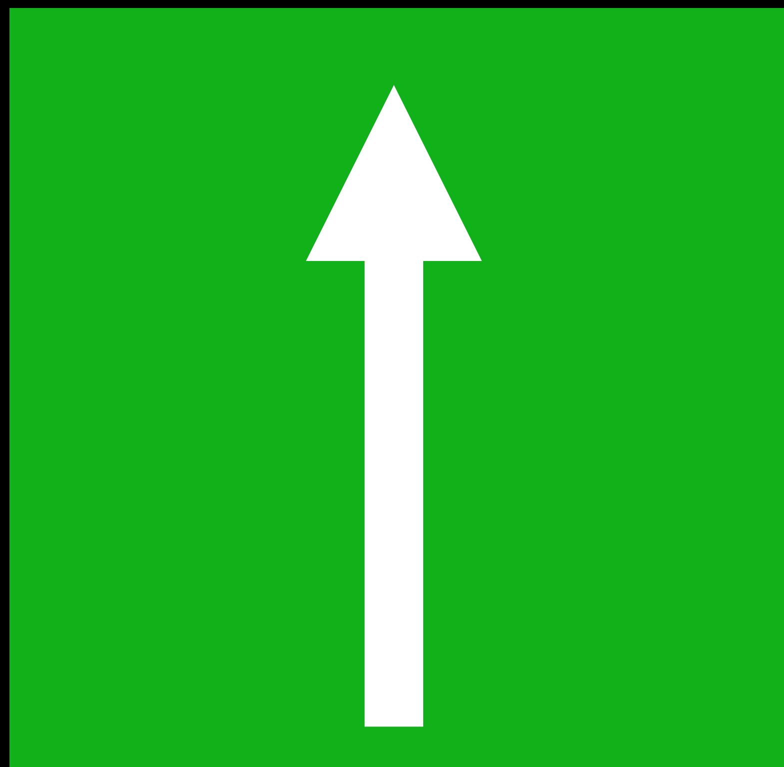


input texture synthesized texture

Non-parametric Approach

Novel texture

Synthesis



Analysis

Sample texture

Motivation from Language

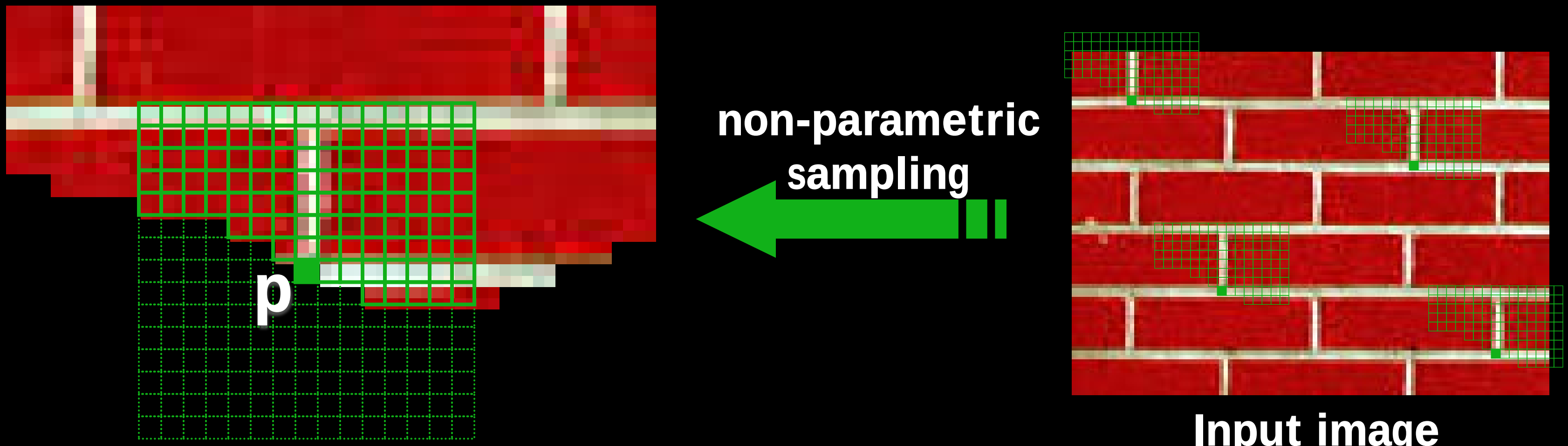
- [Shannon,’48] proposed a way to generate English-looking text using N-grams:
  - Assume a generalized Markov model
  - Use a large text to compute prob. distributions of each letter given N-1 previous letters
  - Starting from a seed repeatedly sample this Markov chain to generate new letters
  - Also works for whole words

WE NEED TO EAT CAKE

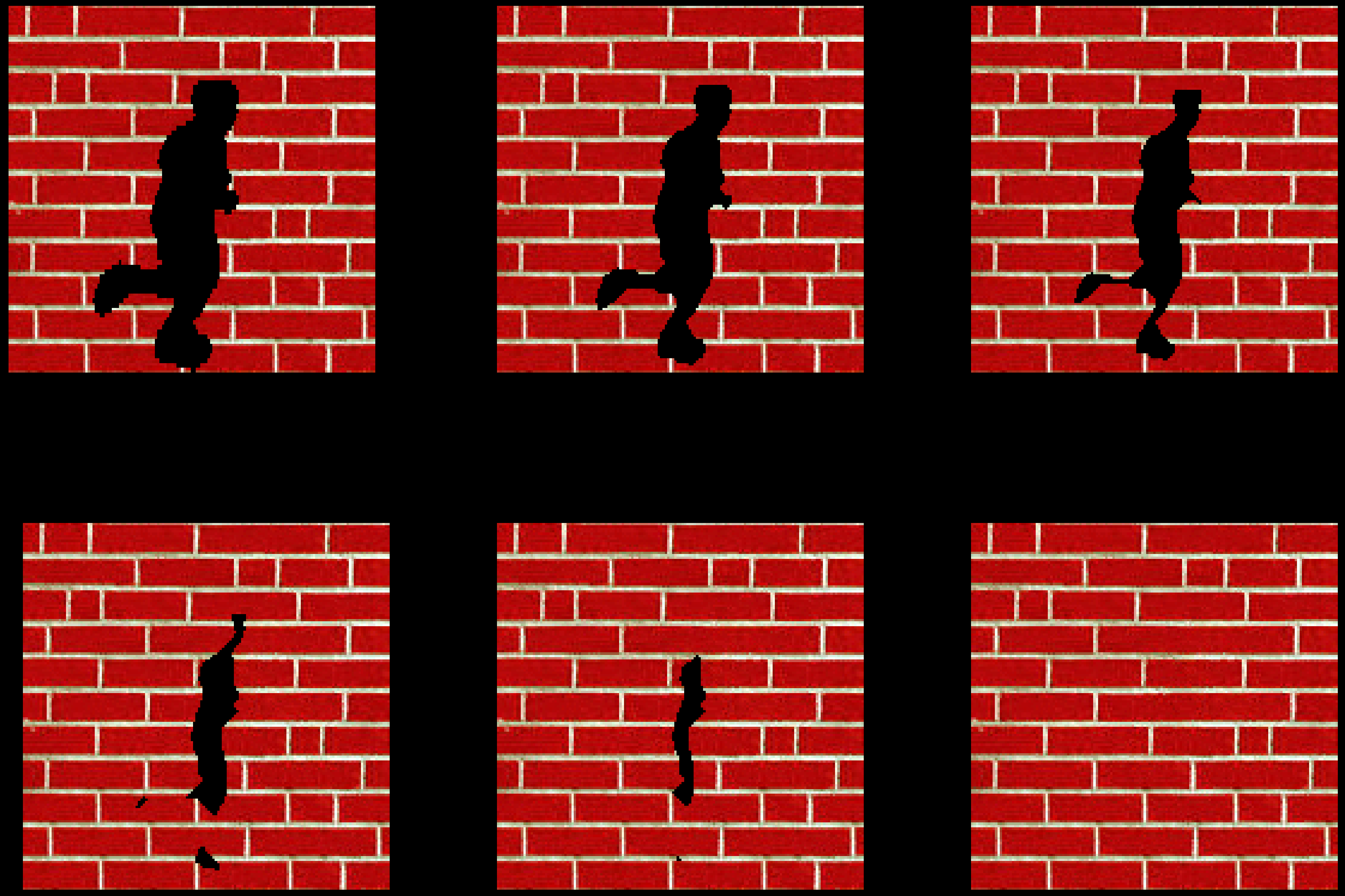
Mark V. Shaney (Bell Labs)

- Results (using `alt.singles` corpus):
  - *"As I've commented before, really relating to someone involves standing next to impossible."*
  - *"One morning I shot an elephant in my arms and kissed him."*
  - *"I spent an interesting evening recently with a grain of salt"*
- Notice how well local structure is preserved!
  - Now, instead of letters let's try pixels…

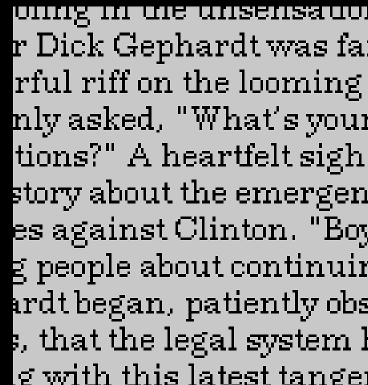
[Efros & Leung, ’99]

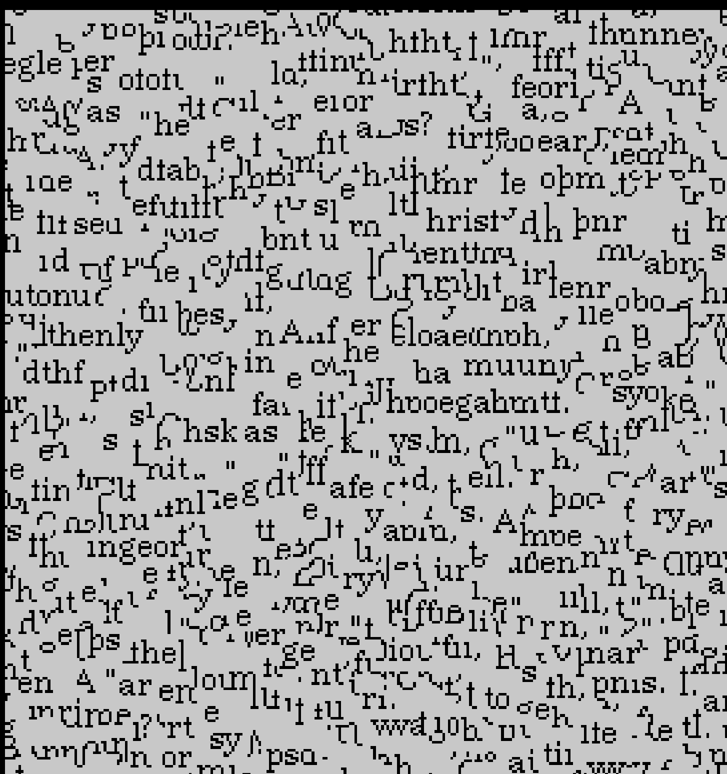


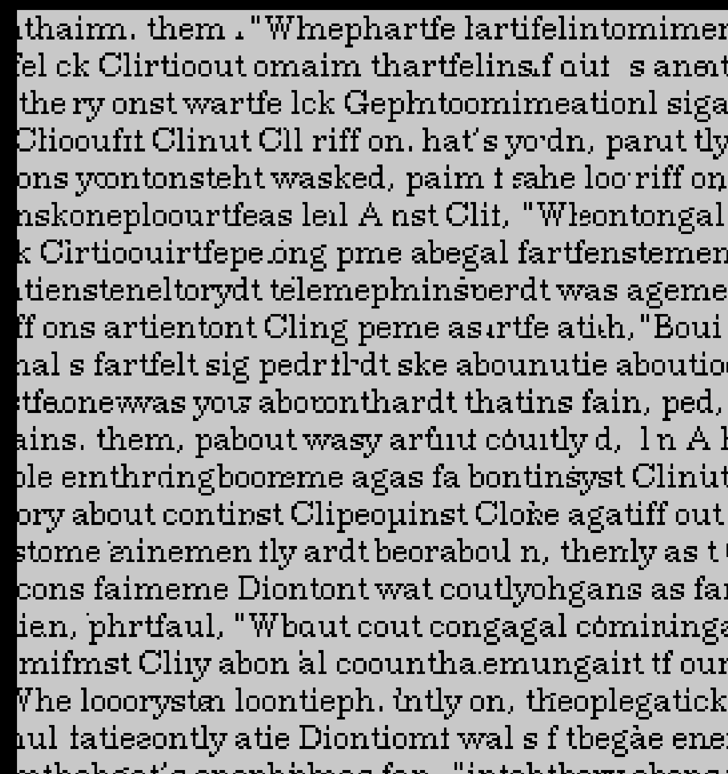
Texture Growing

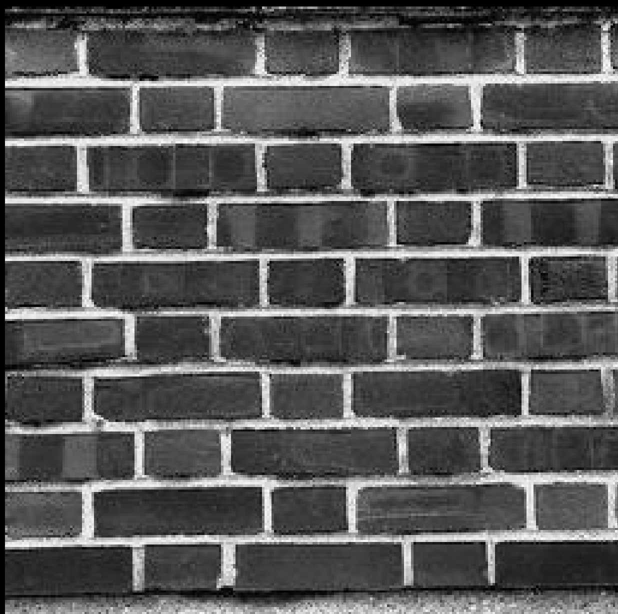


Homage to Shannon



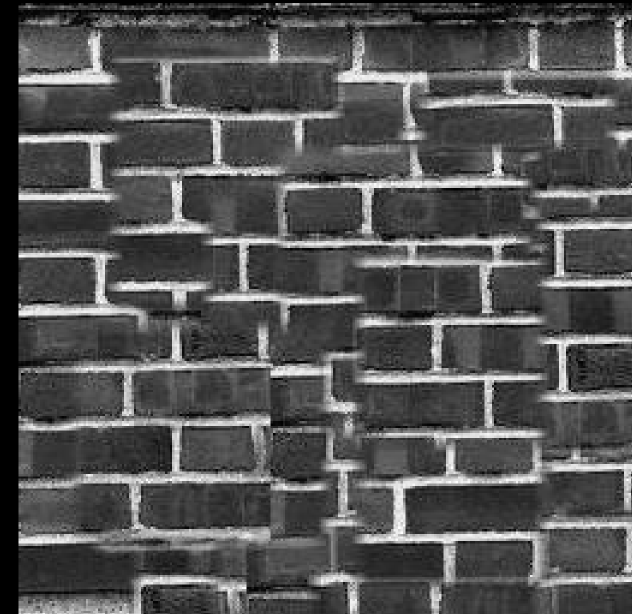




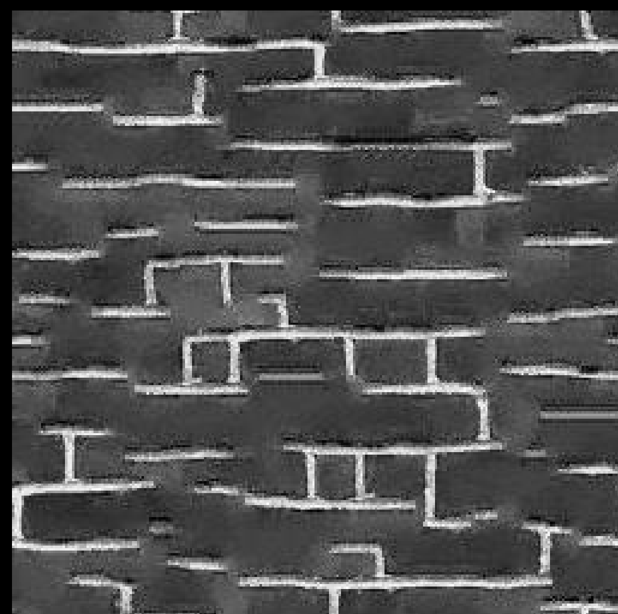


input image

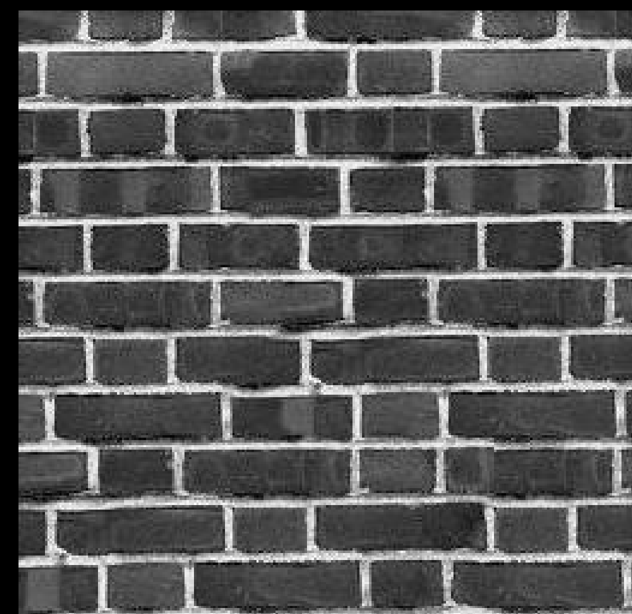
Portilla & Simoncelli



Xu, Guo & Shum



Wei & Levoy

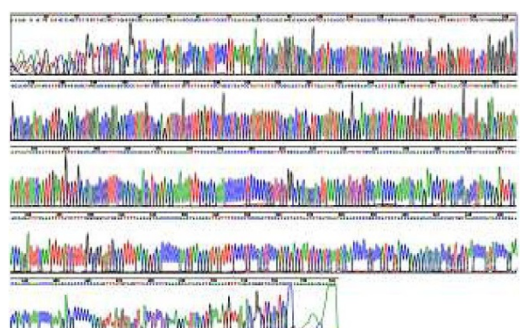


Our algorithm

Two Kinds of Things in the World

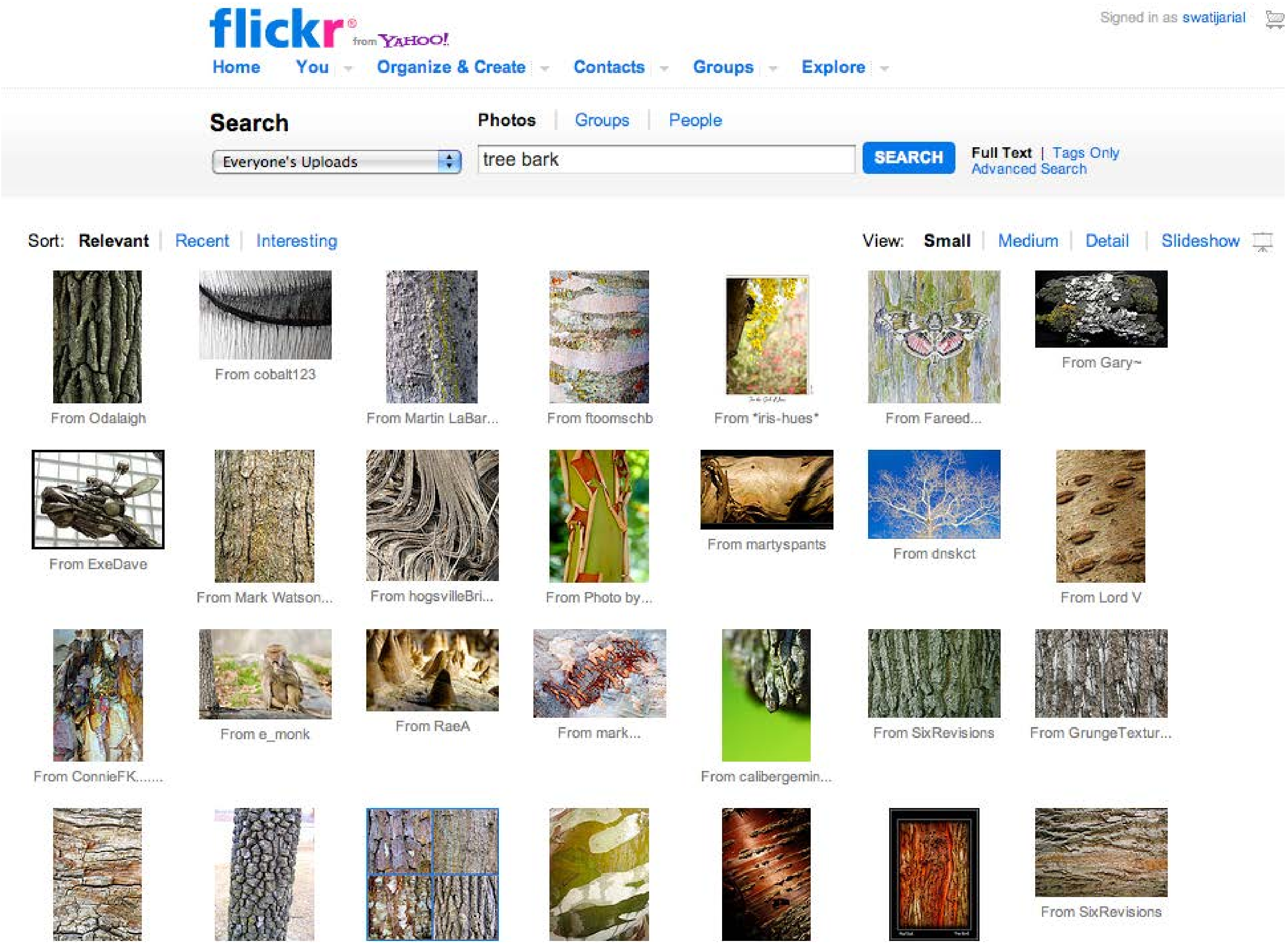
Navier-Stokes Equation

$$\frac{\partial \mathbf{u}}{\partial t} = -\left(\mathbf{u} \cdot \nabla\right)\mathbf{u} + v\nabla^2\mathbf{u} - \frac{1}{d}\nabla p + \mathbf{f}$$



+ weather
+ location
+ …

Lots of data available



“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, genetics, economics, etc.
- Enter: The Magic of **Big Data**
  - Great advances in several fields:
    - e.g. speech recognition, machine translation, Google

- A.I. for the postmodern world:
  - all questions have already been answered…many times, in many ways
  - Google is dumb, the "intelligence" is in the data

The Good News

Really stupid algorithms + Lots of Data

= *“Unreasonable Effectiveness”*

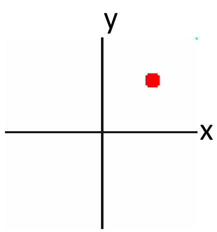
The Bad News

Visual Data is much more difficult

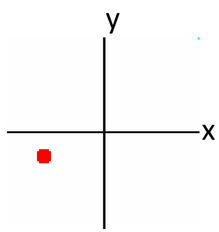
- text:
  - clean, segmented, compact, 1D
- Visual data:
  - Noisy, unsegmented, high entropy, 2D/3D

Distance Metrics

CLIME - CRIME = hamming distance of 1 letter



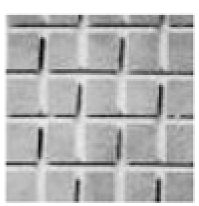
-



= Euclidian distance of 5 units

-

= Grayvalue distance of 50 values



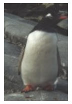
-

= ?

L2 norm says these are not similar

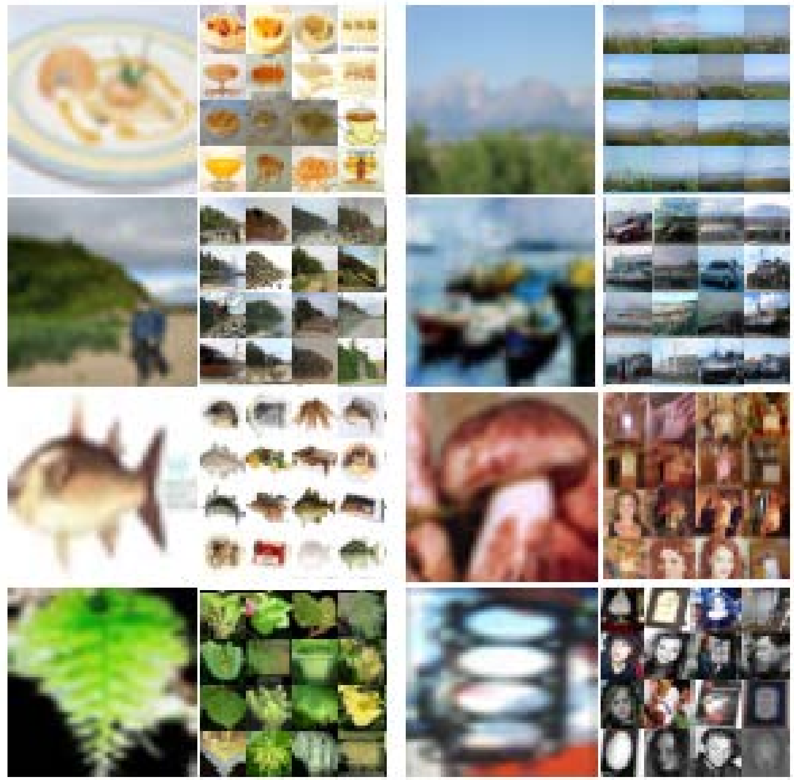
?

Make them tiny!

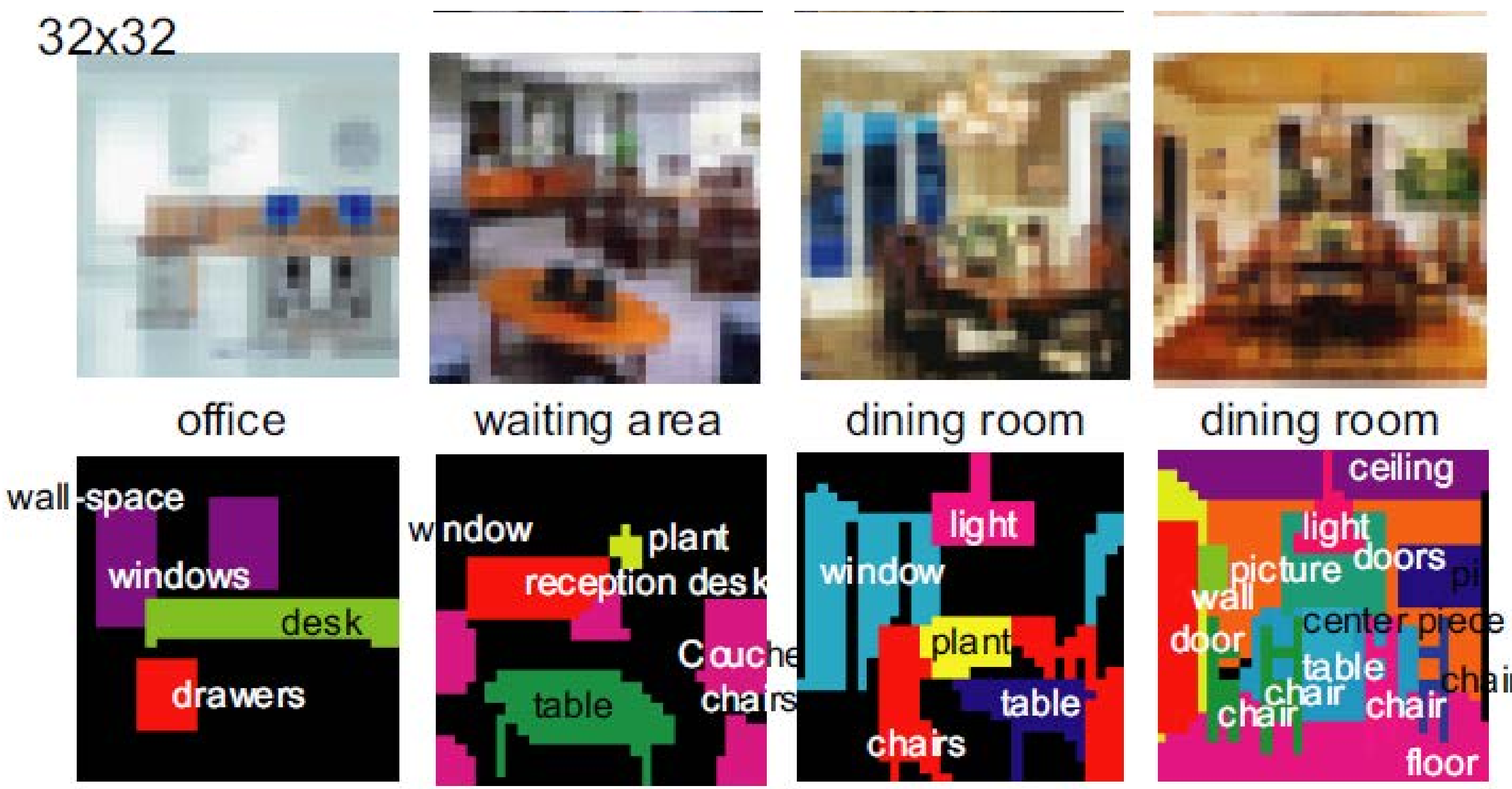


?

Lots of Tiny Images

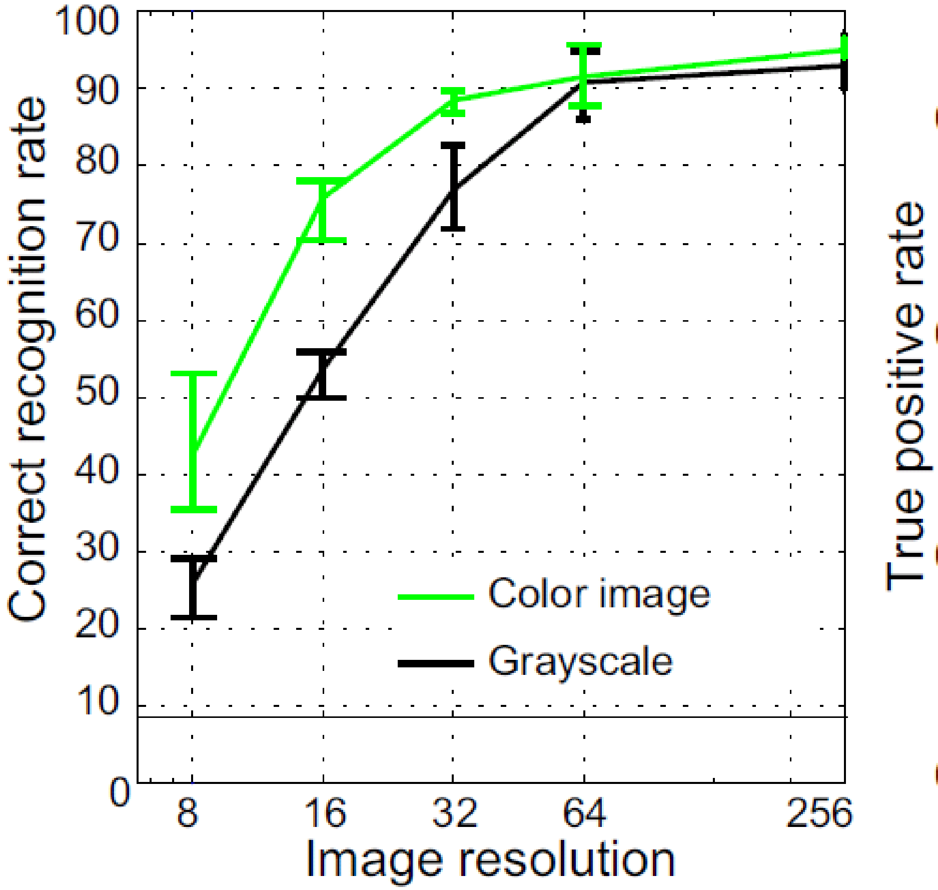


- 80 million tiny images: a large dataset for non-parametric object and scene recognition Antonio Torralba, Rob Fergus and William T. Freeman. PAMI 2008.



c) Segmentation of 32x32 images

Human Scene Recognition

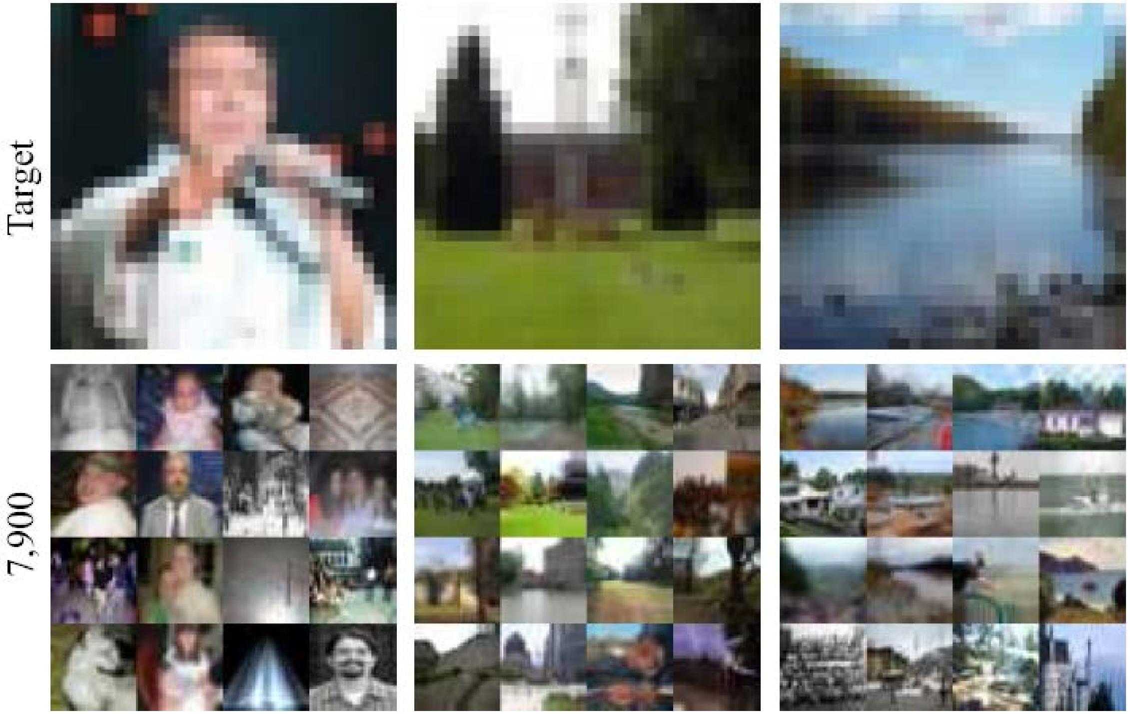


a) Scene recognition

Lots

Of

Images

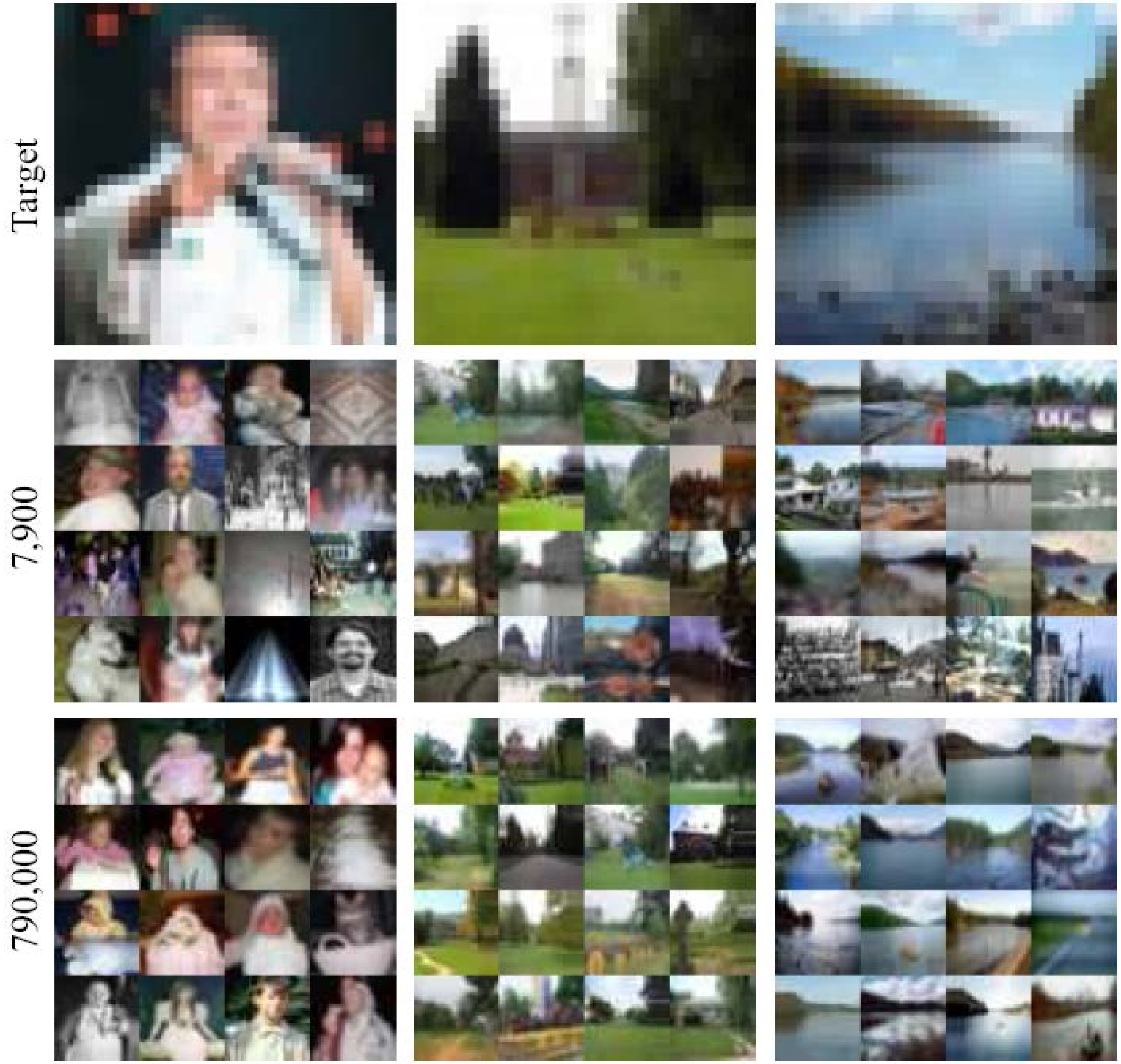


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

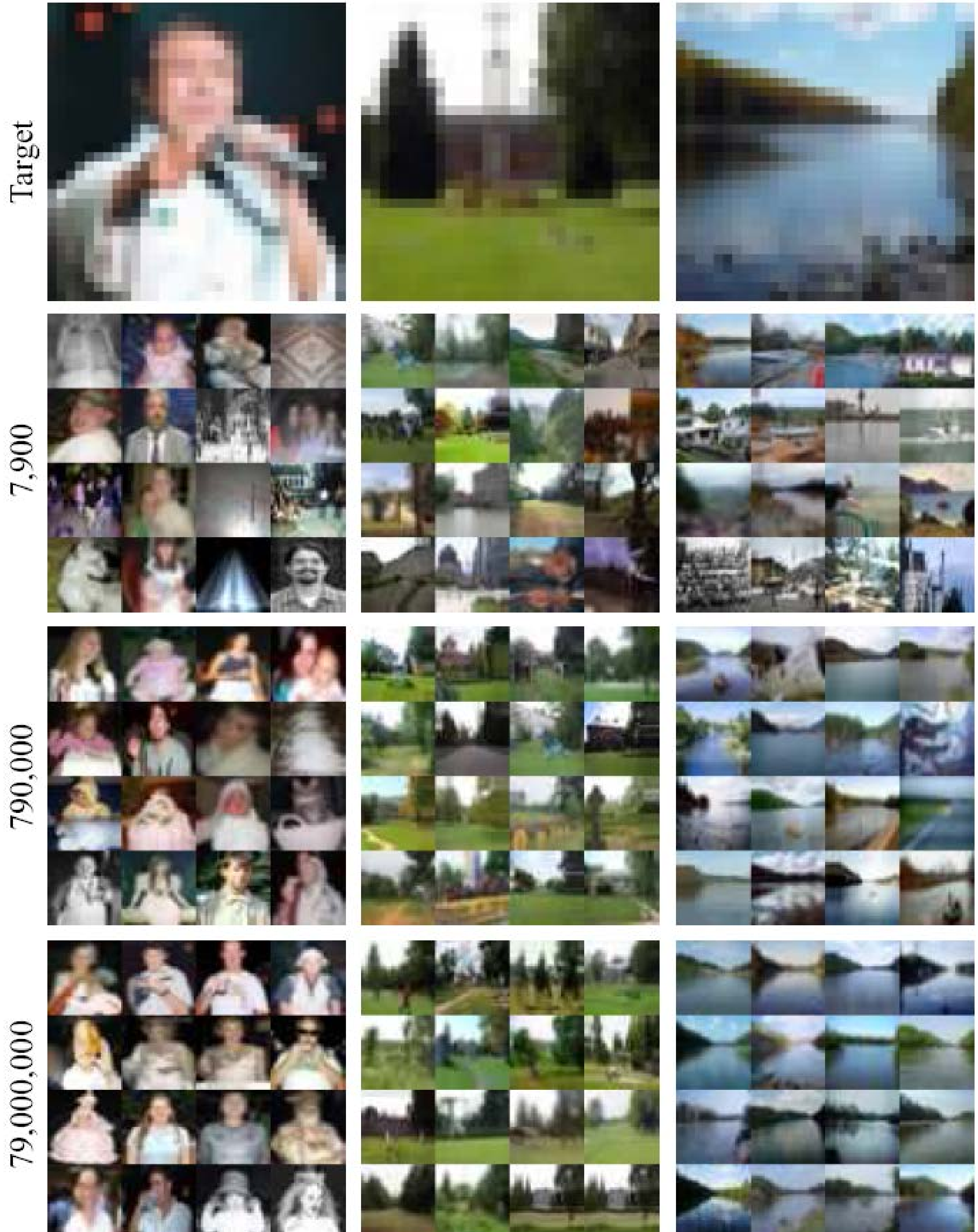


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

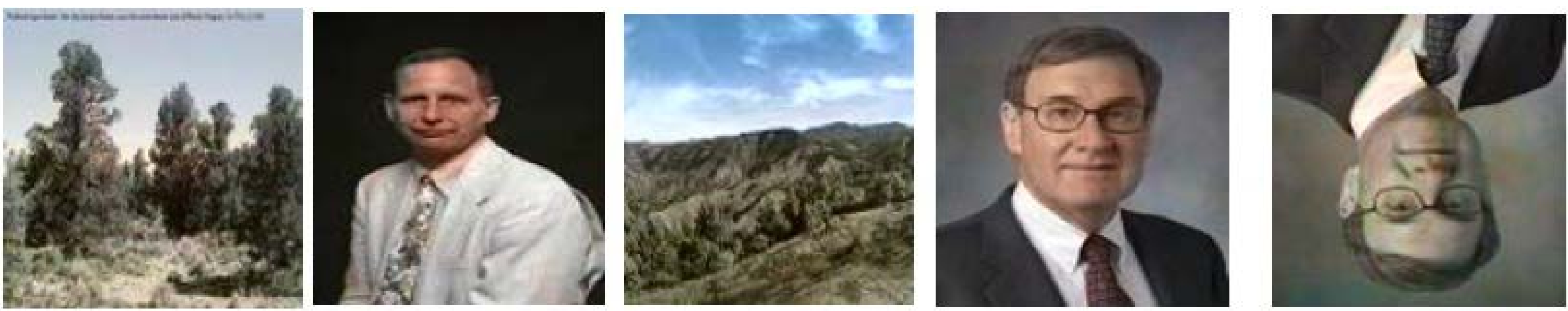
Images



Automatic Colorization

Grayscale input High resolution

Colorization of input using average



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

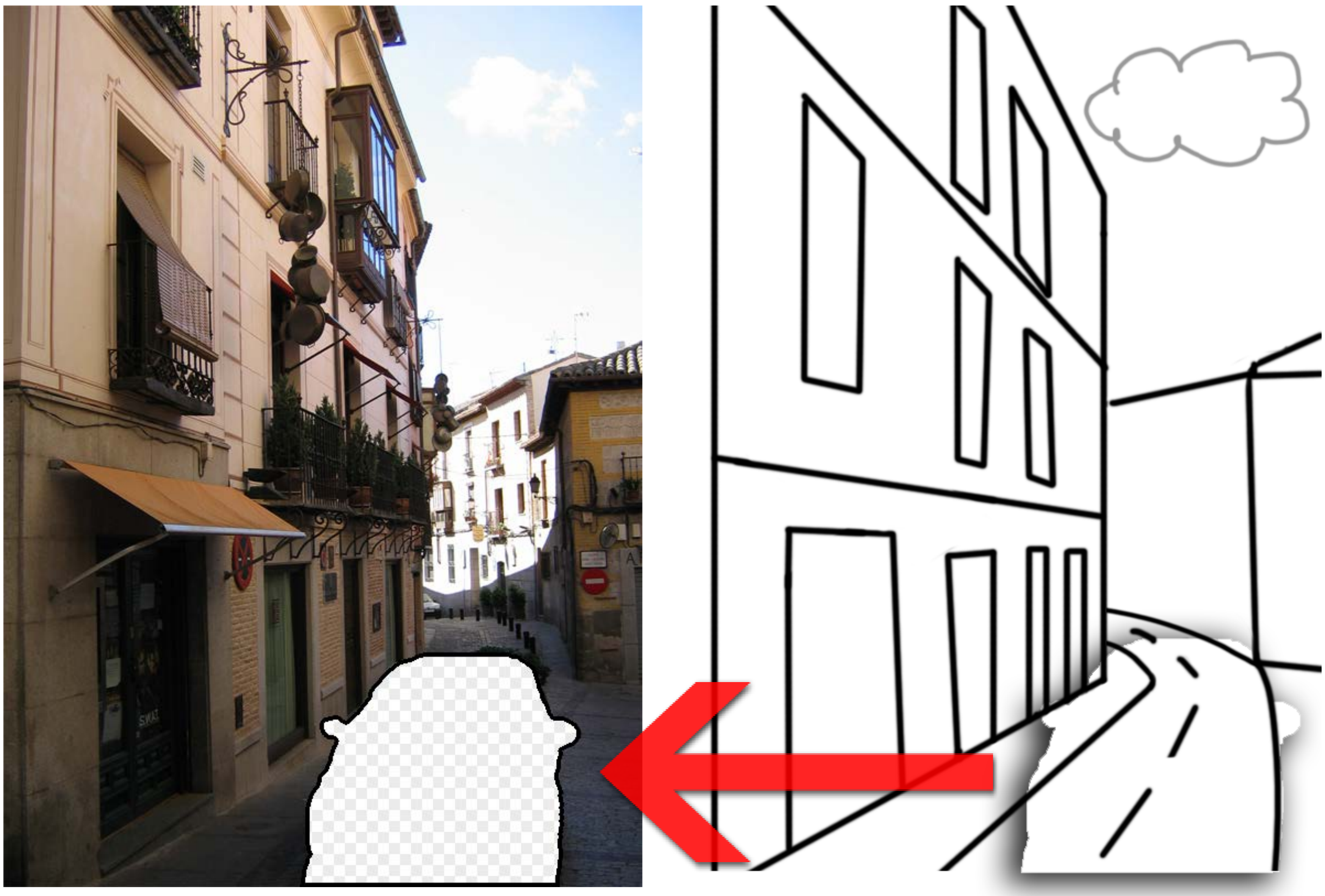
Not a pixel lover? No problem!

Let's match gradients

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

Efros and Leung result

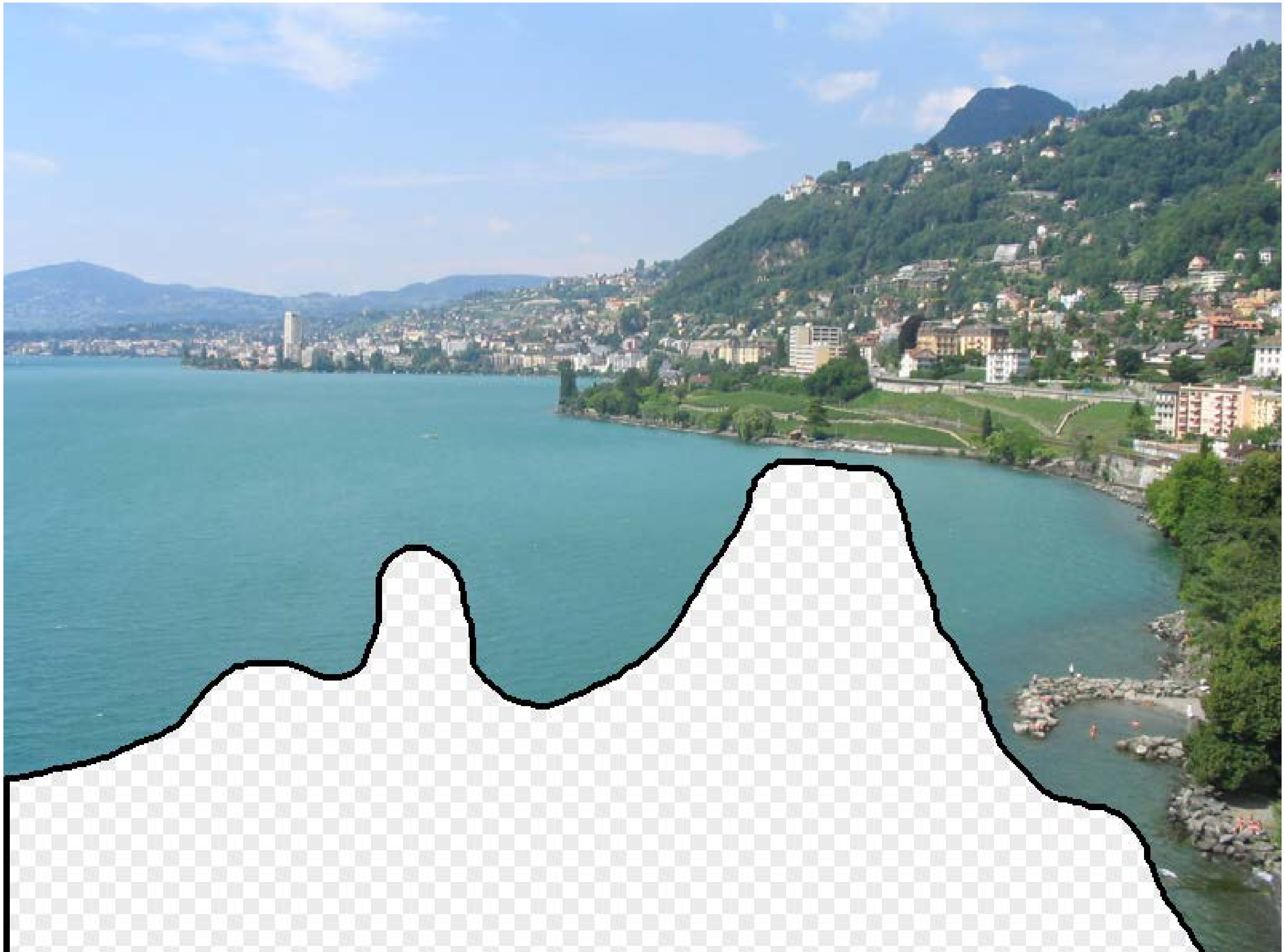
Scene Matching for Image Completion



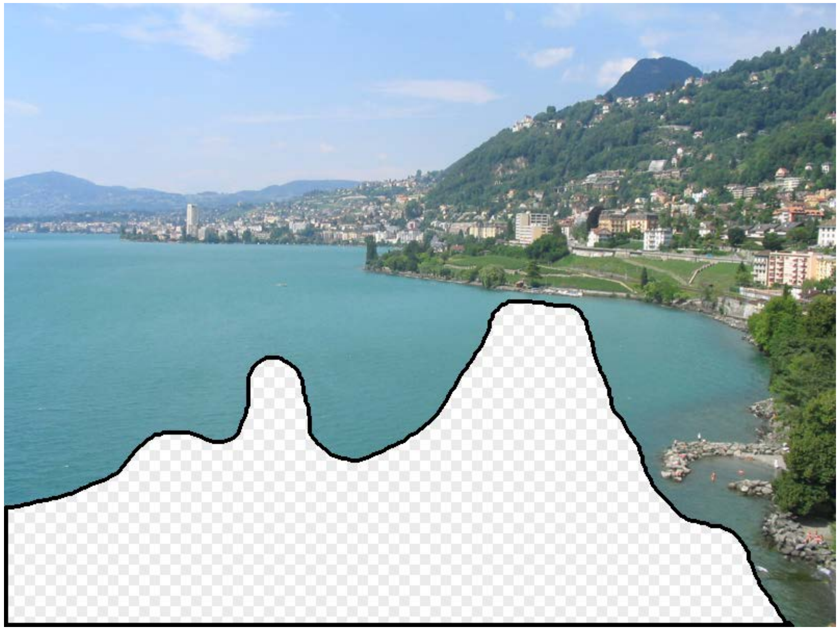
Scene Completion Result

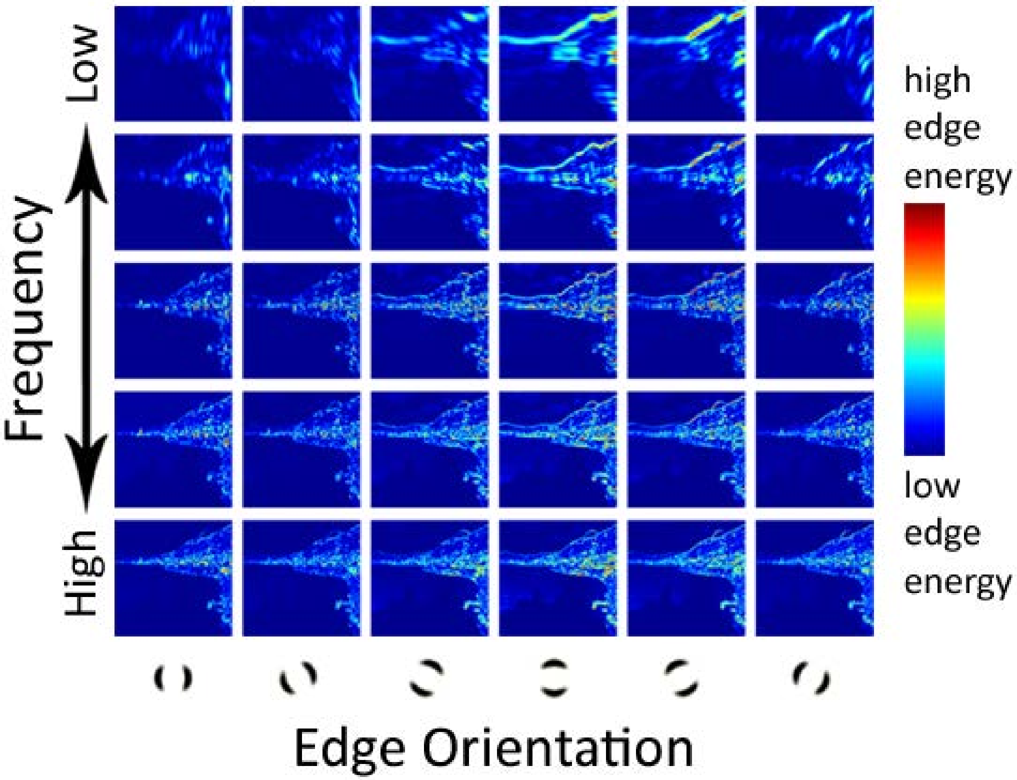
The Algorithm

Scene Matching

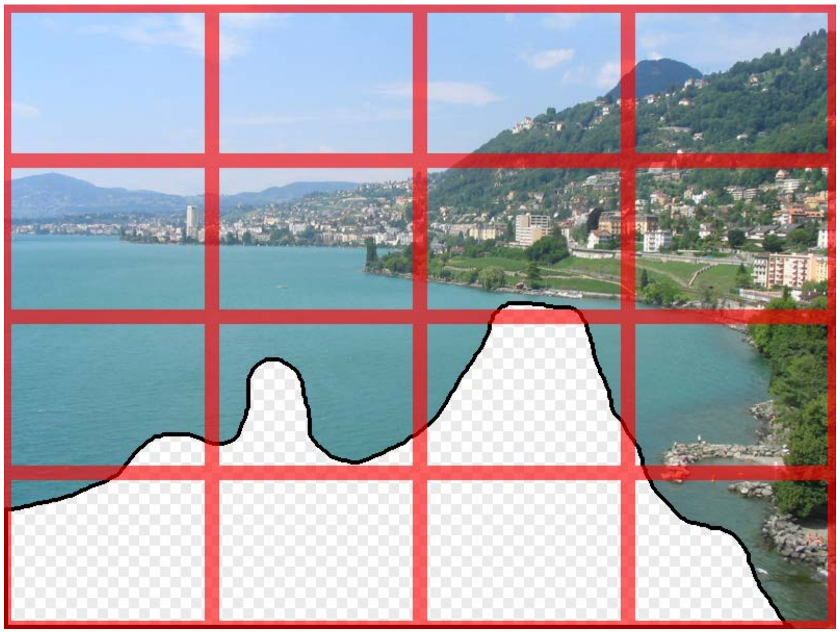


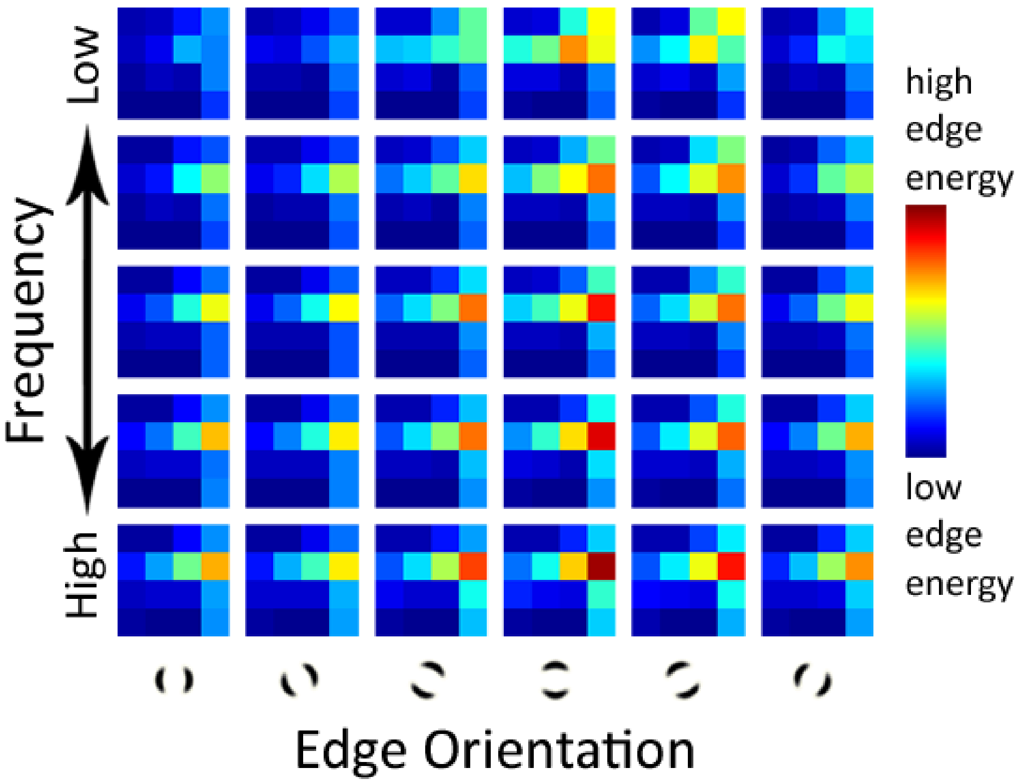
Scene Descriptor





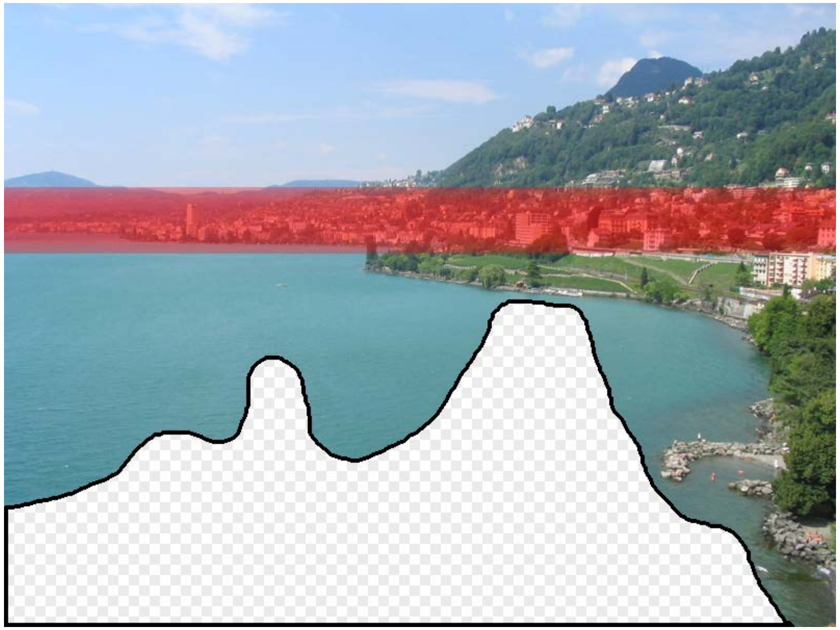
Scene Descriptor

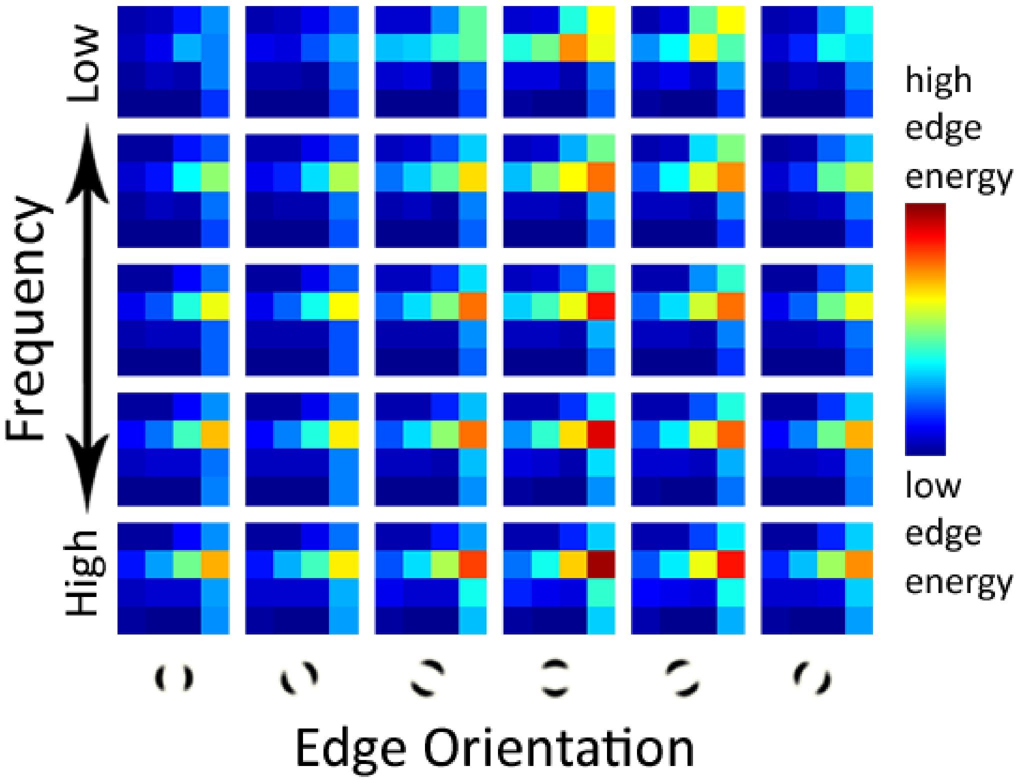




Scene Gist Descriptor
(Oliva and Torralba 2001)

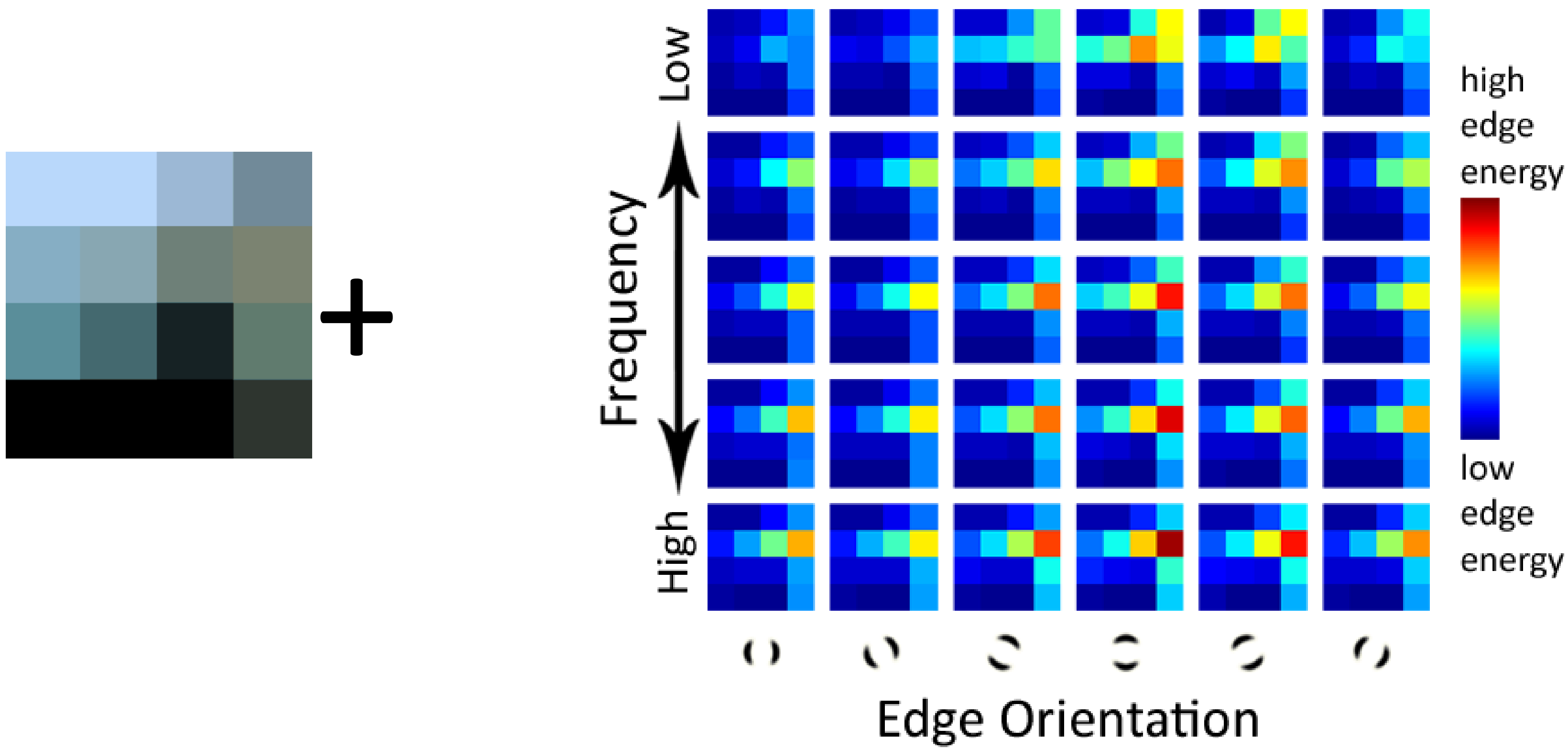
Scene Descriptor



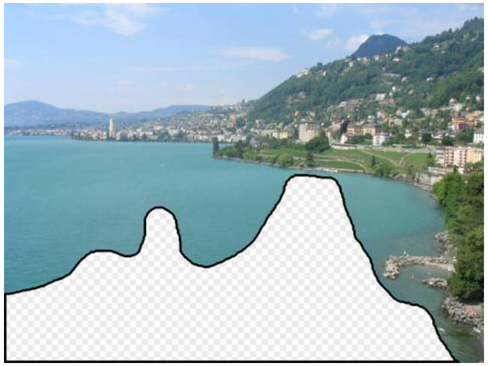


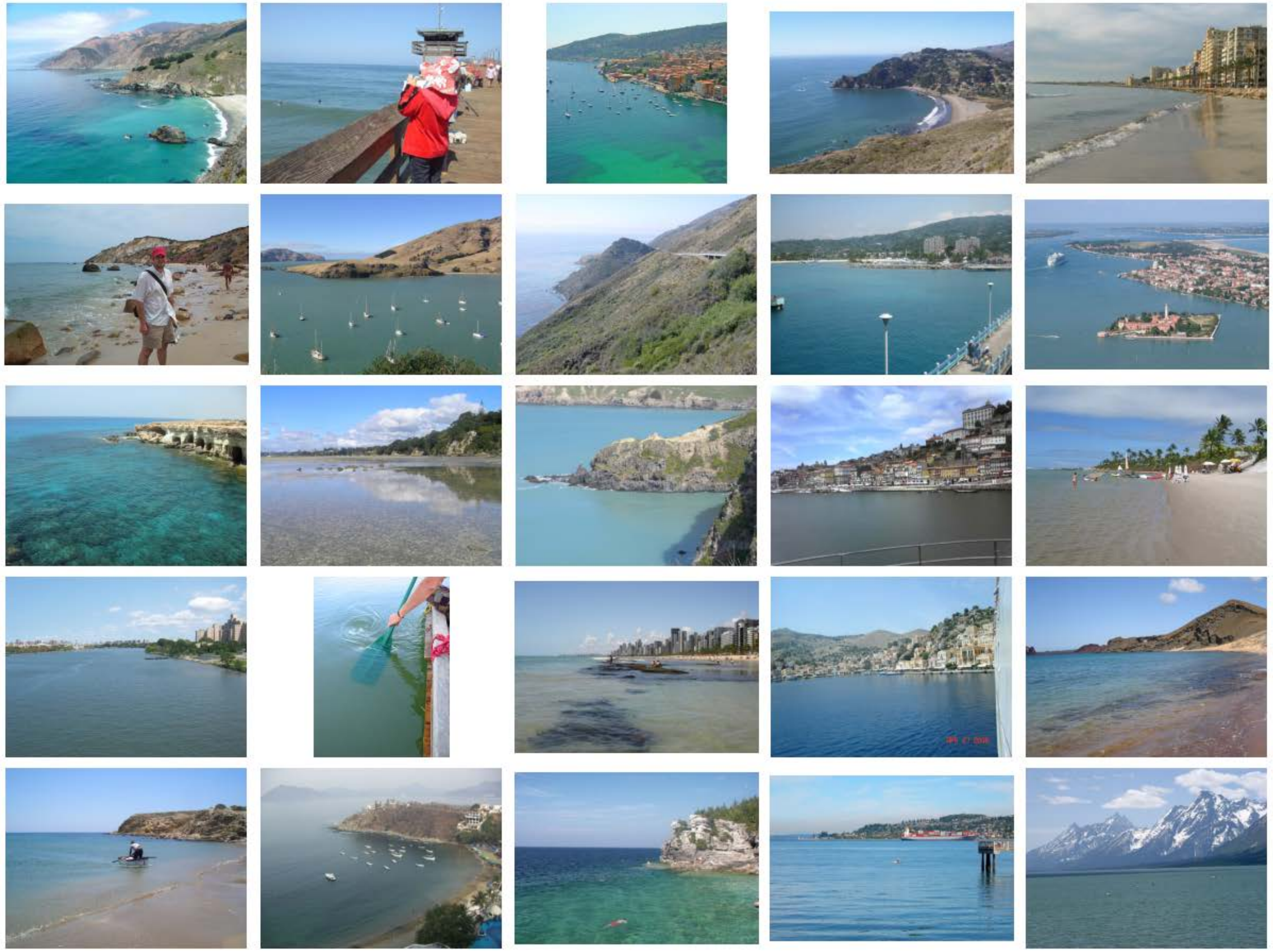
Scene Gist Descriptor
(Oliva and Torralba 2001)

Scene Descriptor



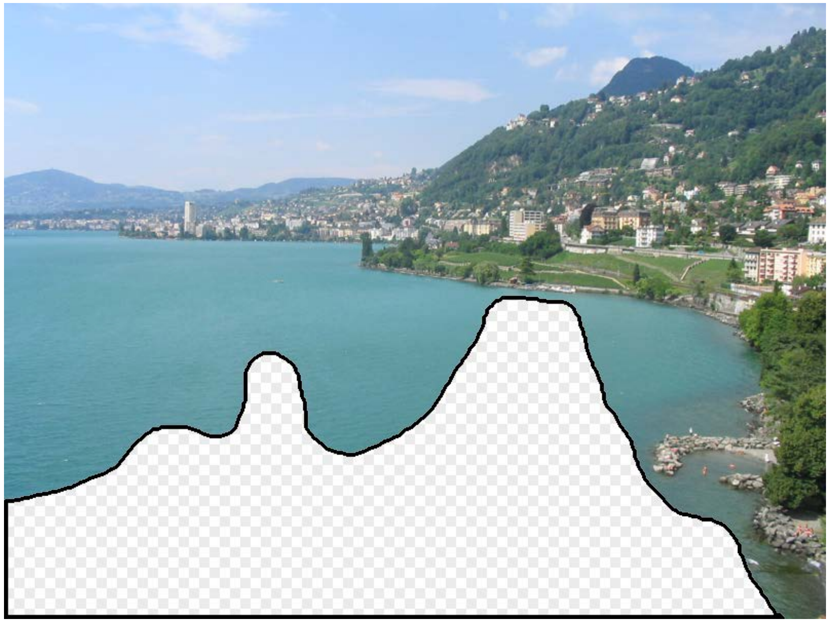
2 Million Flickr Images



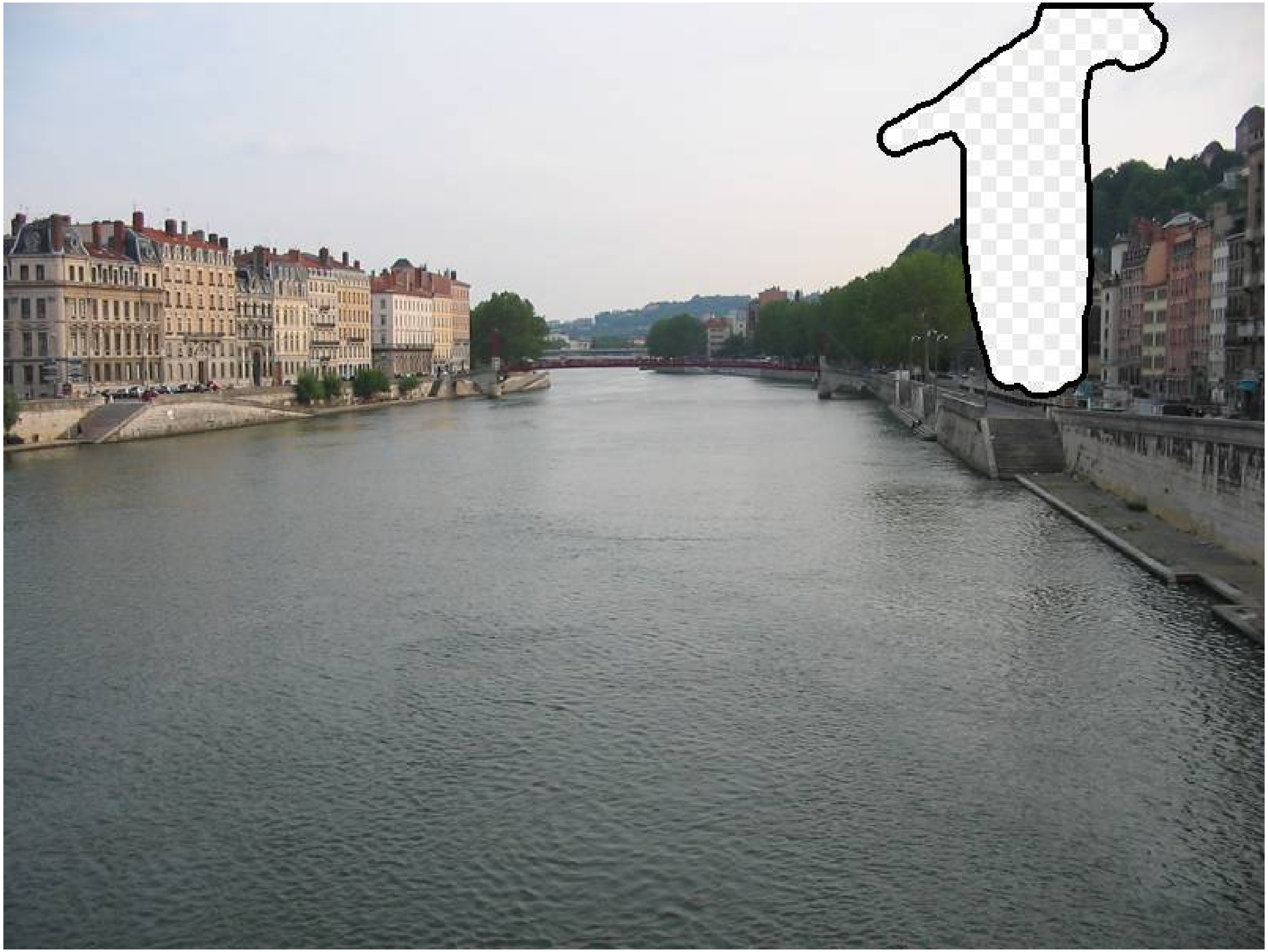


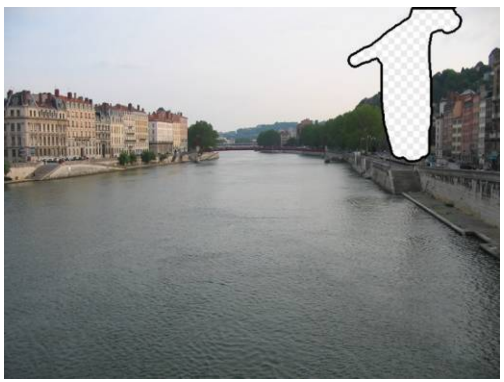
... 200 total

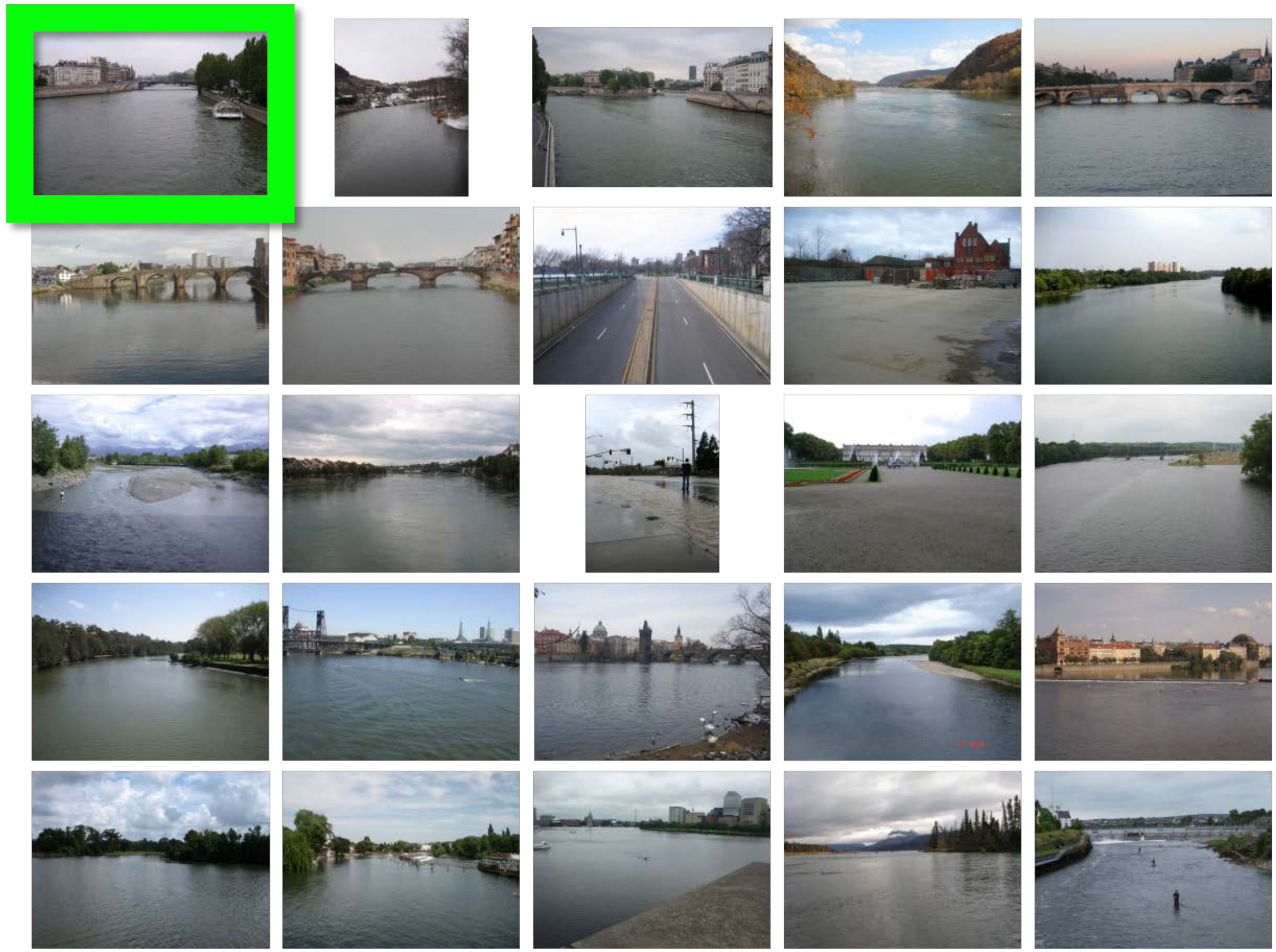
Context Matching



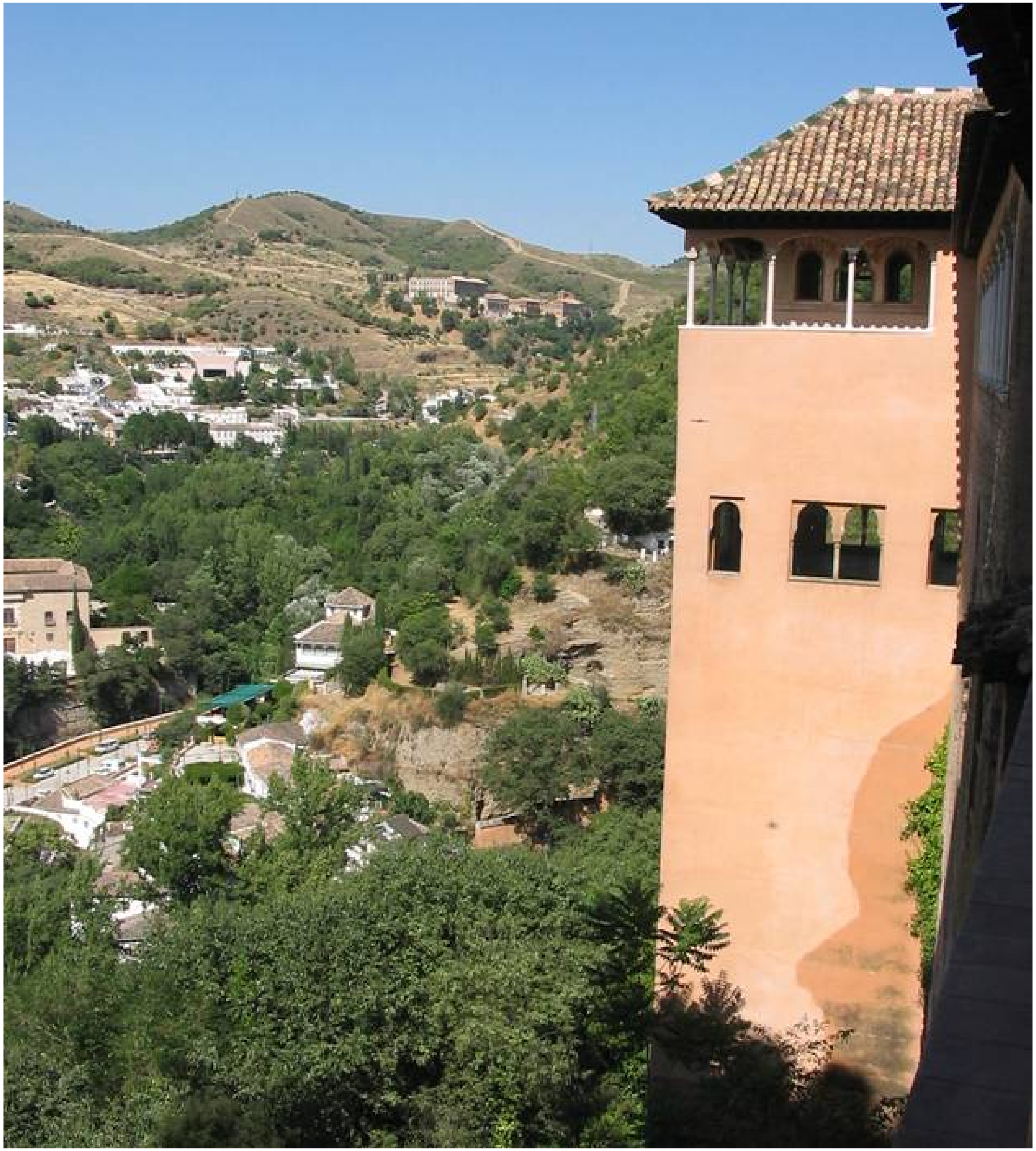
Graph cut + Poisson blending

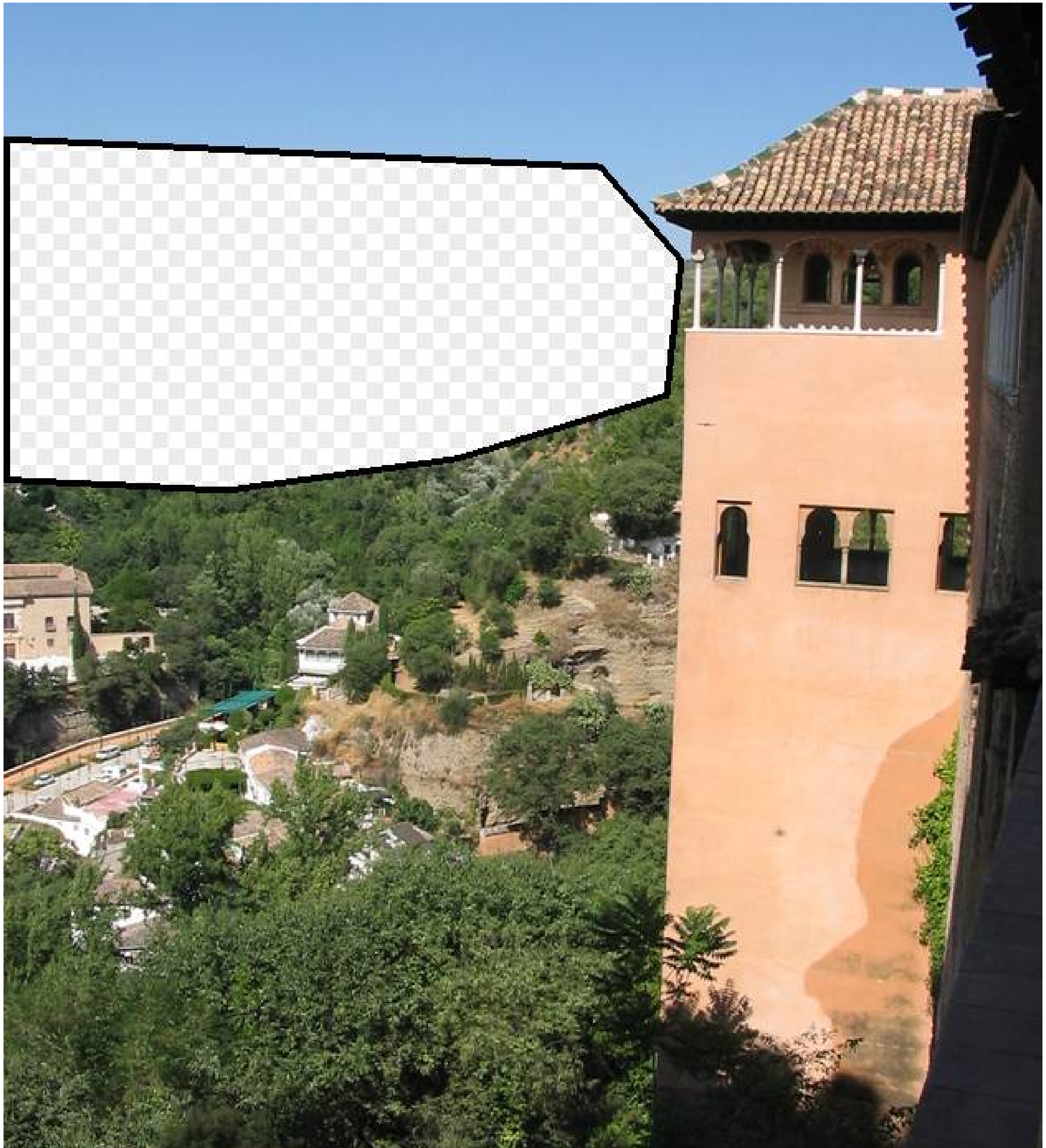


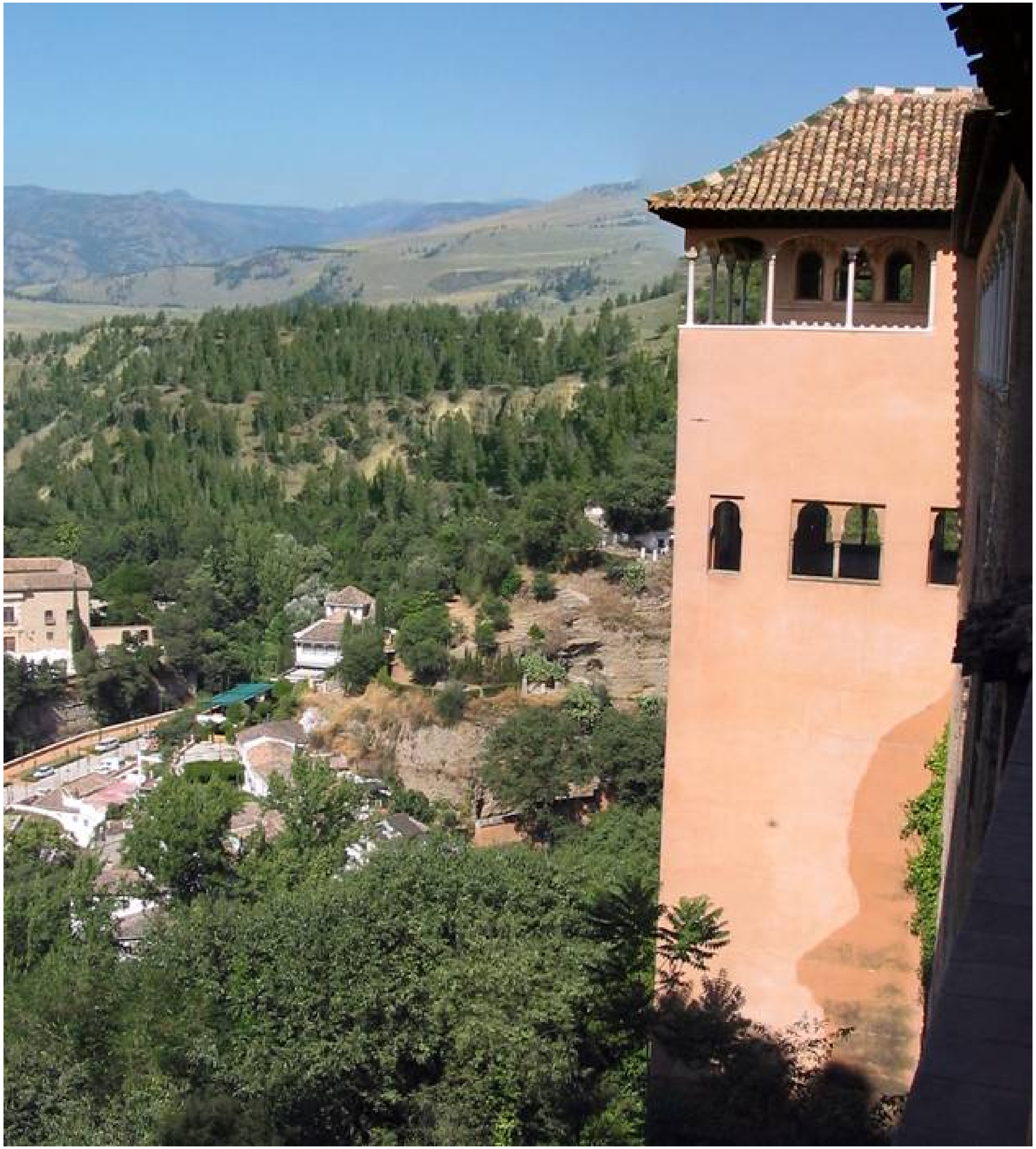




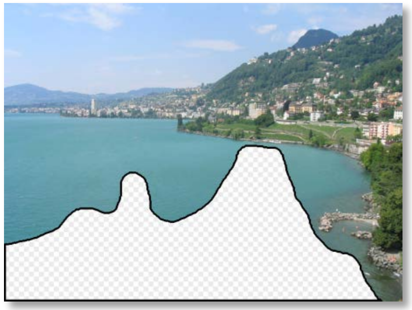
… 200 scene matches

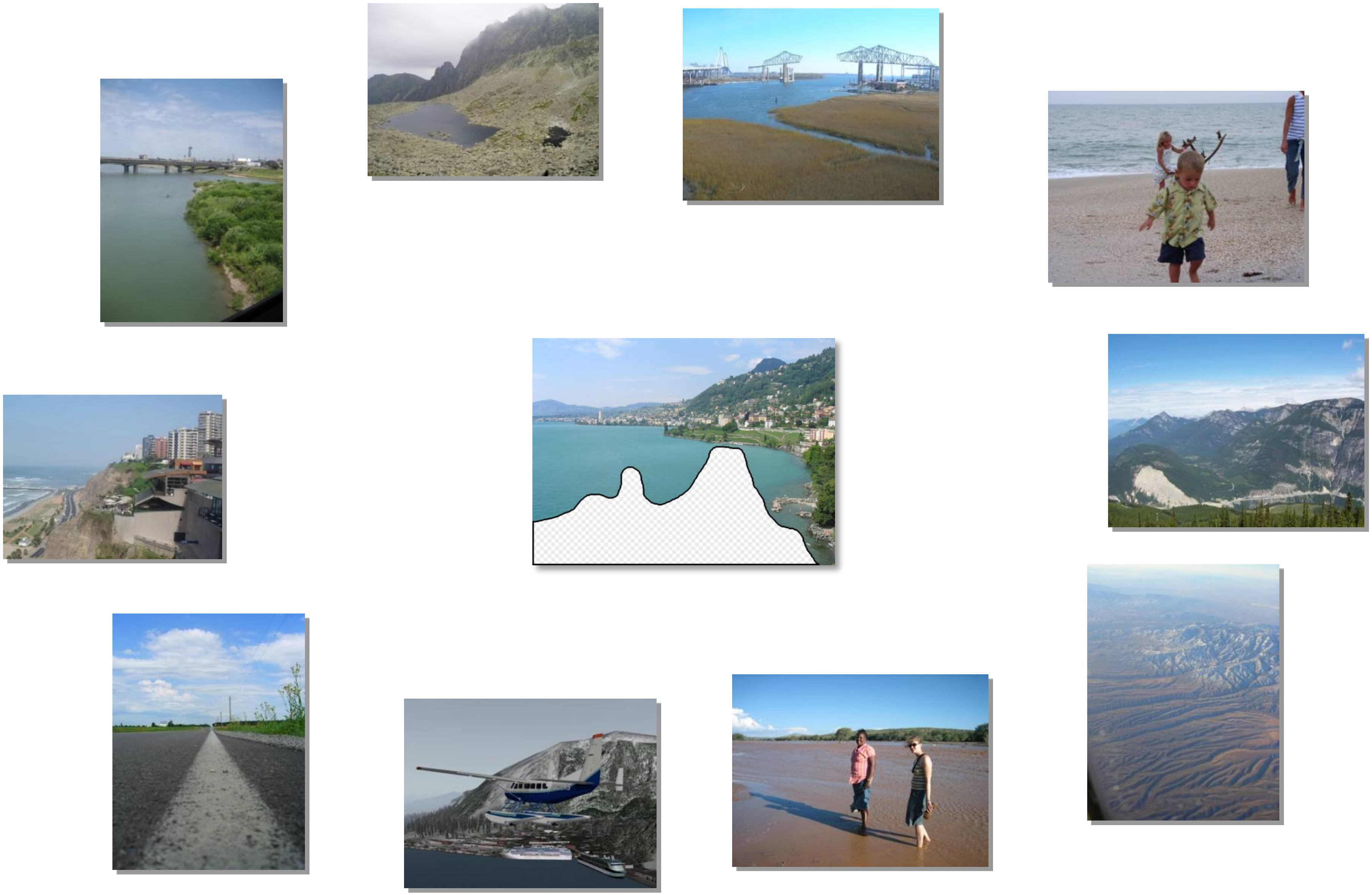




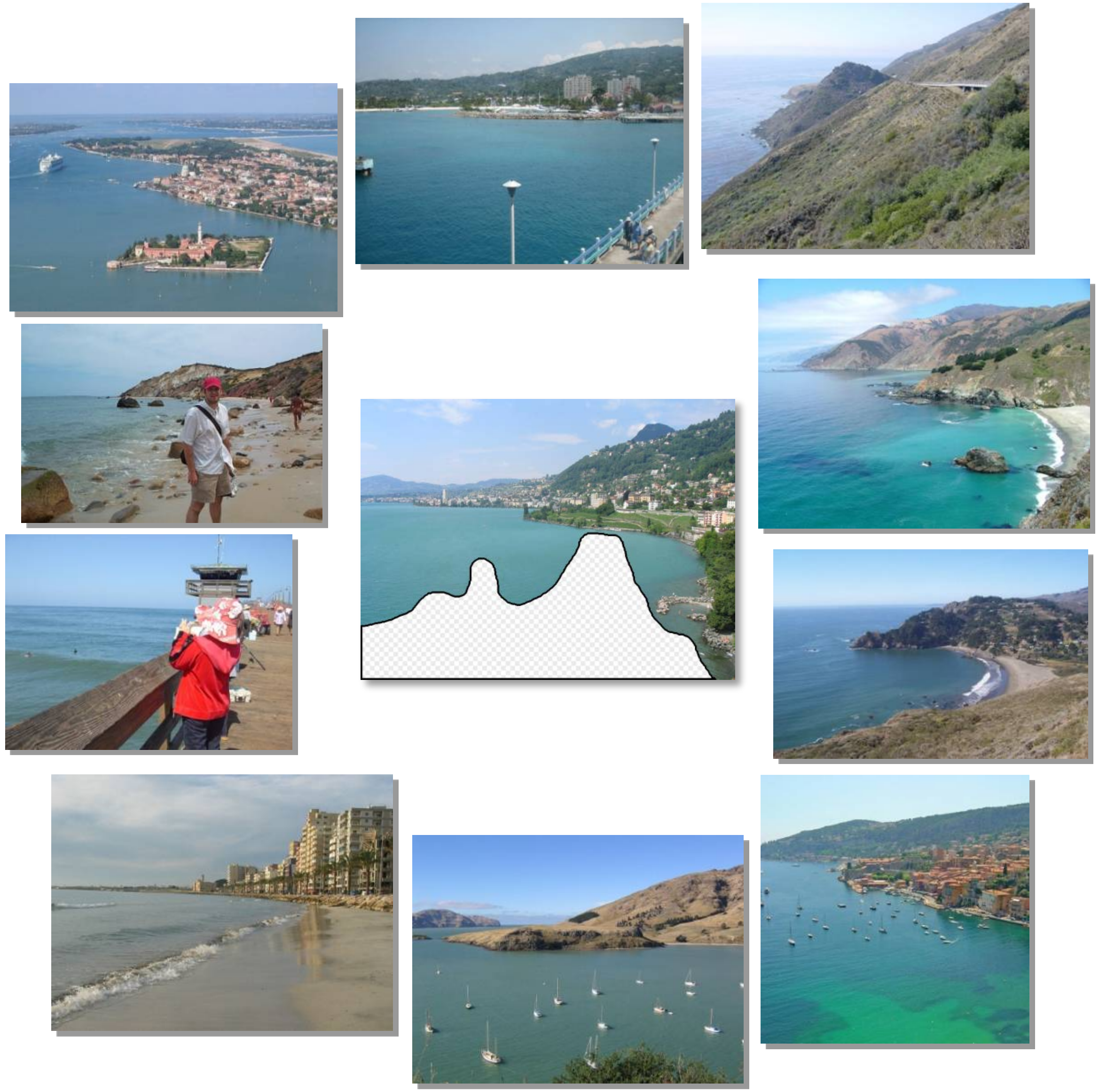


Why does it work?



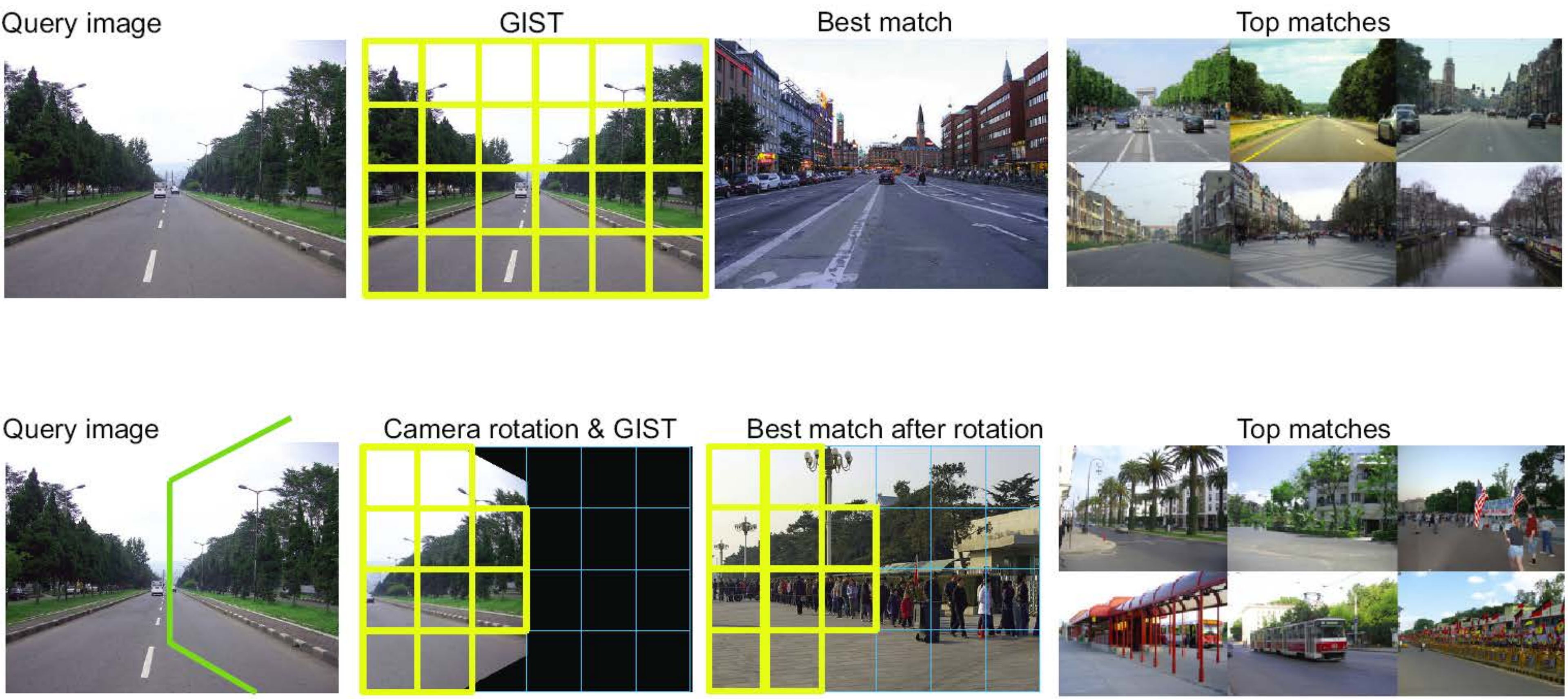


Nearest neighbors from a
collection of 20 thousand images



Nearest neighbors from a collection of 2 million images

Scene matching with camera transformations



Sivic, Kaneva, Torralba, Avidan, Freeman, Internet Vision Workshop, 2008
updated version to appear in Proceedings of the IEEE (2010)

Creating and exploring a large, photorealistic virtual space

Josef Sivic (INRIA/ENS), Biliana Kaneva (MIT), Antonio Torralba (MIT), Shai Avidan (Adobe) and Bill Freeman (MIT)

IEEE Workshop on Internet Vision, 2008

Cross-Domain Matching

“Local – bad

Global – good!”

(after Orwell)

Medici Fountain, Paris

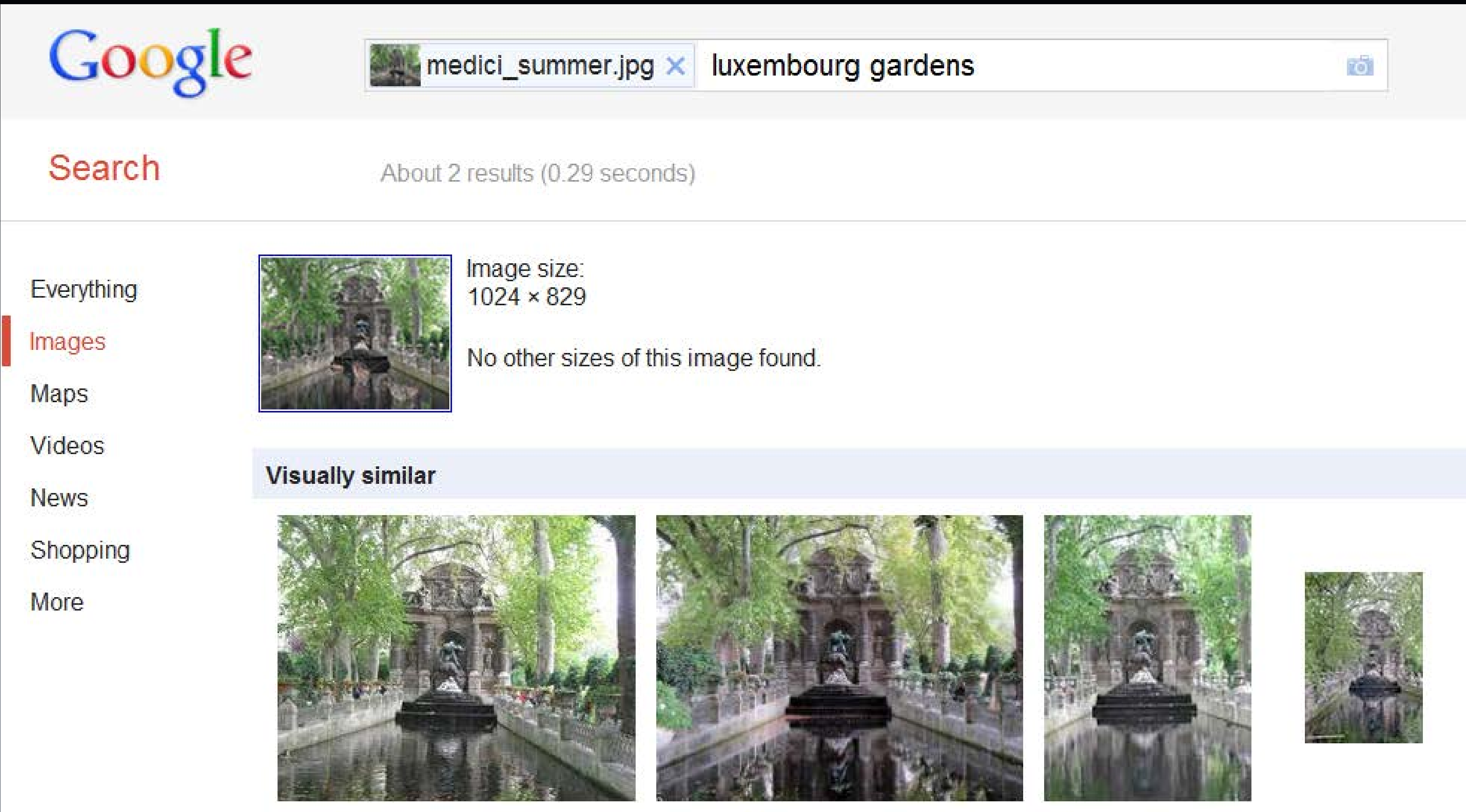
Google images

Search by image

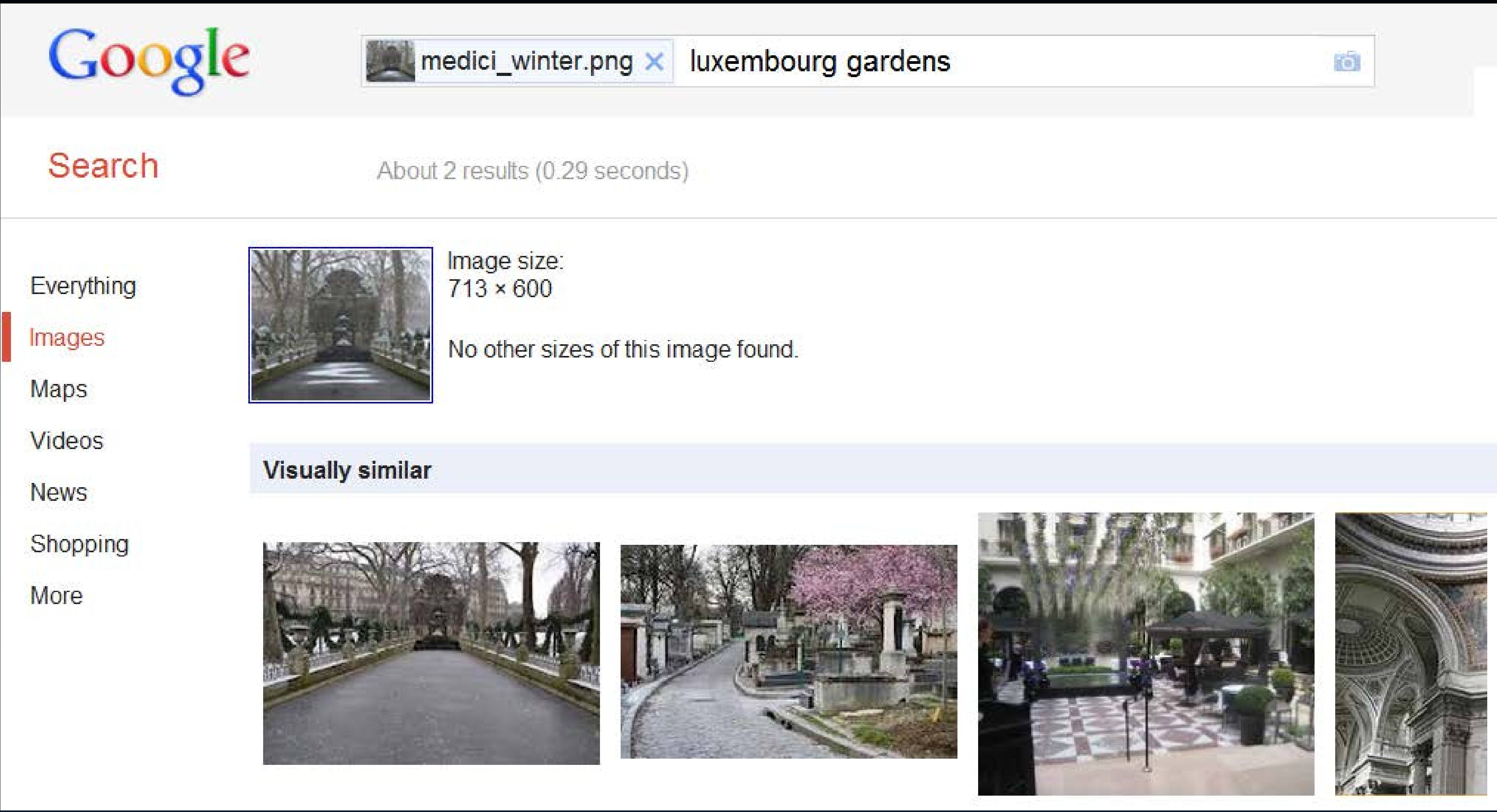
Drop image here

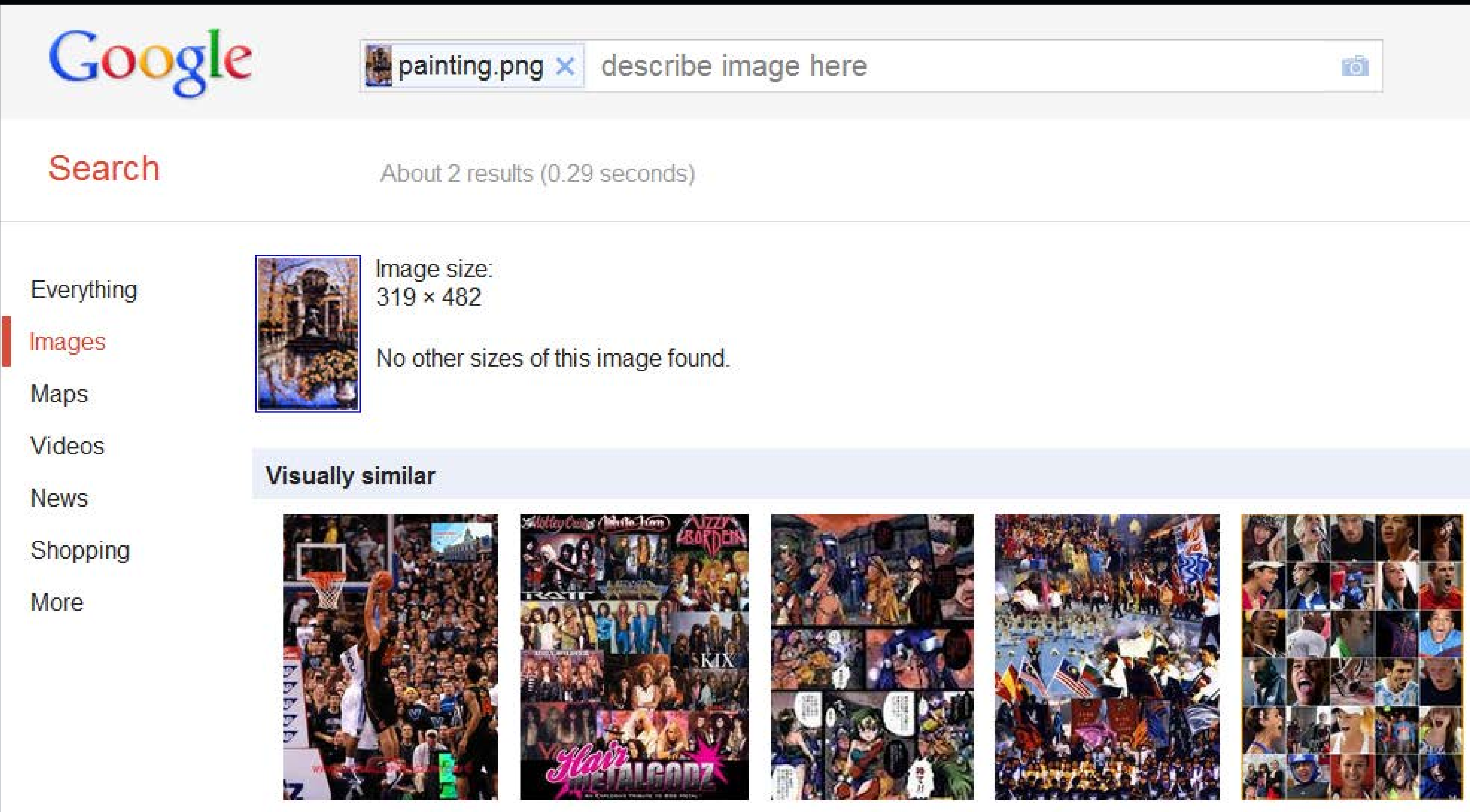
Move

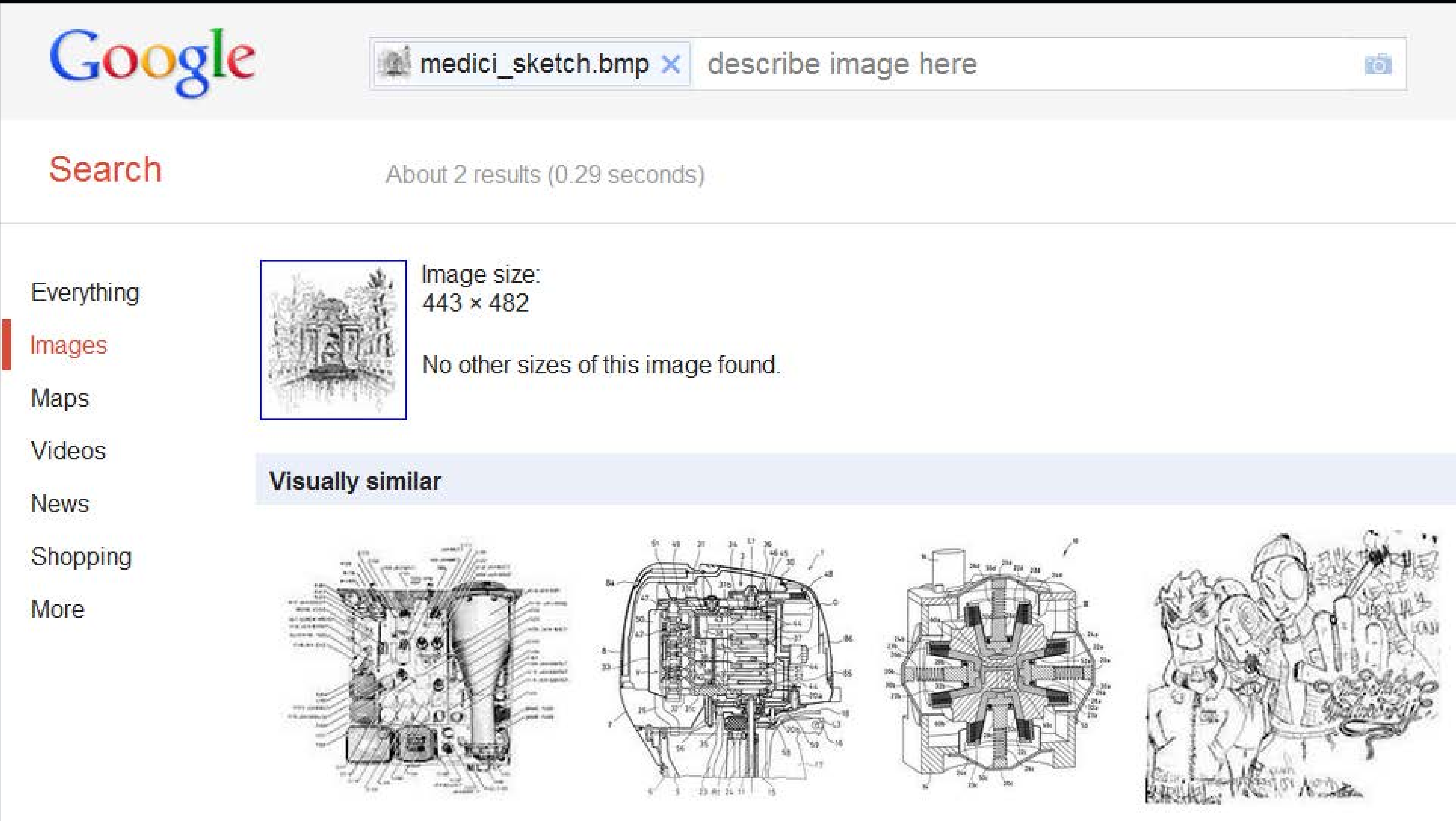
Watch a short video to learn more.


Google
medici_summer.jpg
luxembourg gardens
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
1024 × 829
No other sizes of this image found.
Visually similar

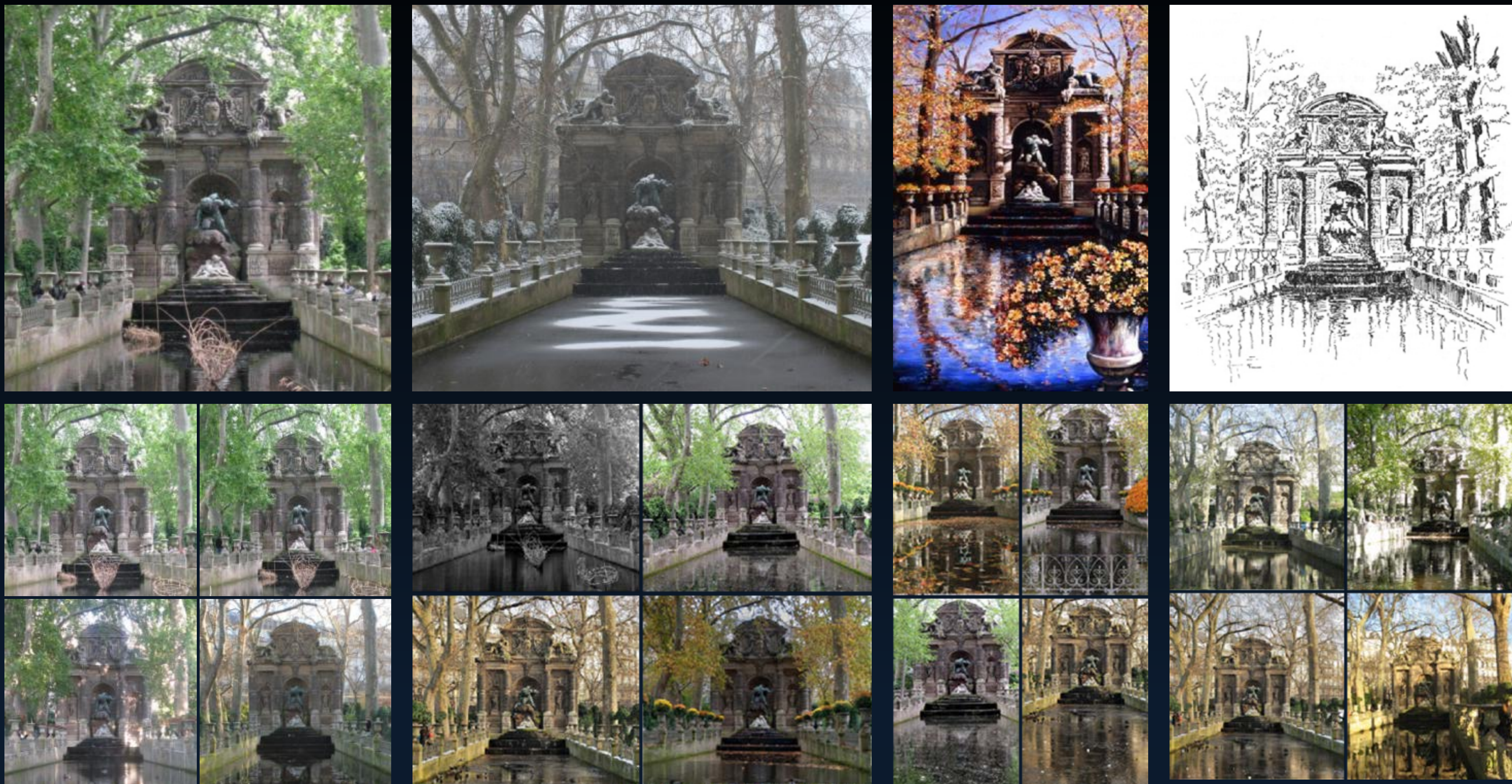
Medici Fountain, Paris (winter)


Google
medici_winter.png
luxembourg gardens
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
713 × 600
No other sizes of this image found.
Visually similar


Google
painting.png
describe image here
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
319 × 482
No other sizes of this image found.
Visually similar
KIX


Google
medici_sketch.bmp
describe image here
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
443 × 482
No other sizes of this image found.
Visually similar

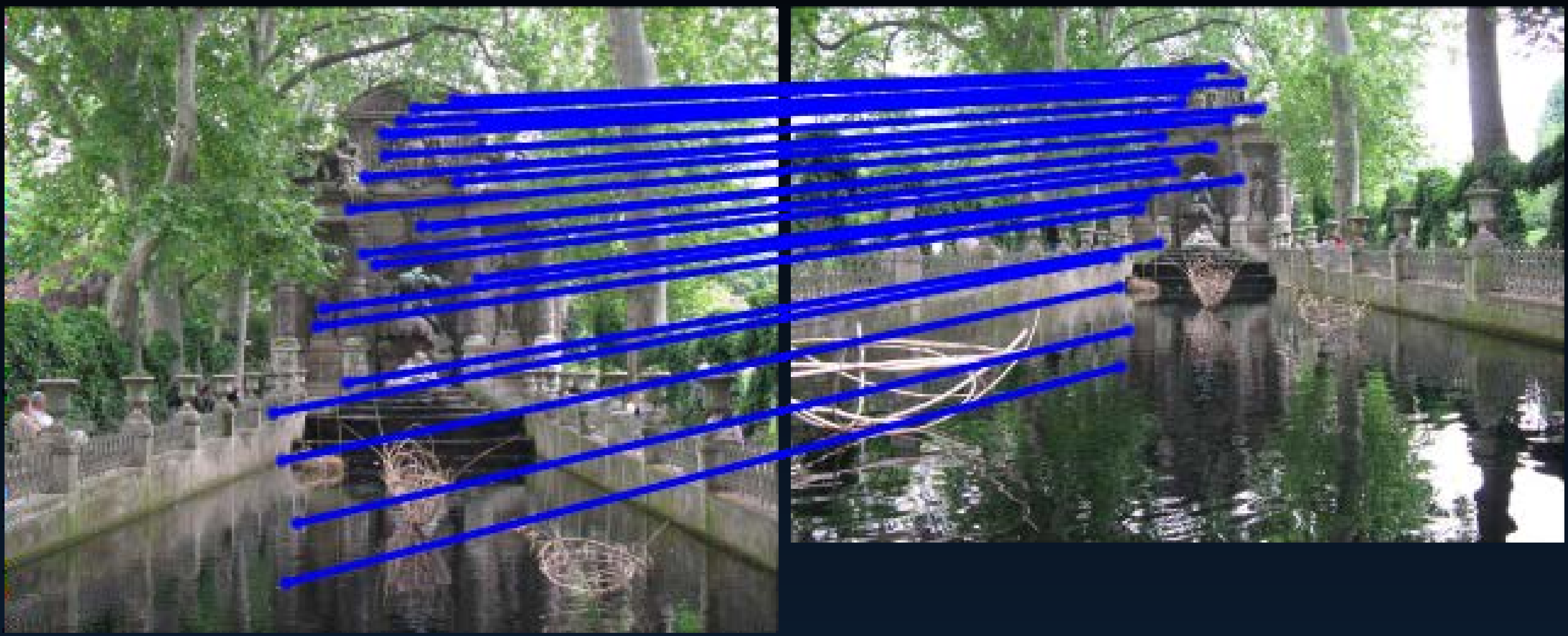
OUR GOAL [SIGGRAPH ASIA'11]



Why is this so Hard?




EXAMPLE: SIFT MATCHING



[SIFT: Lowe, 2004]

Example: SIFT Matching

[SIFT: Lowe, 2004]


STOCKMORE STREET

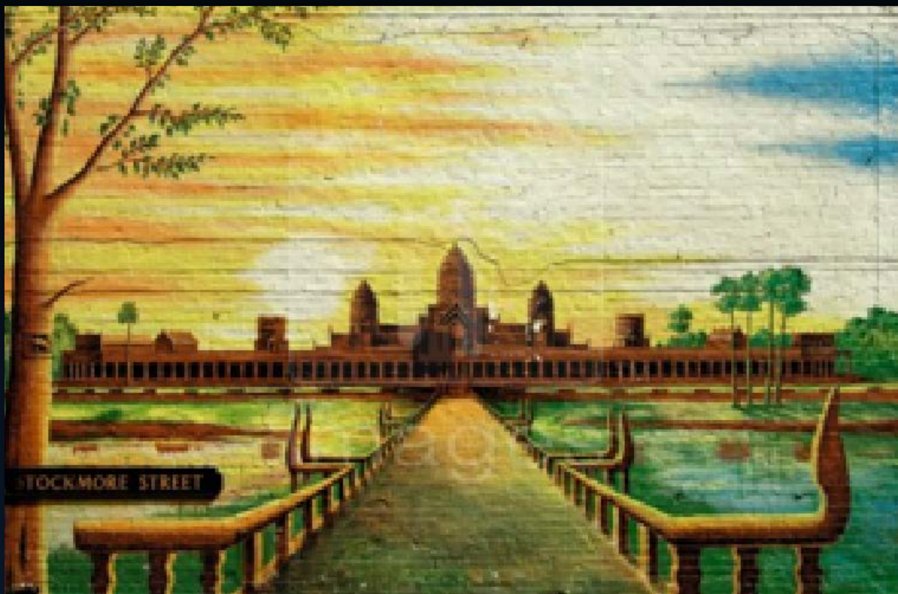
Input Query





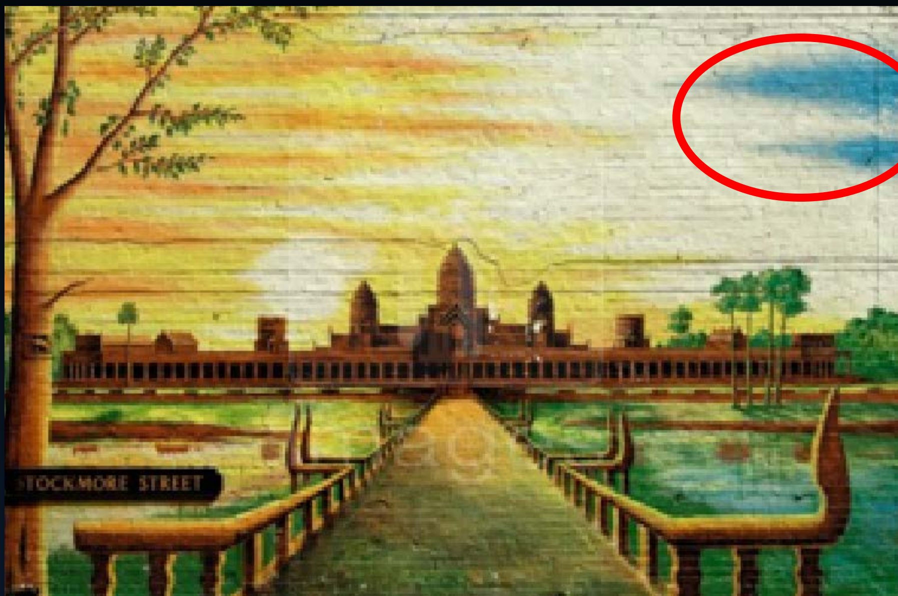
Top Matches

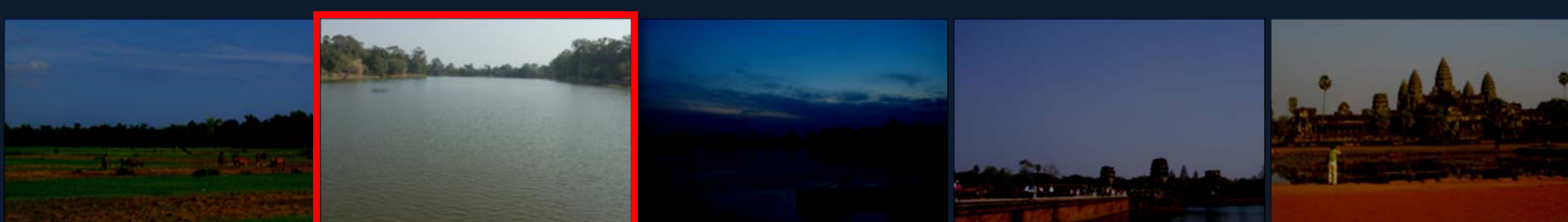
Input Query



Top Matches

Input Query

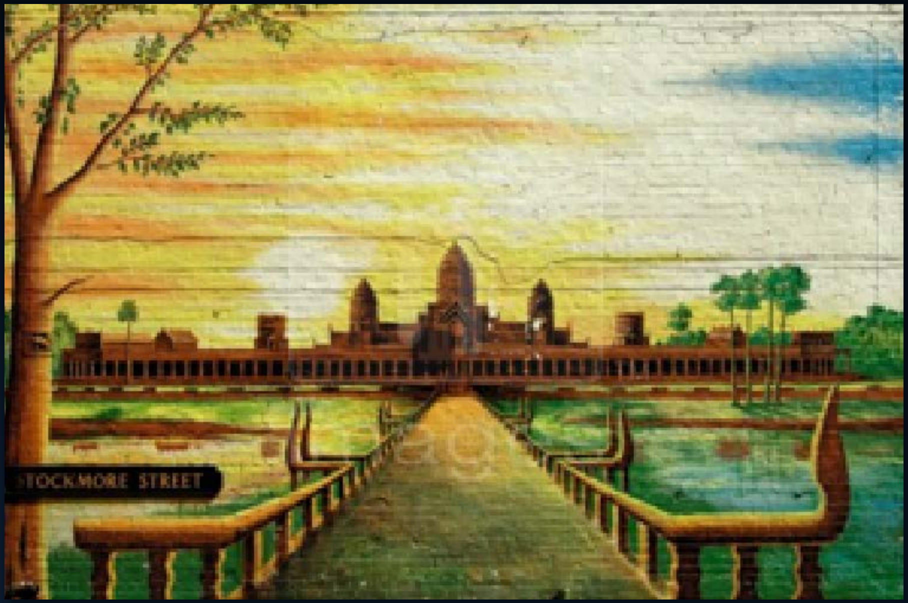




Top Matches

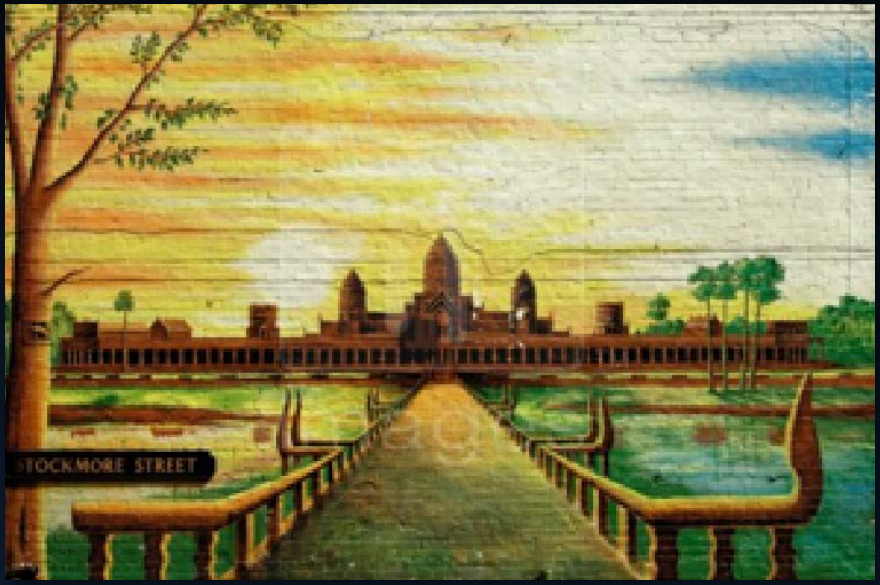
IMPORTANT PARTS?

Input Query

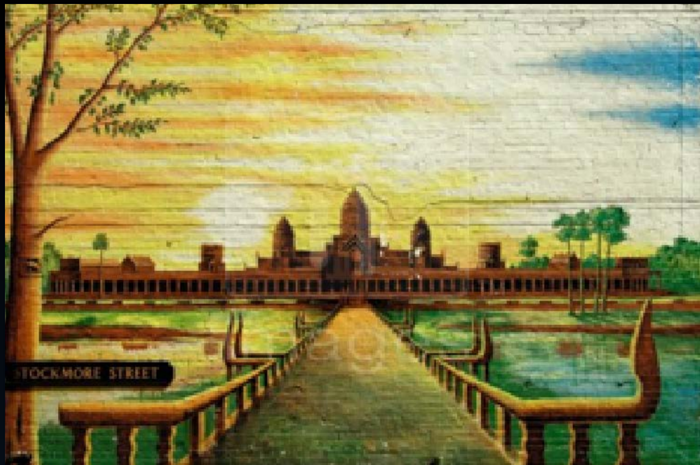


Important Parts

Input Query



Top Matches



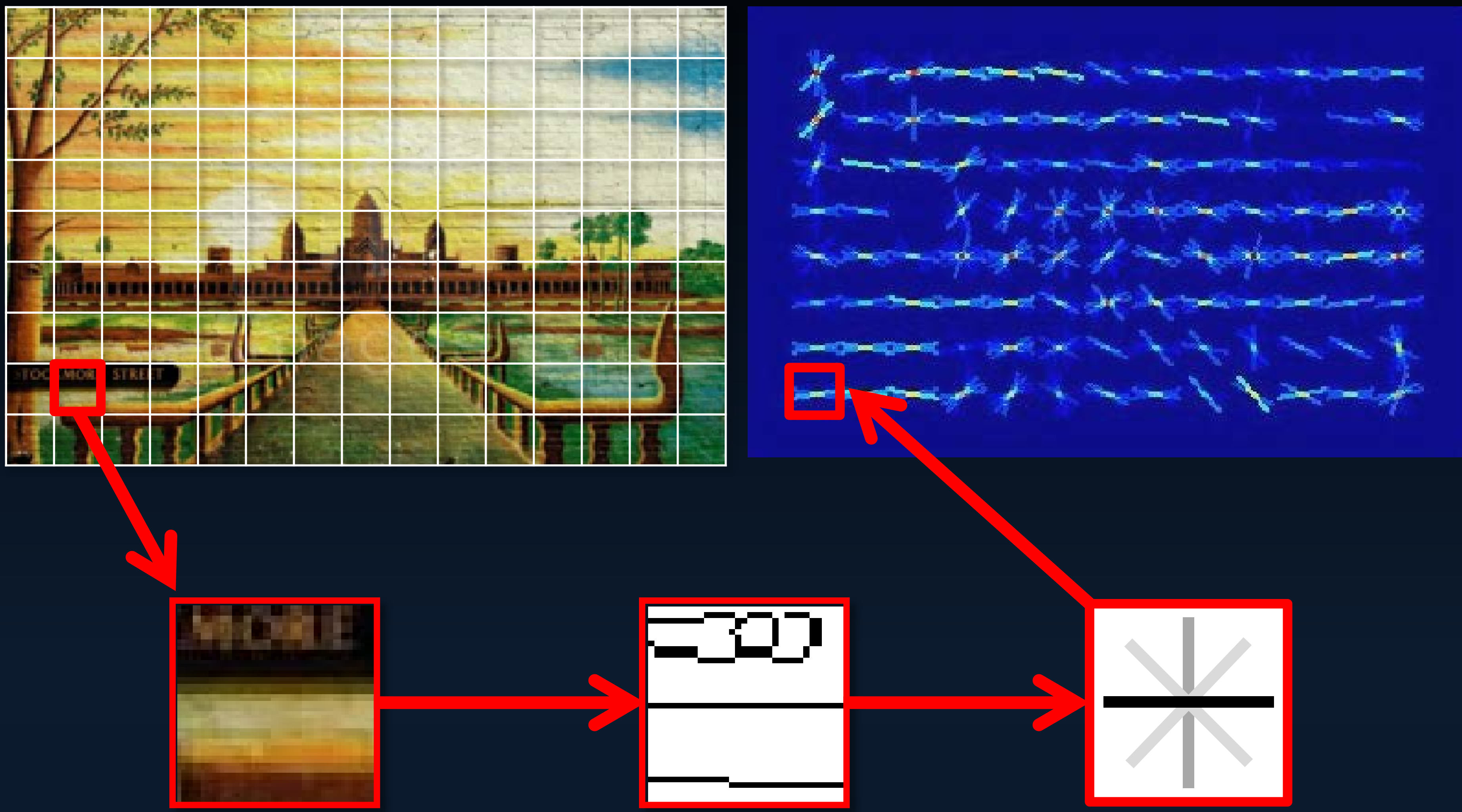
“Data-driven Uniqueness”


STOCKMORE STREET

Possible Explanation

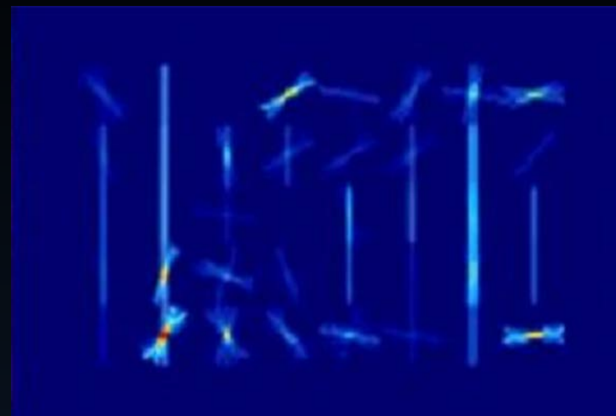
- If the space of images was uniform, nearest neighbor would work perfectly well
- But the space is very non-linear, non-Eucledian
- The Exemplar-SVM is trying to make a small, linear rescaling of visual space, near the query point
  - i.e. capturing the natural image statistics near the exemplar
  - Or maybe global is good enough (see Deva)

Feature Representation
Histogram of Oriented Gradients (HOG)



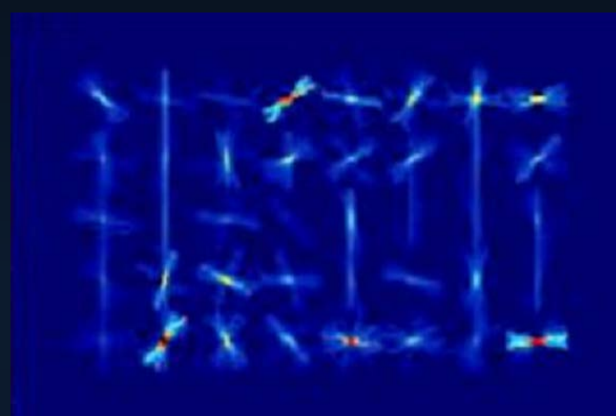
[Dalal and Triggs, CVPR, 2005]

Input Query



HOG

Top Match

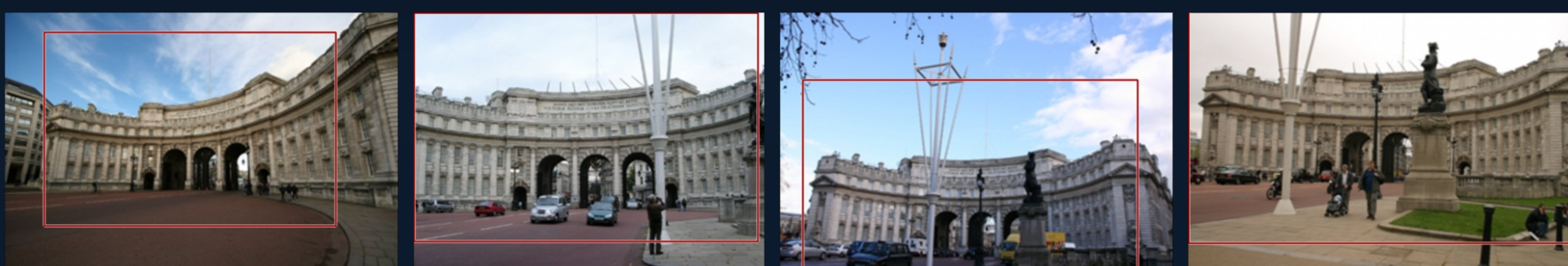


Learnt Weights

Top Match

SEARCH USING IMAGES

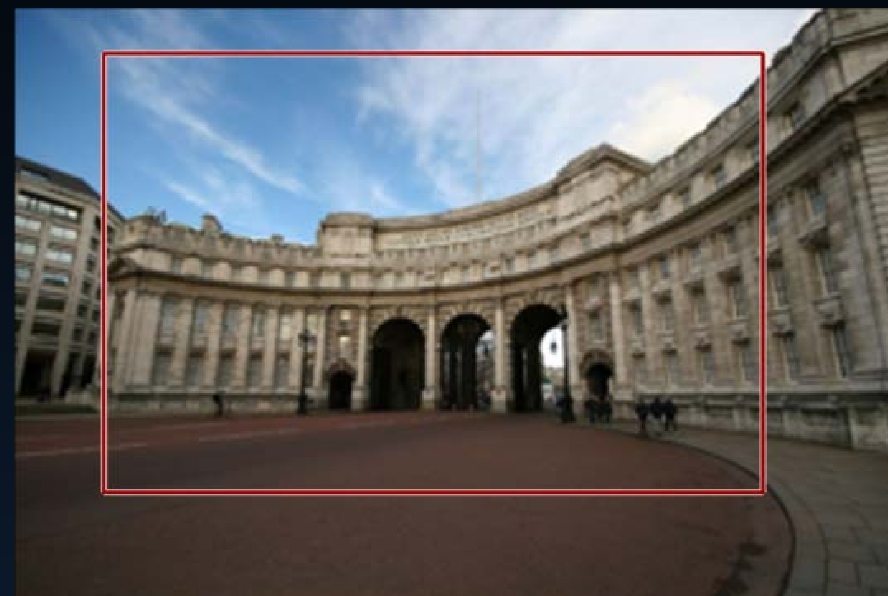
Input Query

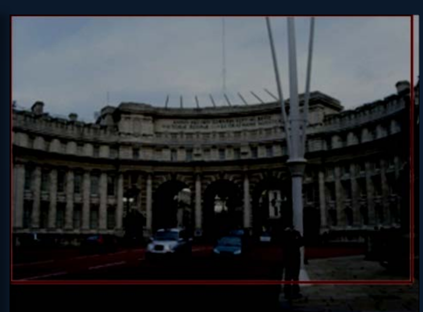


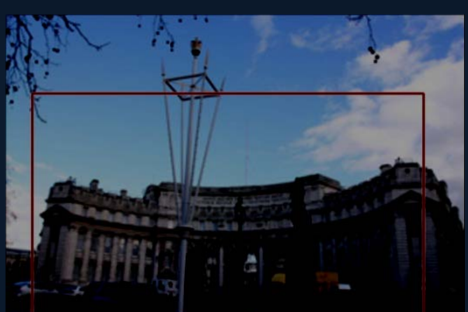
Top Matches

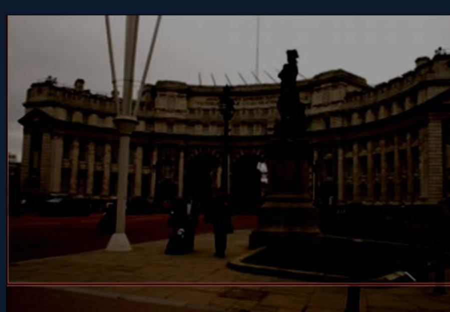
SEARCH USING IMAGES

Input Query





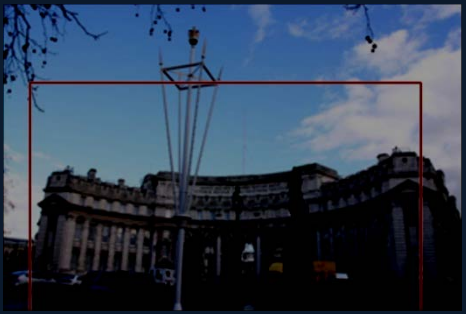


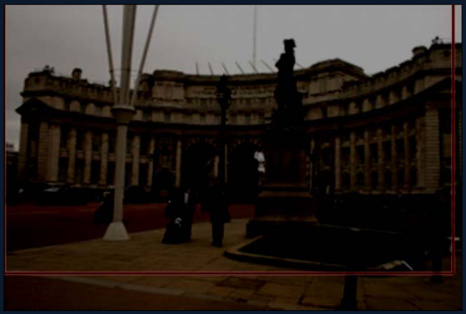


Top Matches

SEARCH USING IMAGES

Input Query





Top Matches

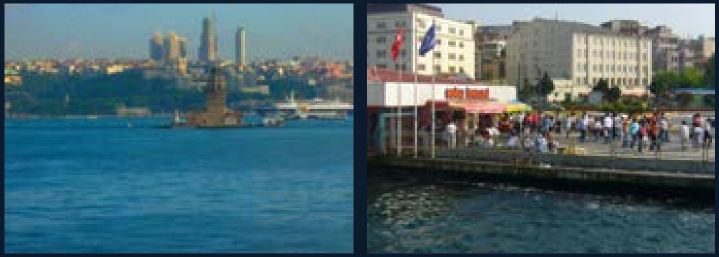
SEARCH USING PAINTINGS

Input Painting

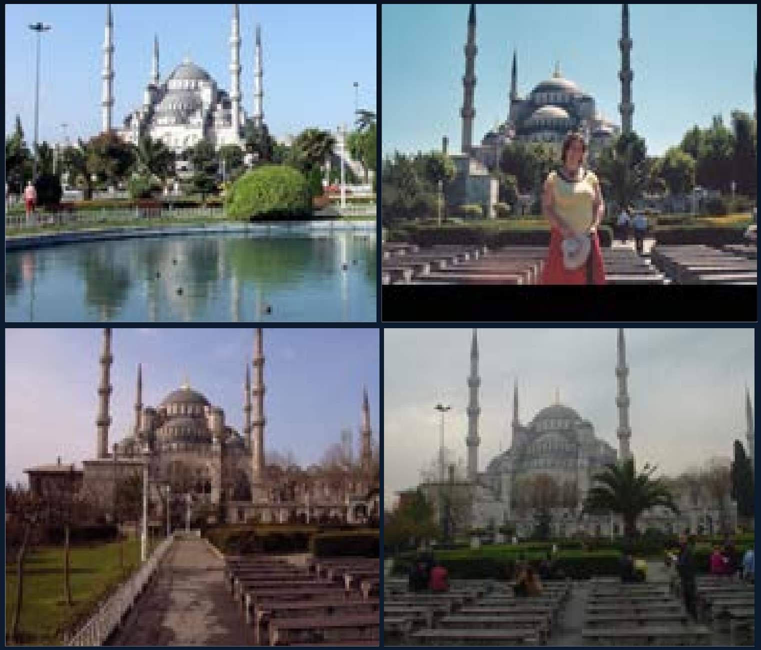
GIST

Bag-of-Words

Tiny Images



HOG



Our
Approach

SEARCH USING PAINTINGS

Input Painting

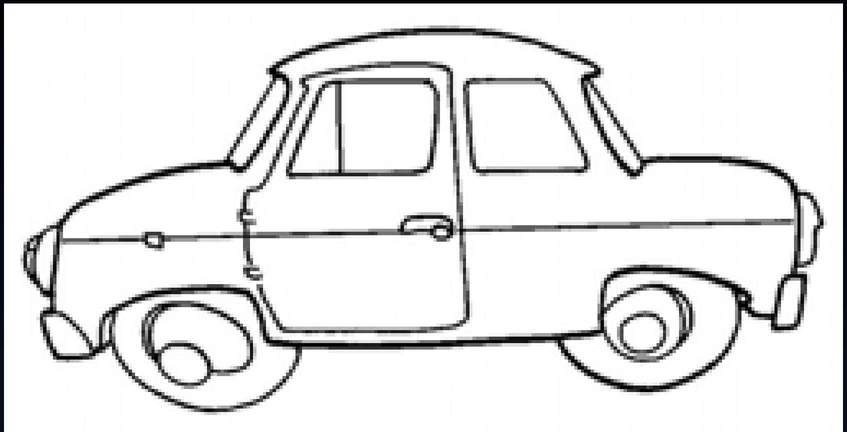
Top Matches

SEARCH USING PAINTINGS

Input Painting

Top Matches

SEARCH USING SKETCHES



Input Sketch

Our Approach

Tiny Images

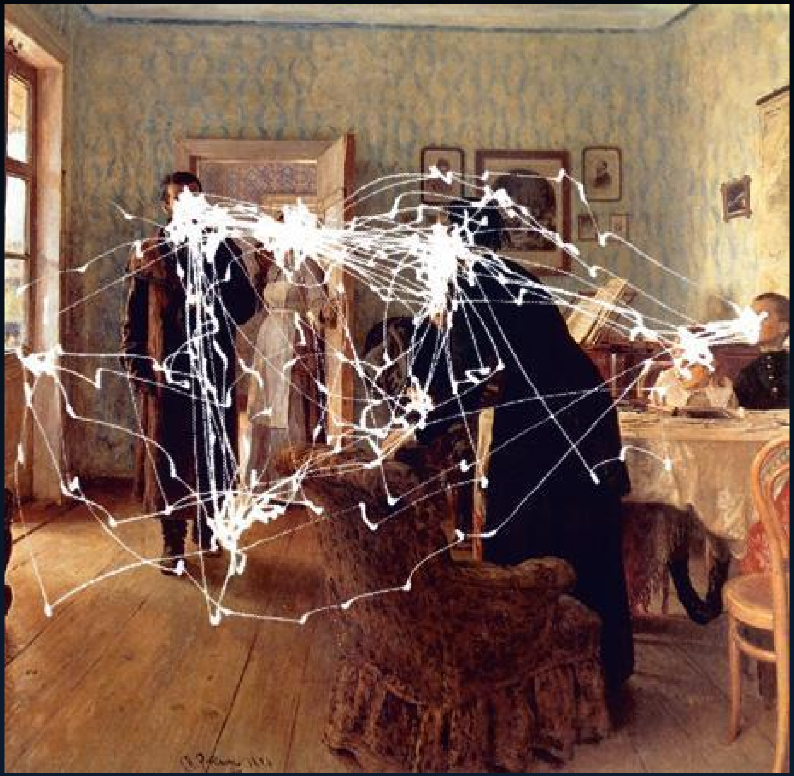
GIST

Bag-of-Words

HOG

SEARCH USING SKETCHES

SALIENCY

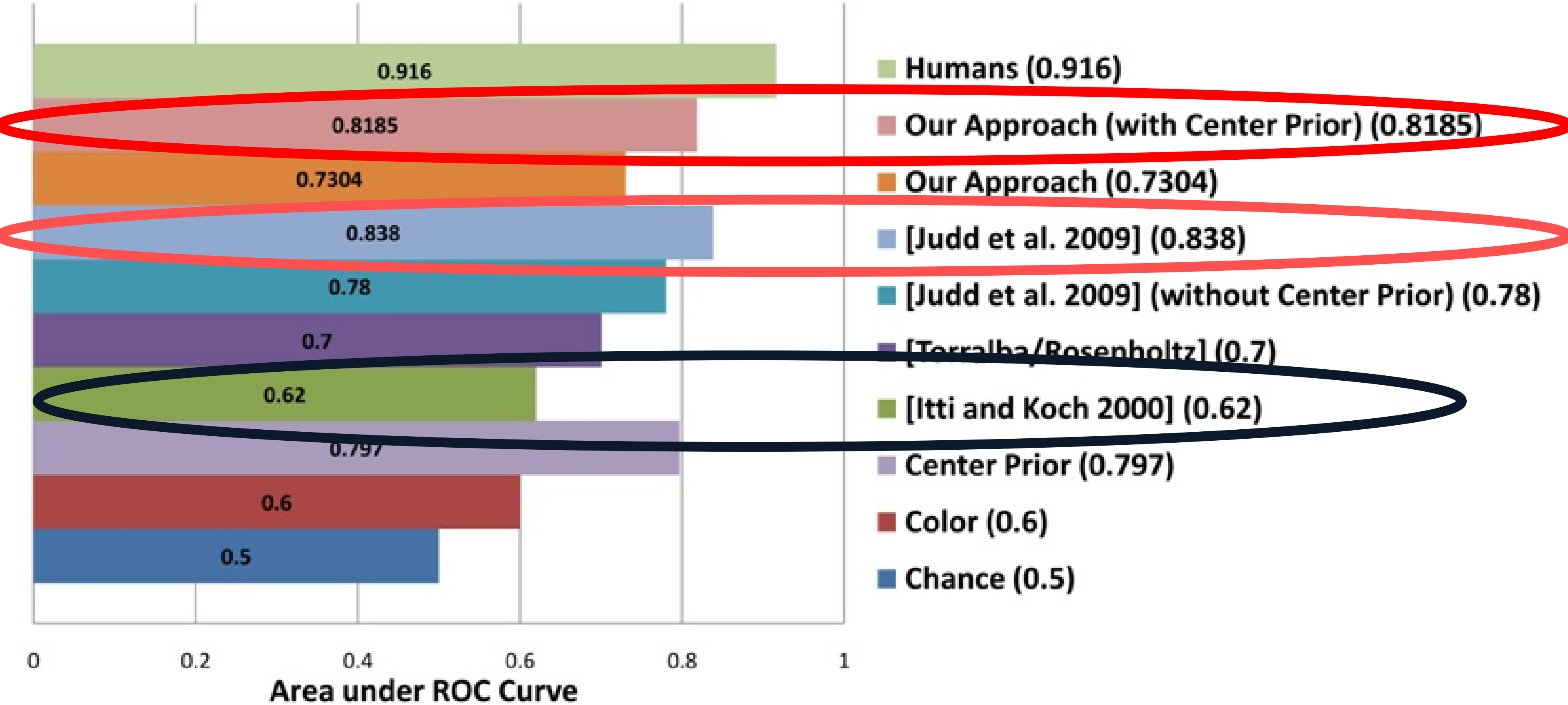


[Yarbus, 1962]

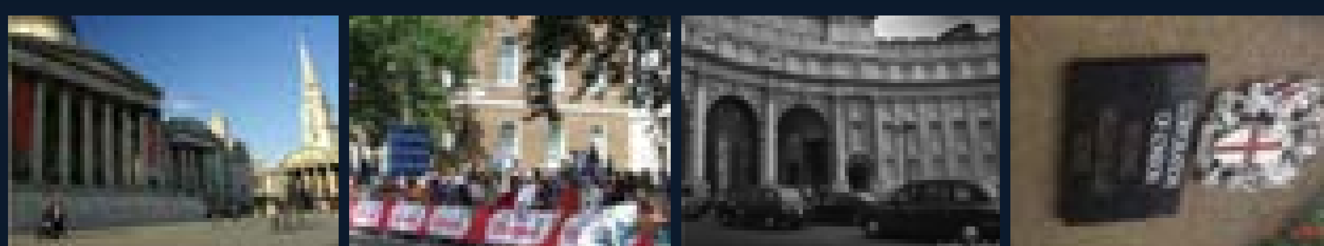
PROXY FOR SALIENCY?

PREDICTING SALIENCY

SALIENCY DATASET [Judd et al., 2009]



WHERE DOES IT FAIL?



Top Matches

APPLICATIONS

WHERE WAS THE PAINTER STANDING?

Input Painting

PAINTING2GPS

Input Painting

Retrieval set
10,000 Geo-tagged Flickr Images

100 top matches used to estimation

PAINTING2GPS

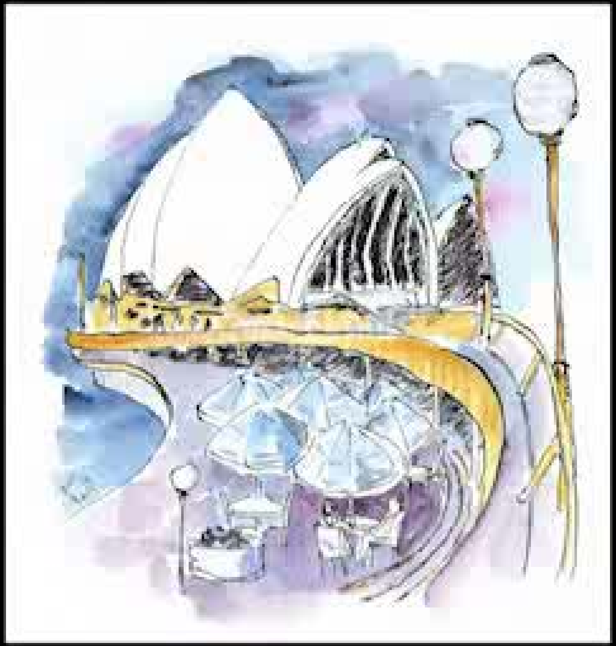
Input Painting

Estimated Geo-location

Estimated using 100 top matches

PAINTING2GPS

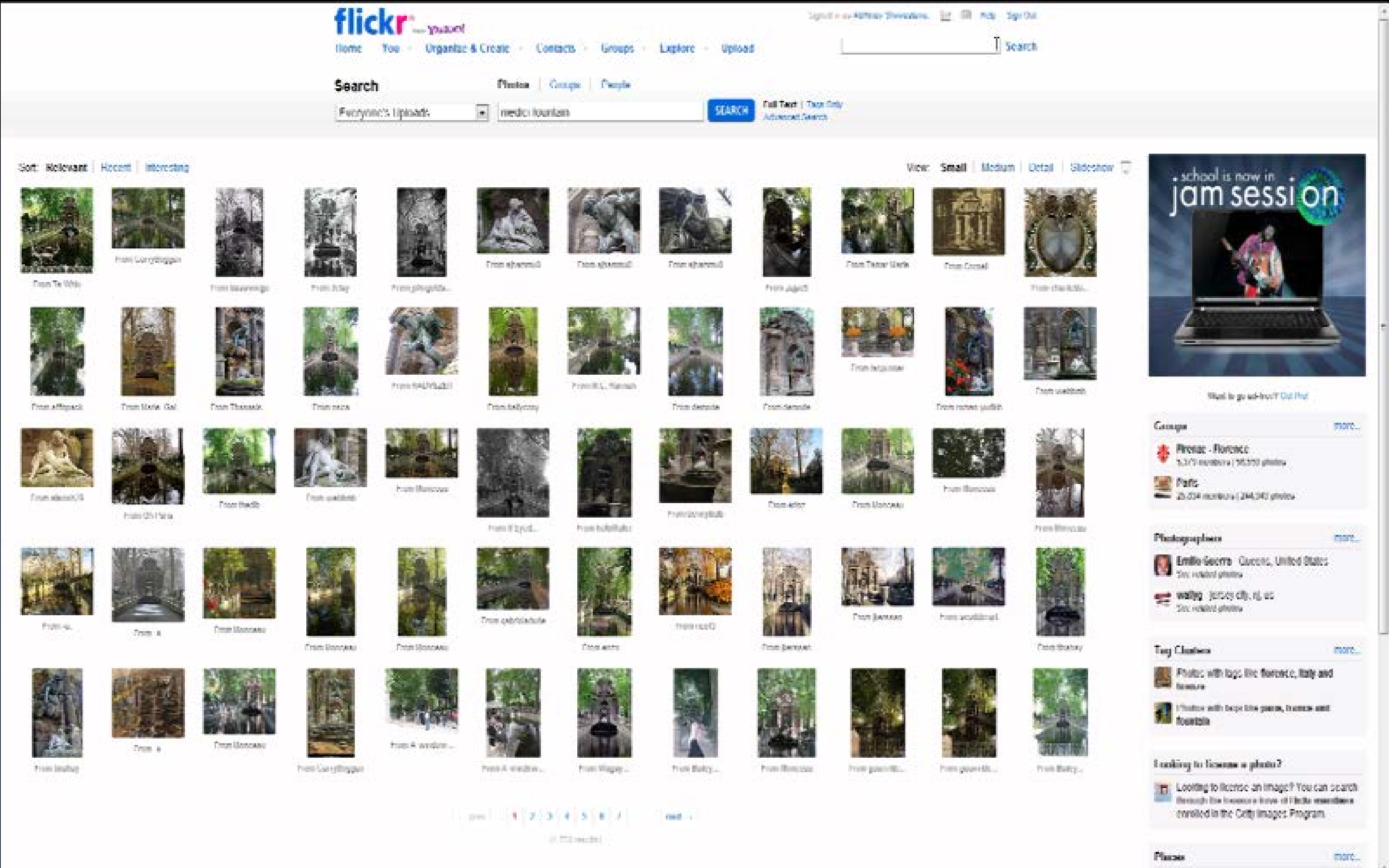
Input Painting



Sydney Opera House

VISUAL SCENE EXPLORATION

VISUAL SCENE EXPLORATION



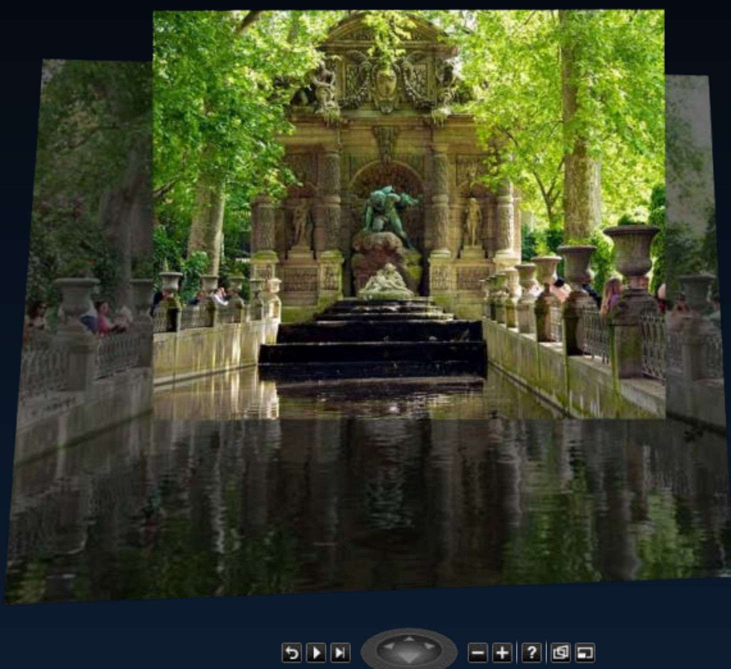
Organizing Visual "Garbage Heap"

"It irritated him that the "dog" of 3:14 in the afternoon, seen in profile, should be indicated by the same noun as the dog of 3:15, seen frontally…"

"My memory, sir, is like a garbage heap."

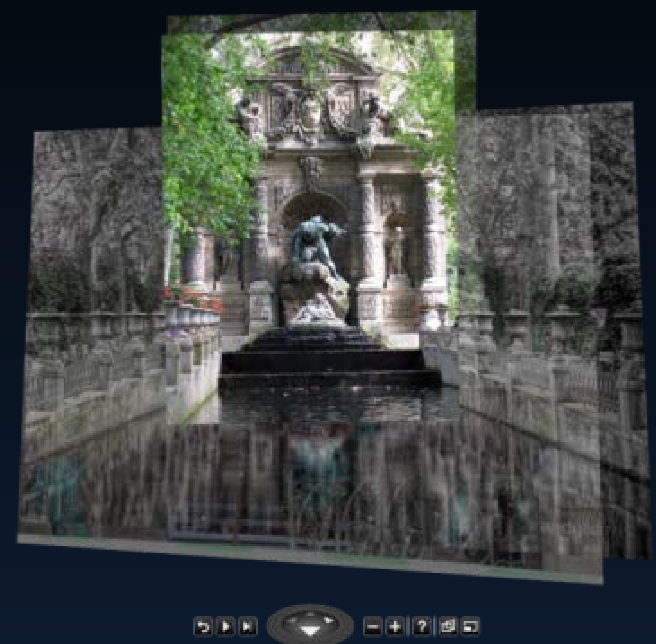
-- Jorge Luis Borges, *Funes the Memorious*

PHOTOSYNTH
[Snavely et al., 2006]



Synth 1

Synth 2

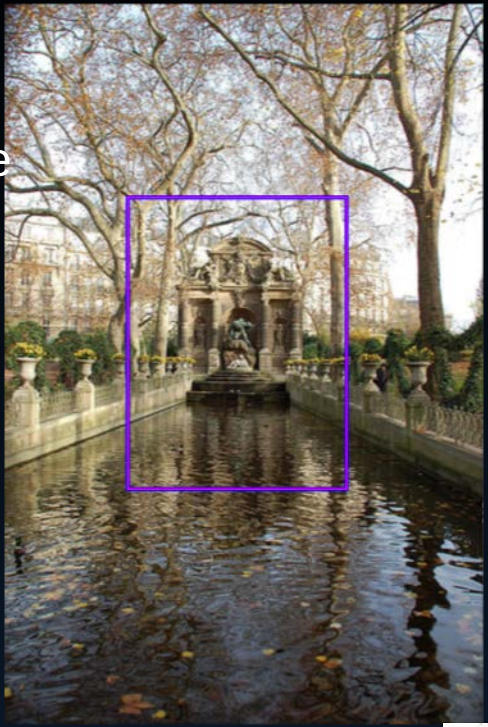


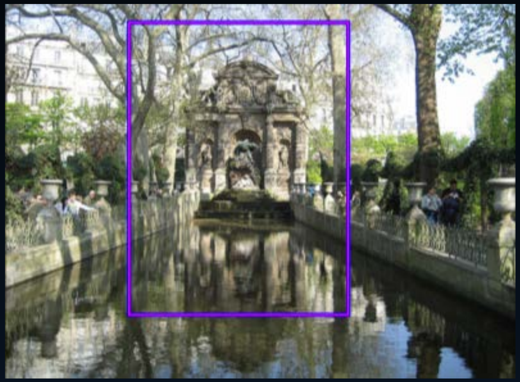
Synth 3 **…**

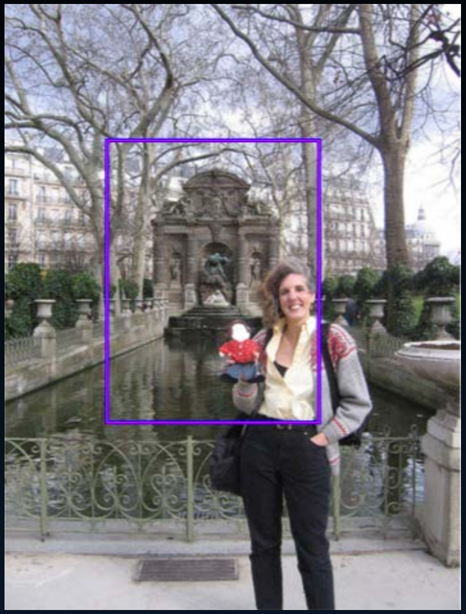
Dataset size: 136 photos (from flickr)
# of discovered synths: 14
82 photos not part of any synth

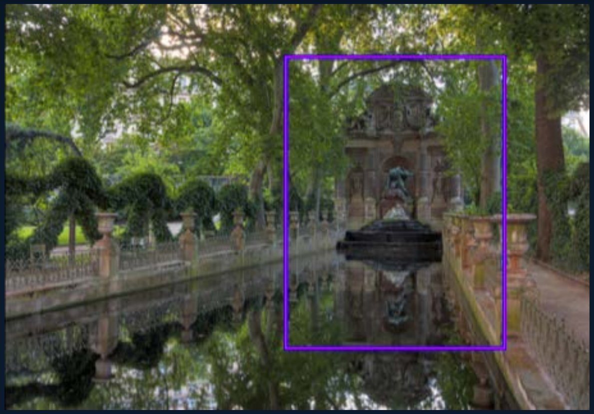
Finding Similar Images

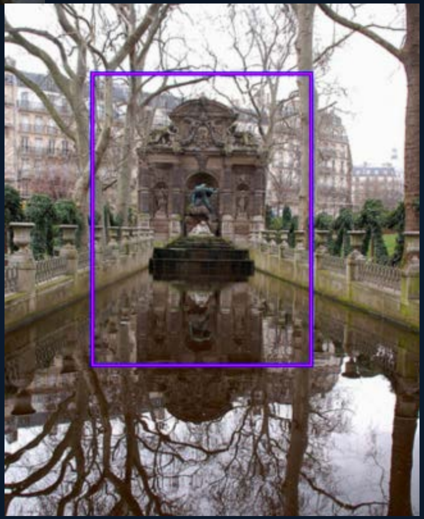
Query image



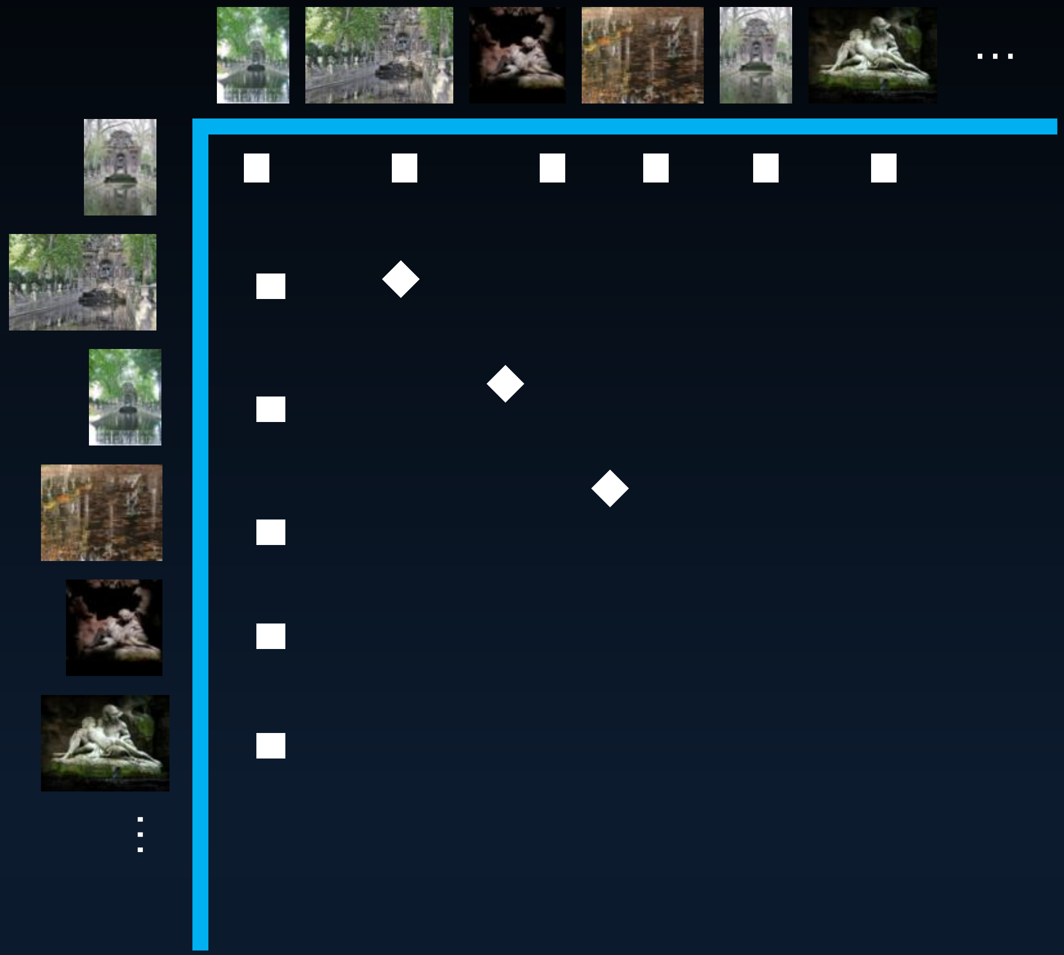




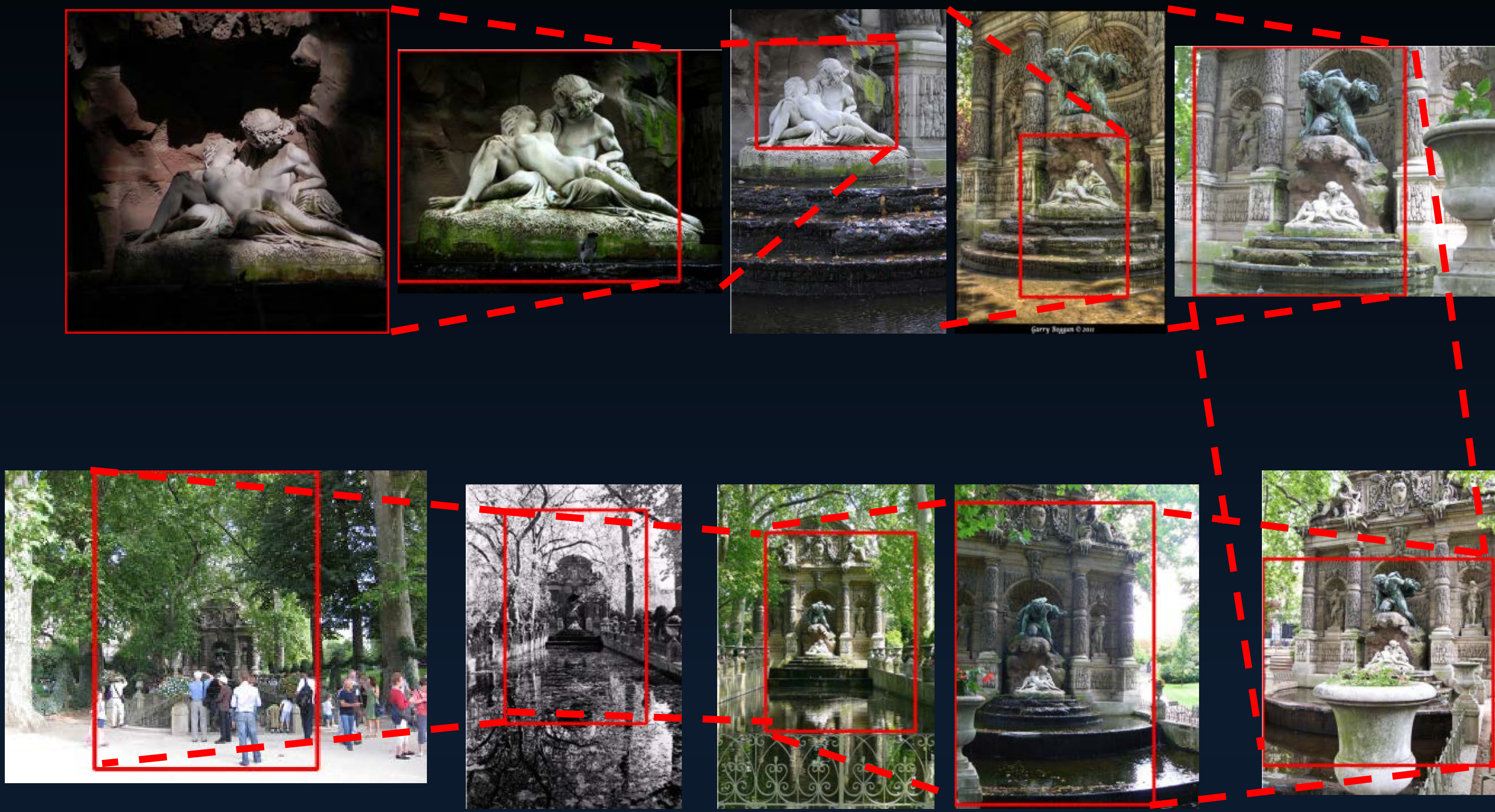




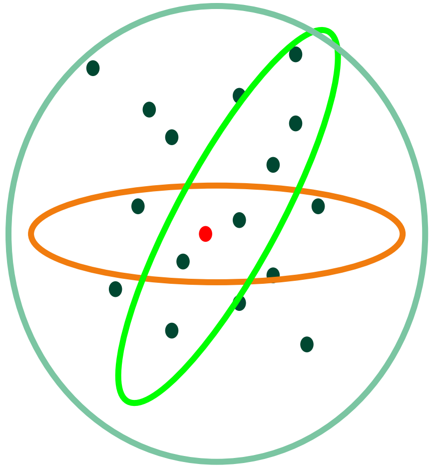
Pairwise Similarity Matrix



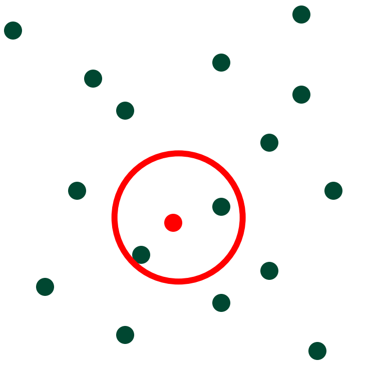
TRAVERSING THE GRAPH



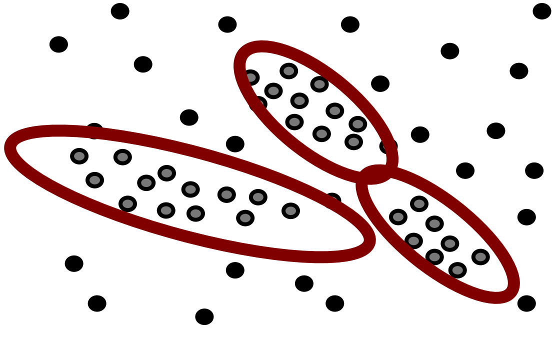
Ways to use Big Data



1. See what different subsets of data think of you



2. Use kNN to make a smaller, cozier sub-problem



3. Visual Data Mining: find needles in a haystack

4. Ditch Categories – keep all instances and connect them

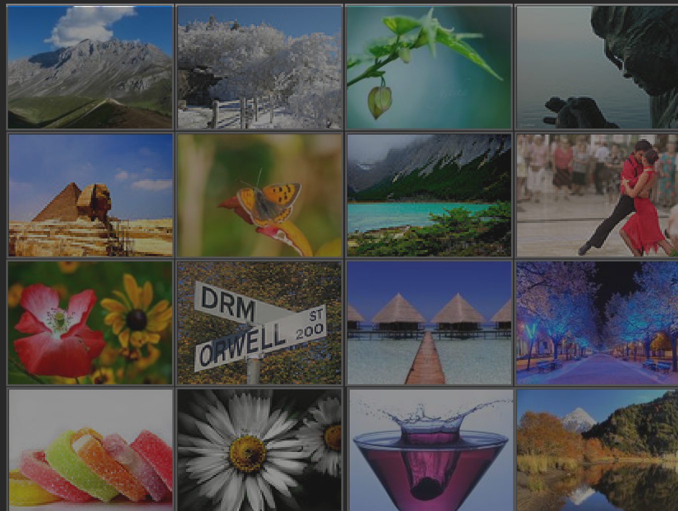
Priors for Large Photo Collections
&
What they Reveal about Cameras

Sujit Kuthirummal — Columbia University

Aseem Agarwala — Adobe Systems, Inc.

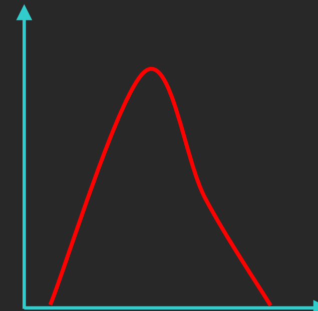
Dan B Goldman — Adobe Systems, Inc.

Shree K. Nayar — Columbia University



Camera Distortion Free

Compute
Aggregate
Statistic

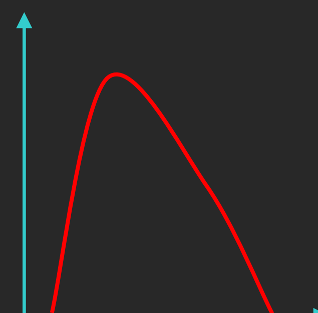


Independent of
Scenes, Photographers
& Cameras

Recover
Camera Properties

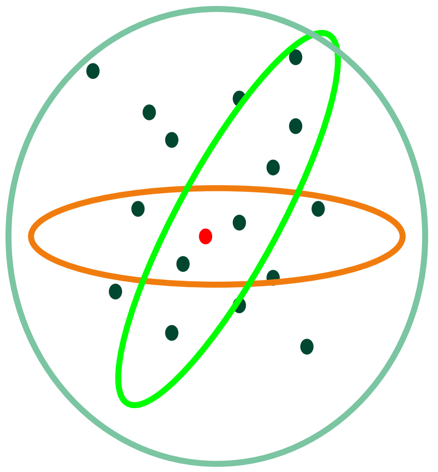
One Camera's Distortion

Compute
Aggregate
Statistic

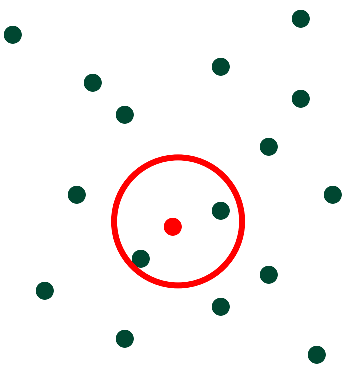


Independent of
Scenes & Photographers
Dependent on
Camera

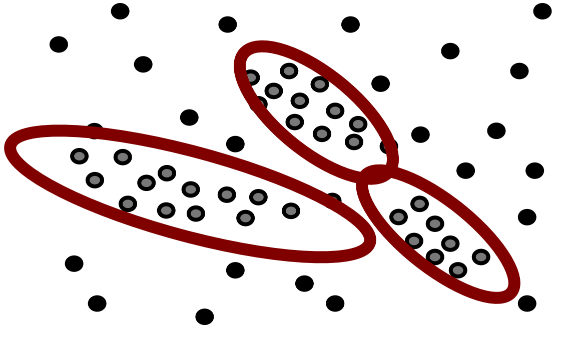
Ways to use Big Data



1. See what different subsets of data think of you



2. Use kNN to make a smaller, cozier sub-problem



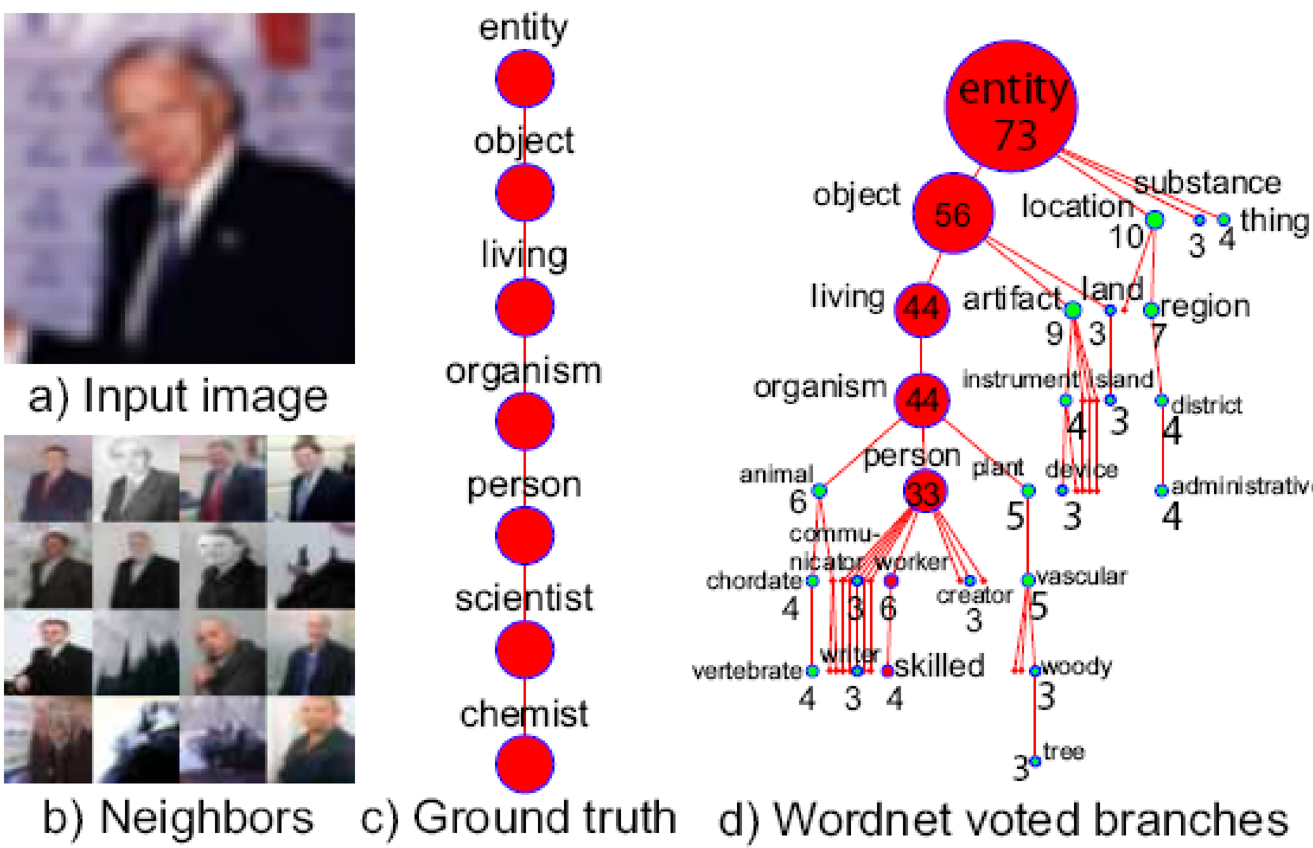
3. Visual Data Mining: find needles in a haystack

4. Ditch Categories – keep all instances and connect them

1. kNN + Label Transfer

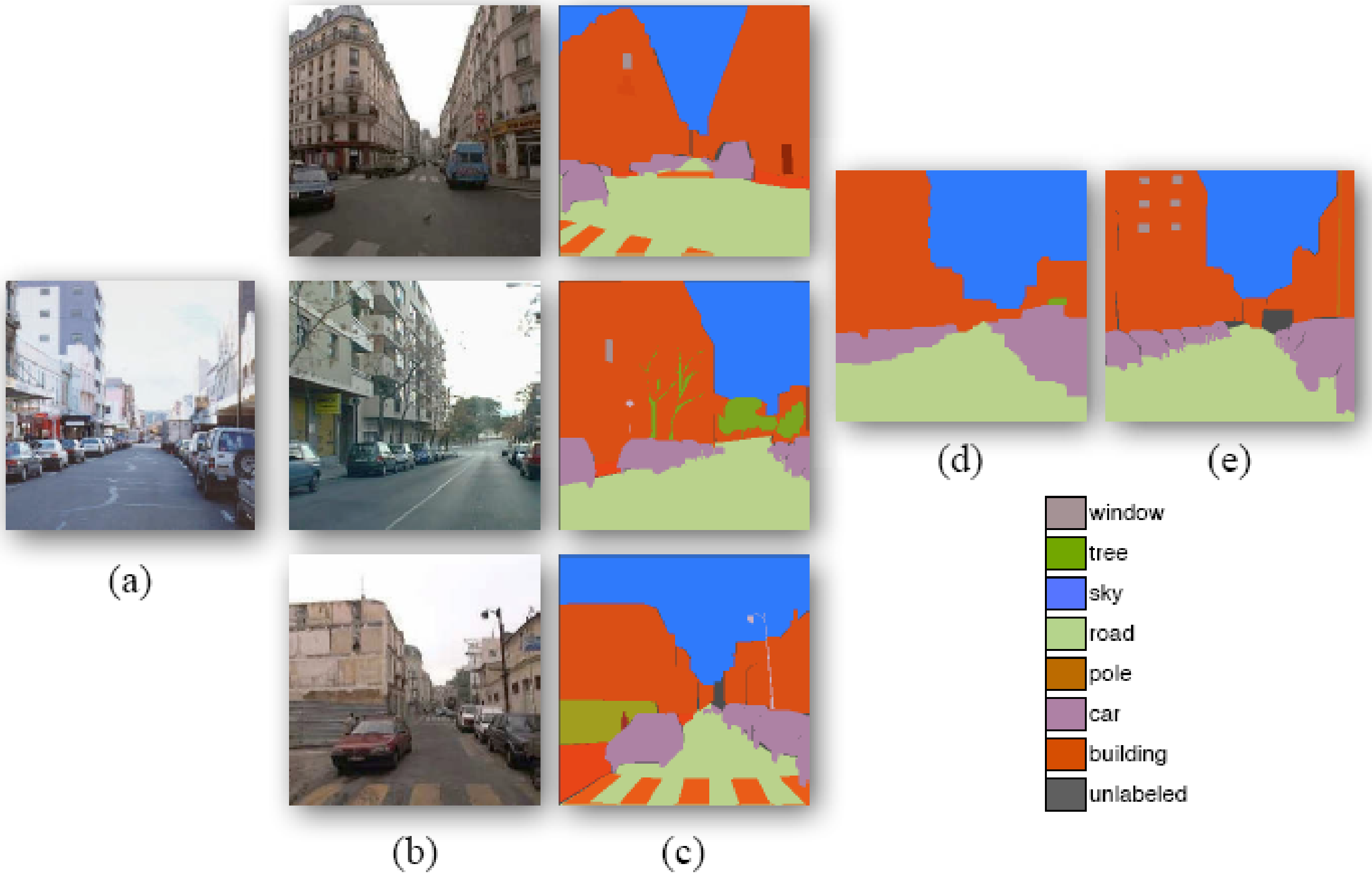
Sky, Water, Hills, Beach,
Sunny, mid-day

80 Million Tiny Images [PAMI'08]



Torralba, Fergus, Freeman, PAMI 2008

Non-parametric Scene Parsing [CVPR'09]



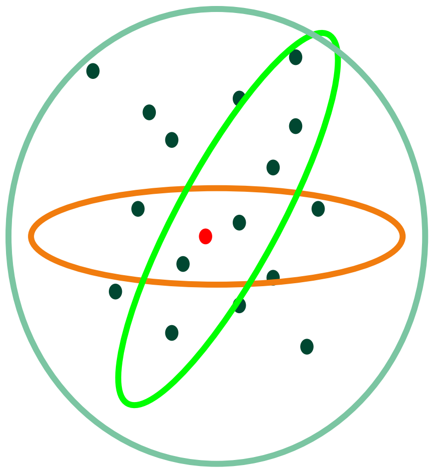
Liu, Yuen, Torralba, CVPR 2009

im2gps [CVPR'08]

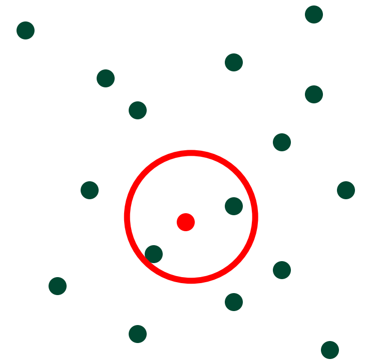
Query Photograph

Hays & Efros, CVPR 2008

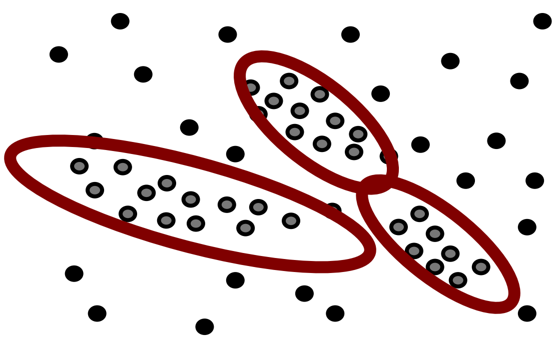
Ways to use Big Data



1. See what different subsets of data think of you



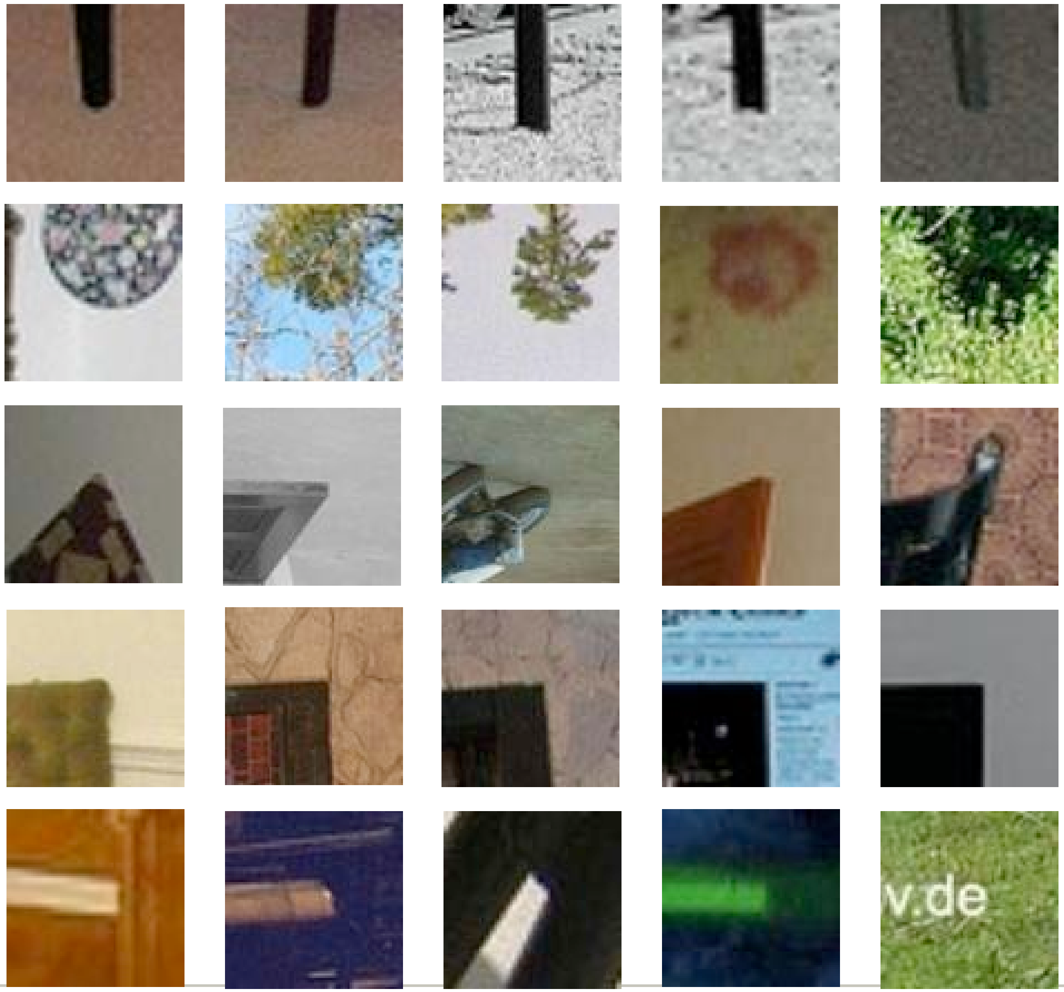
2. Use kNN to make a smaller, cozier sub-problem



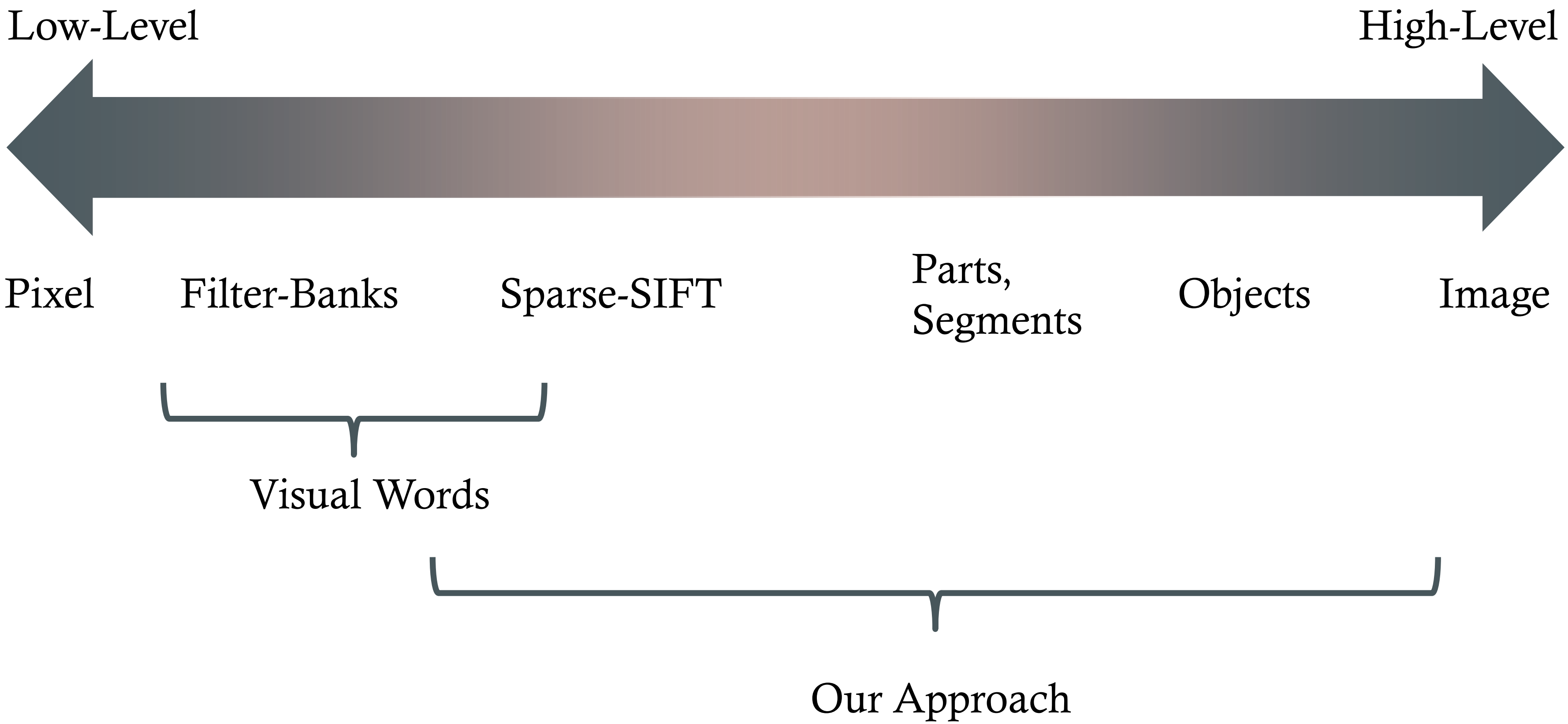
3. Visual Data Mining: find needles in a haystack

4. Ditch Categories – keep all instances and connect them

Visual Words or Letters?



Spectrum of Visual Features



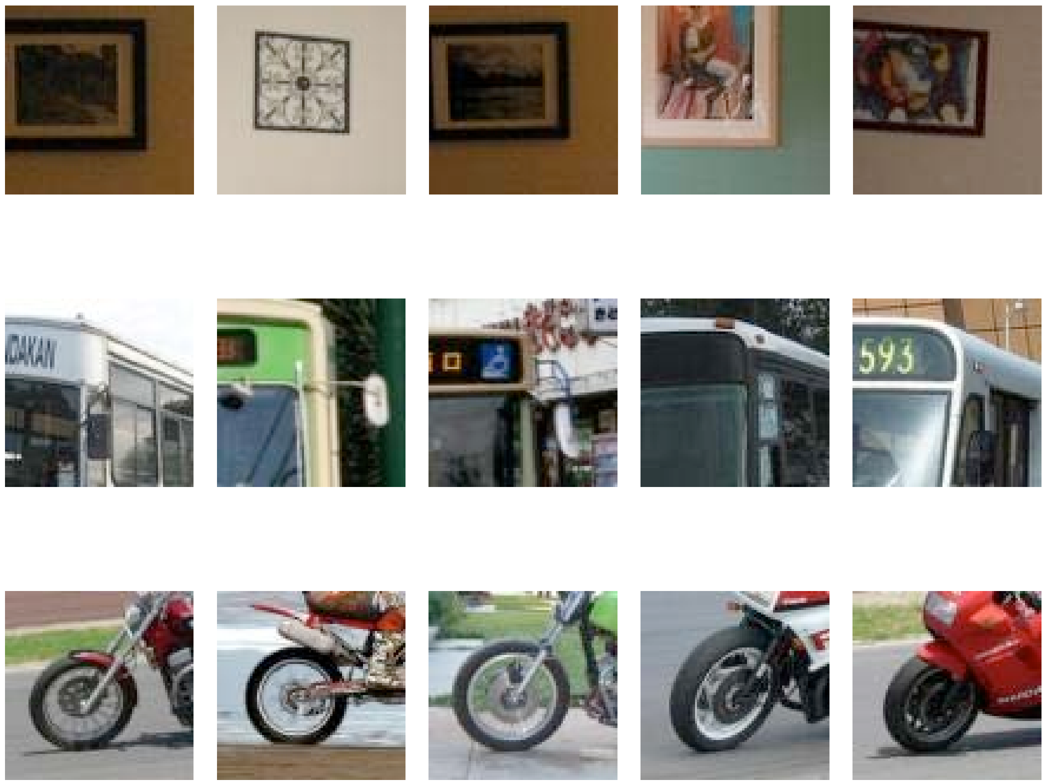
Discriminative Patches

Two key requirements

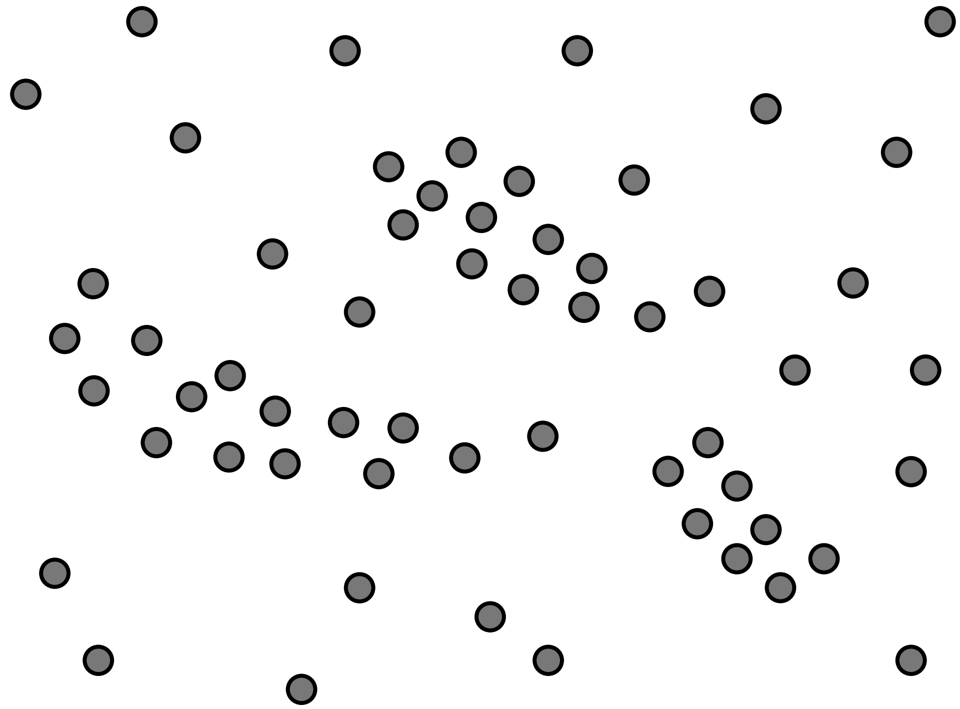
1. Need to occur frequently (representative)
2. …but not **too** frequently

   Discriminative: Need to be different enough from the rest of the visual world.

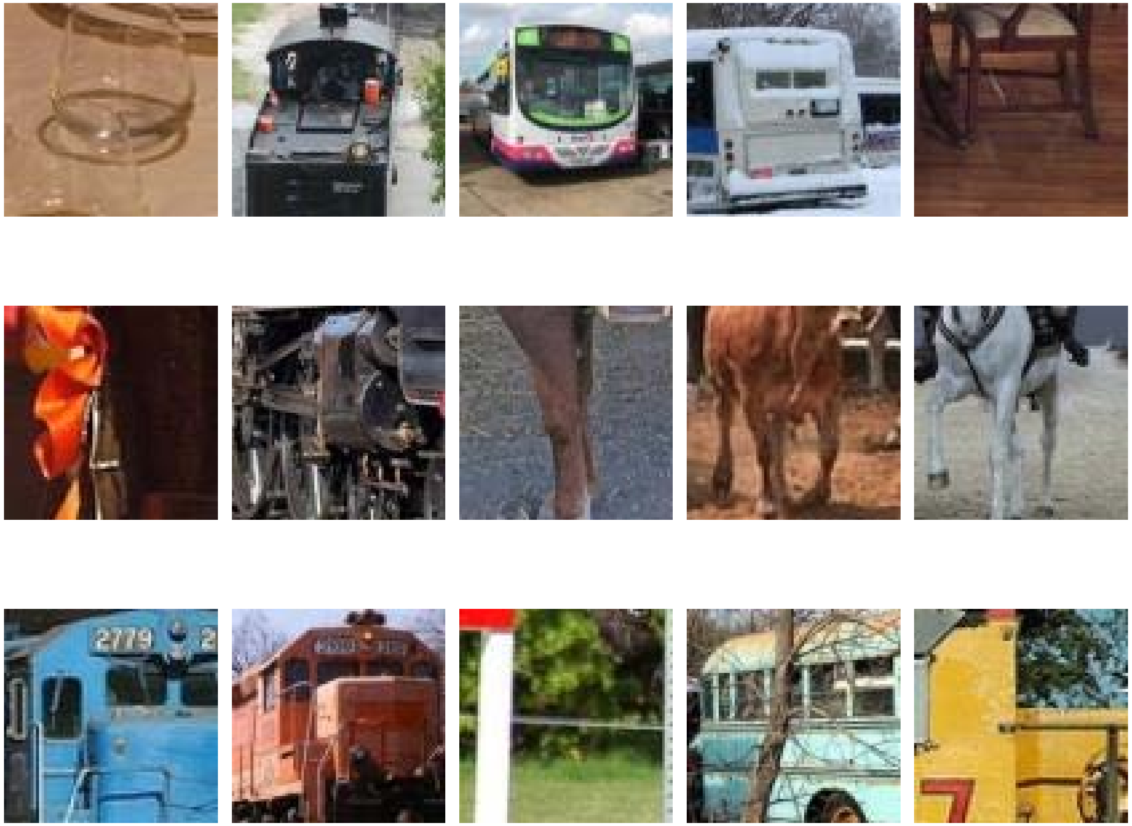
First some examples



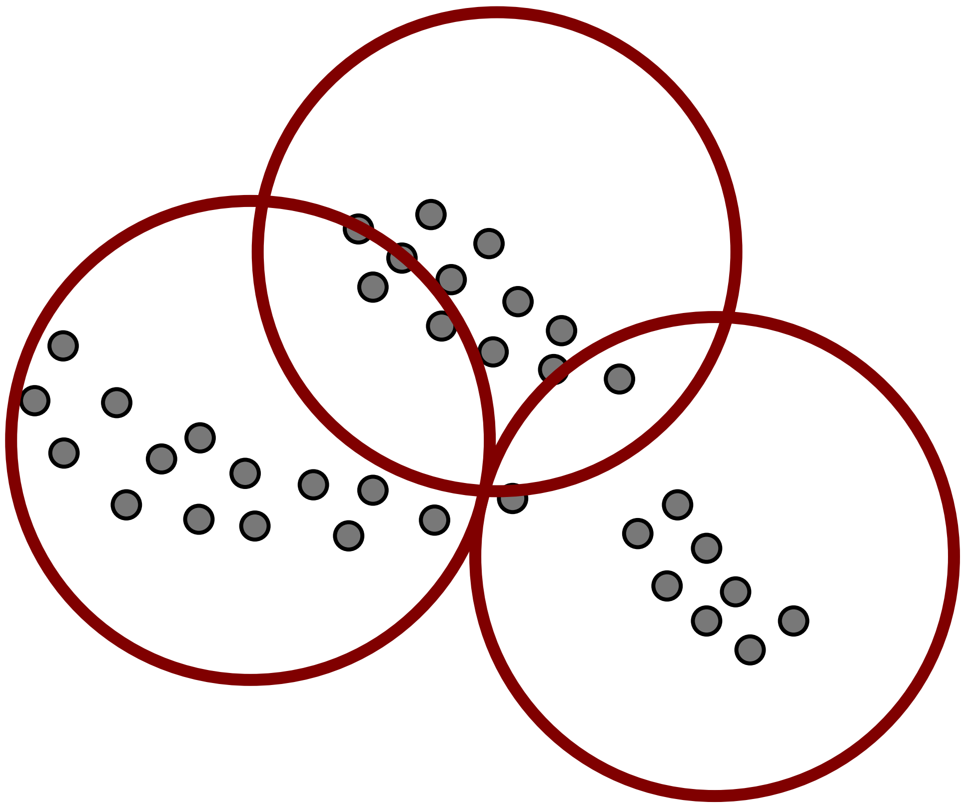
Finding needles in a haystack



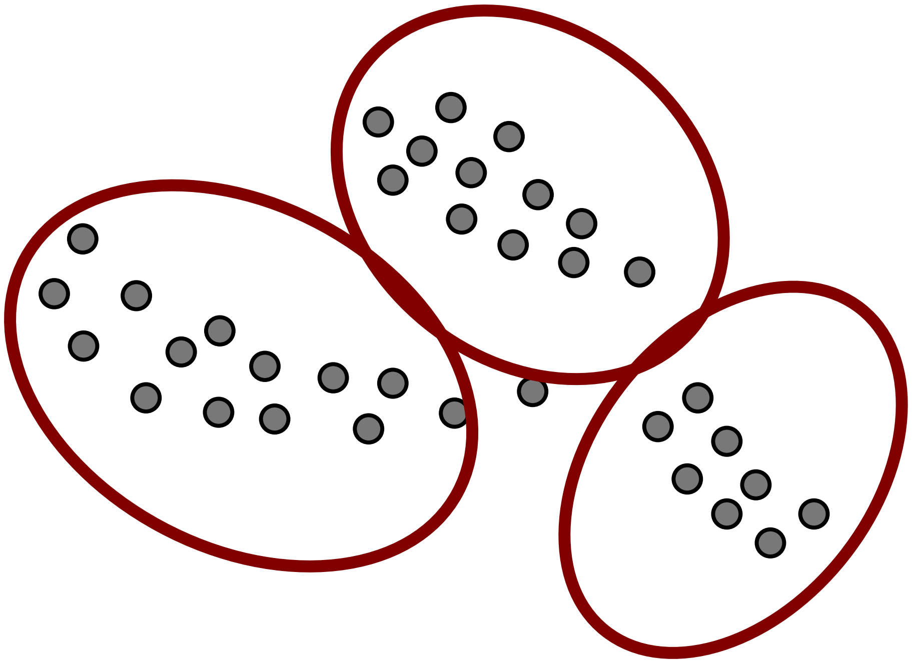
K-Means Clusters



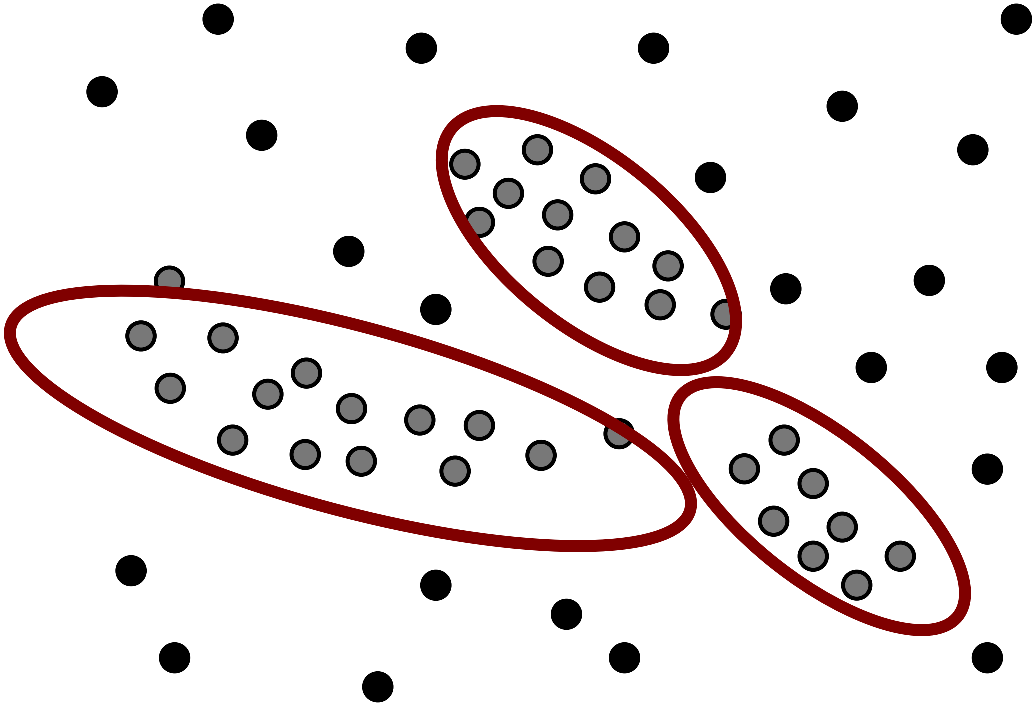
K-Means



Discriminative K-means



Discriminative Clustering+



Discriminative Clustering+

Initial

Final

Initial

Final

Discriminative Clustering++

KMeans

Iter 1

Iter 2

Iter 3

Iter 4

Discriminative Clustering++

KMeans

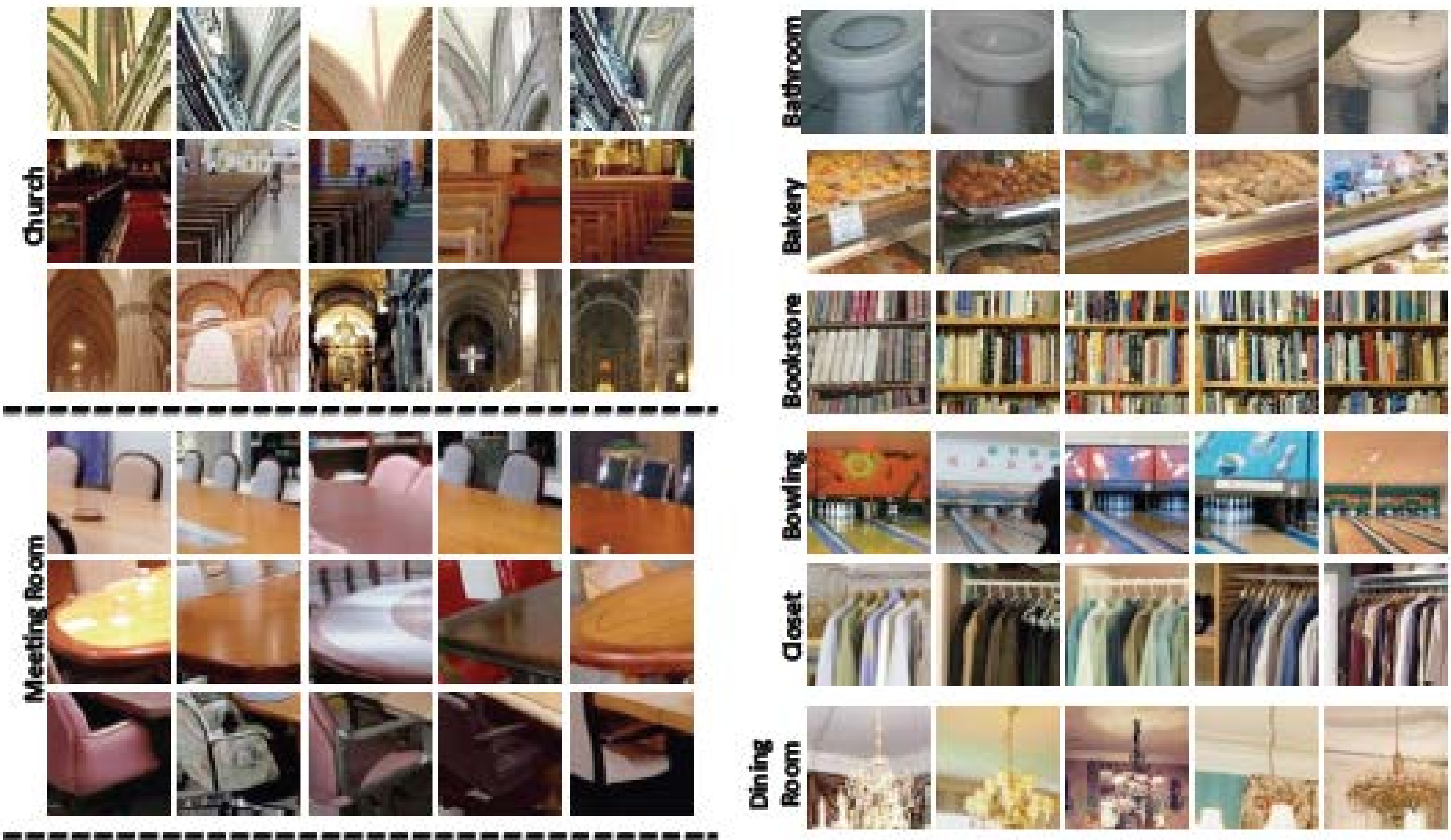
Iter 1

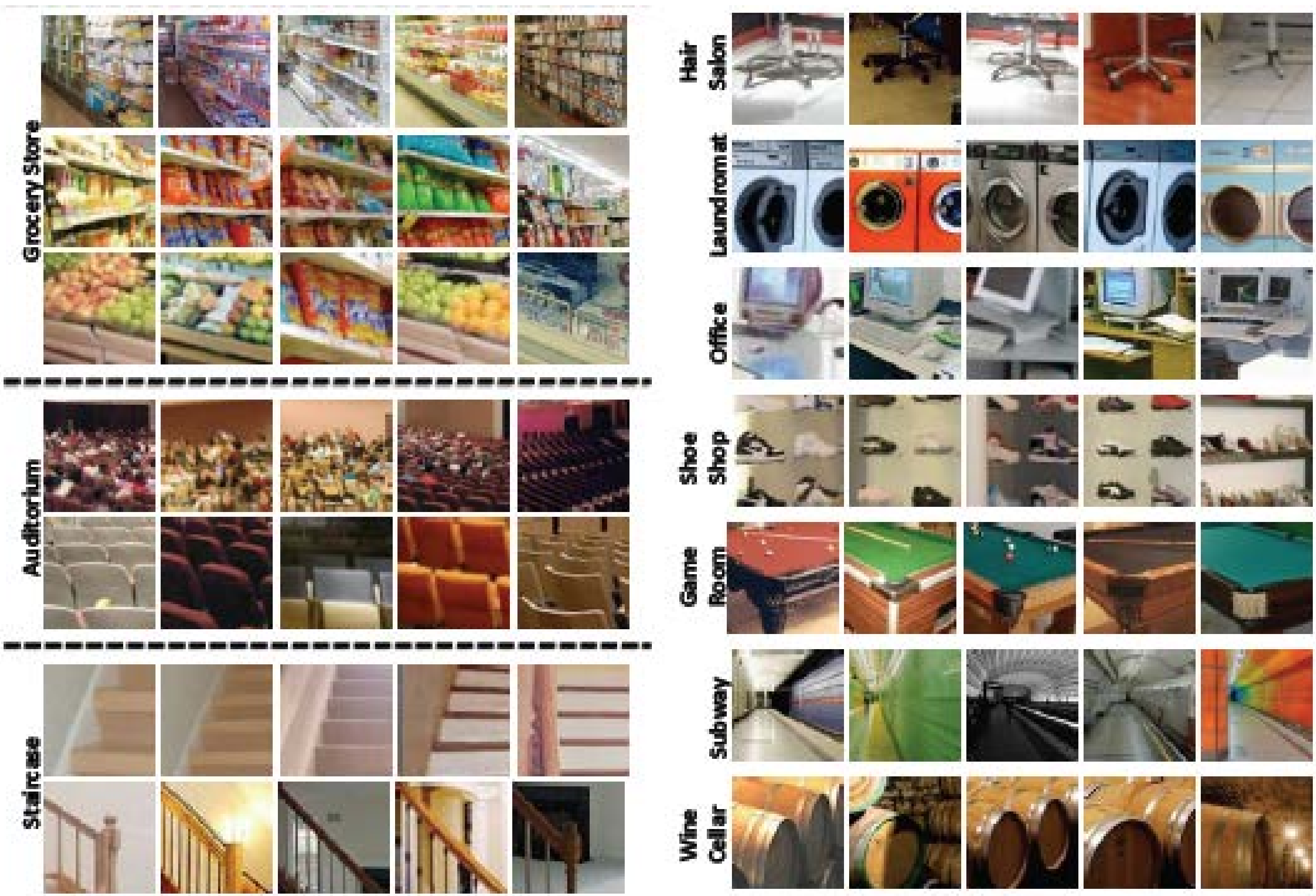
Iter 2

Iter 3

Iter 4

More Discovered Patches


Church
Meeting Room
Bathroom
Bakery
Bookstore
Bowling
Closet
Dining Room



Grocery Store
Auditorium
Staircase
Hair Salon
Laundromat
Office
Shoe Shop
Game Room
Subway
Wine Cellar

What makes Paris look like Paris?

Carl Doersch, Saurabh Singh, Abhinav Gupta,
Josef Sivic and Alexei Efros, [SIGGRAPH ’12]

How well can people do?

- http://baikal.graphics.cs.cmu.edu/cdoersch/im2gps2/corr/test2.html

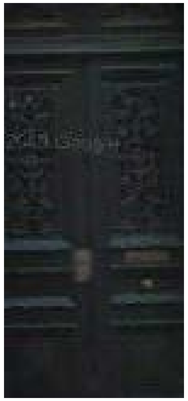
How well can people do?

- http://baikal.graphics.cs.cmu.edu/cdoersch/im2gps2/corr/test2.html
- Mean performance = 79%
  - Subjects who have been to Paris

  (up to 90% if allowed to scrutinize the images)

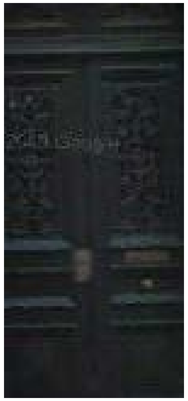
What makes Paris look like Paris?

What makes Paris look like Paris?



Goal

To automatically discover geo-informative visual elements, that (hopefully) capture the “look and feel” of a place



Goal

To automatically discover geo-informative visual elements, that are:

- **Representative:** frequently occur in Paris.

- **Discriminative:** occur in Paris but not at other places.

Need both conditions

Discriminative only:

Need both conditions

Frequently occurring only:

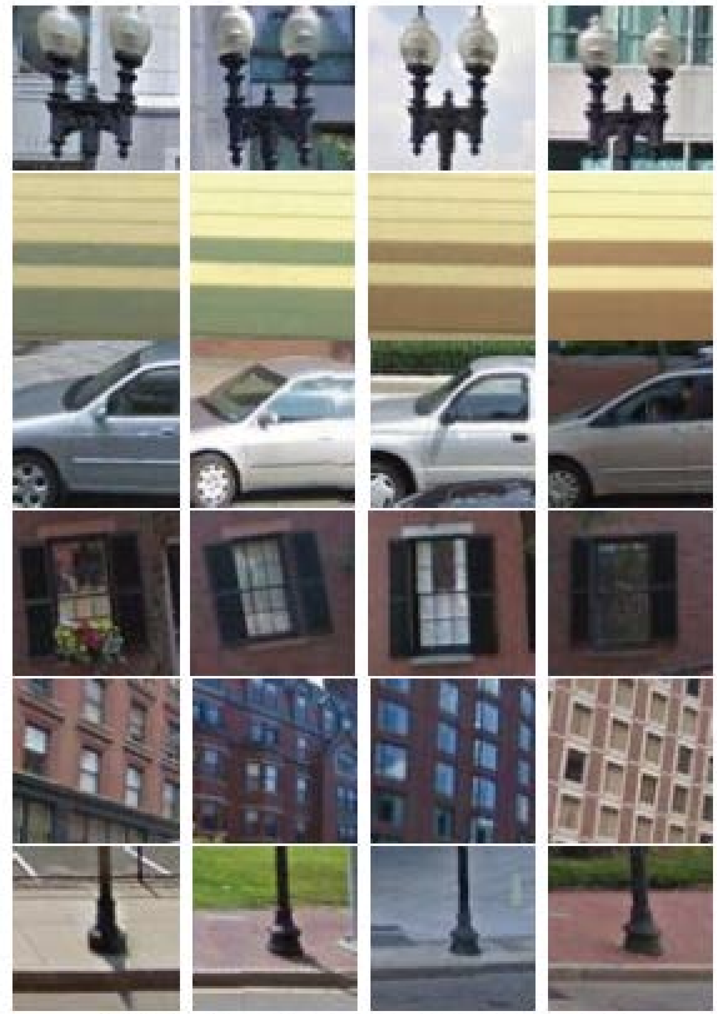

RESTAURAN
PASCAL NAIM
ALéGRIA
bar

Extracted Visual Elements from Prague

Extracted Elements from London

Extracted Elements from Barcelona

Extracted Elements from SF

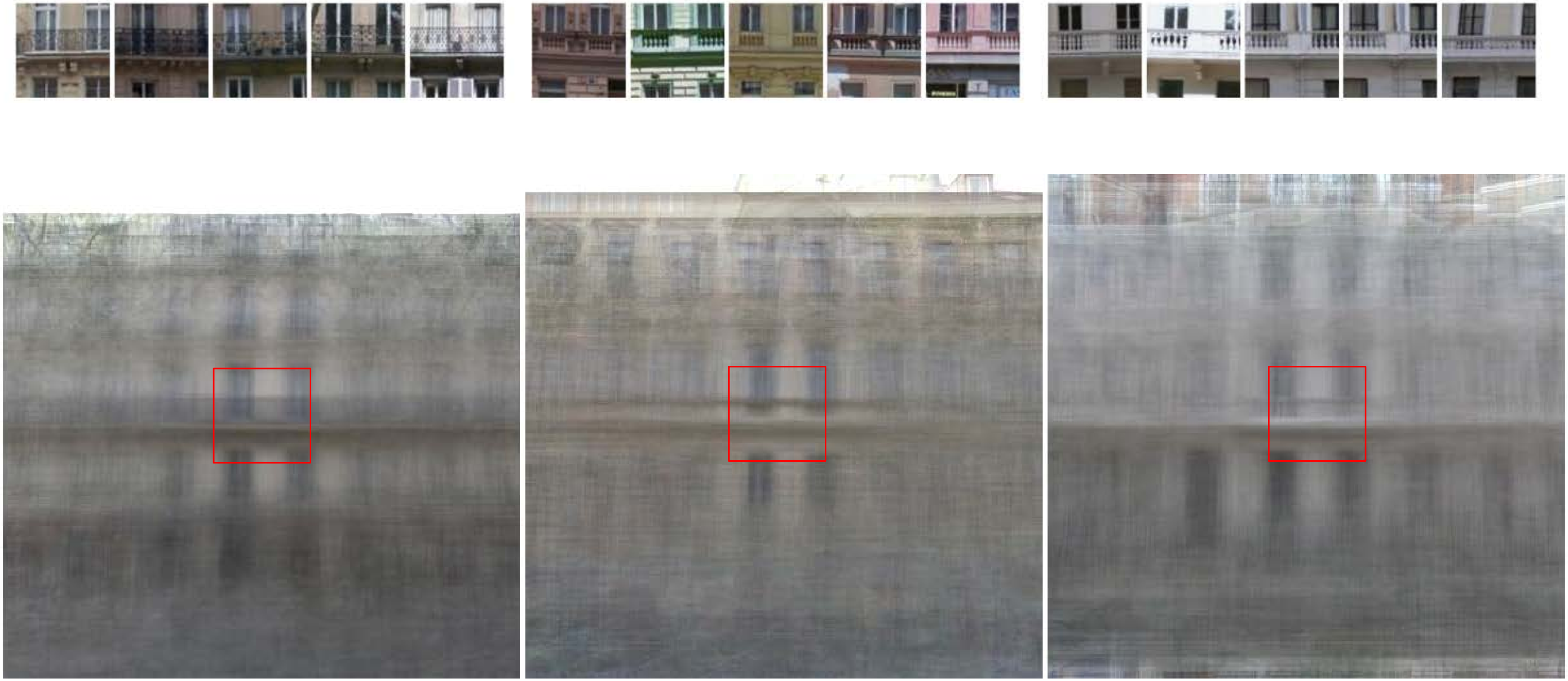


Extracted Elements from Boston

Correspondence

Paris, France   Prague, Czech Republic   London, England

Many Elements Capture Context



Mapping architectural features

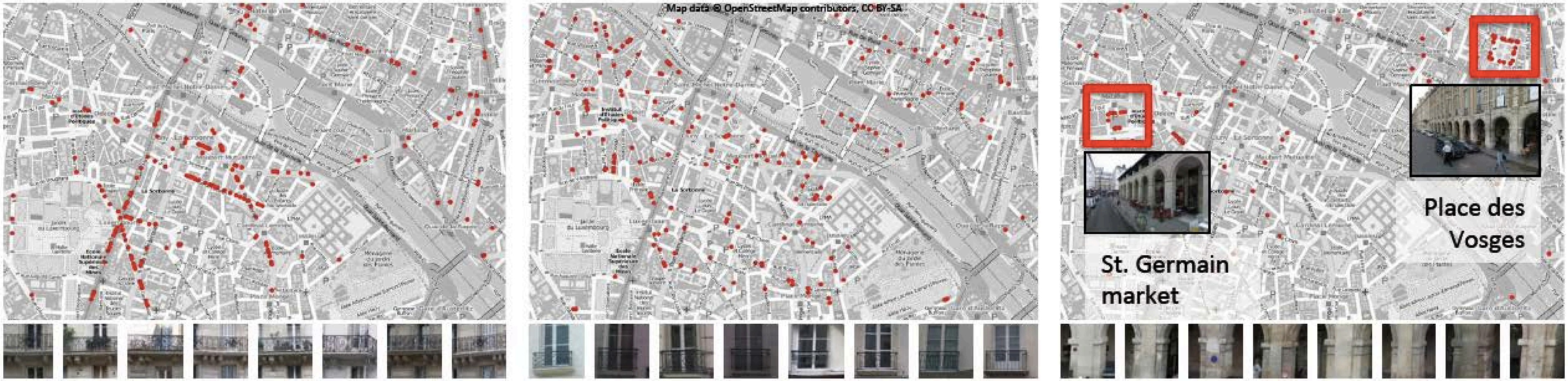
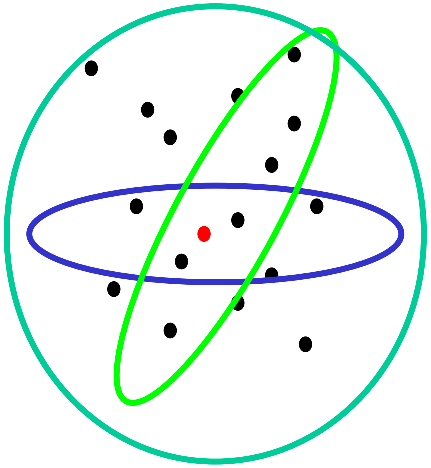
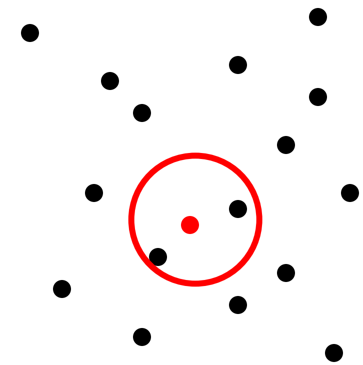


Figure 6: *Examples of geographic patterns in Paris (shown as red dots on the maps) for three discovered visual elements (shown below each map). Balconies with cast-iron railings are concentrated on the main boulevards (left). Windows with railings mostly occur on smaller streets (middle). Arch supporting columns are concentrated on Place des Vosges and the St. Germain market (right).*

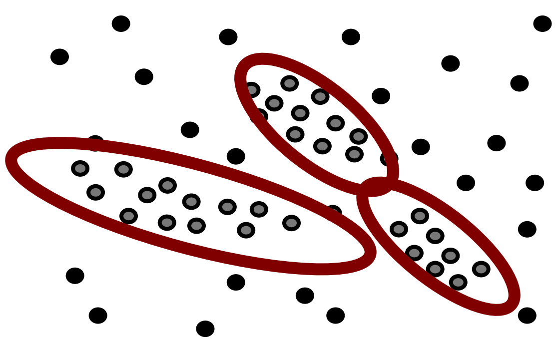
Ways to use Big Data



1. See what different subsets of data think of you



2. Use kNN to make a smaller, cozier sub-problem



3. Visual Data Mining: find needles in a haystack

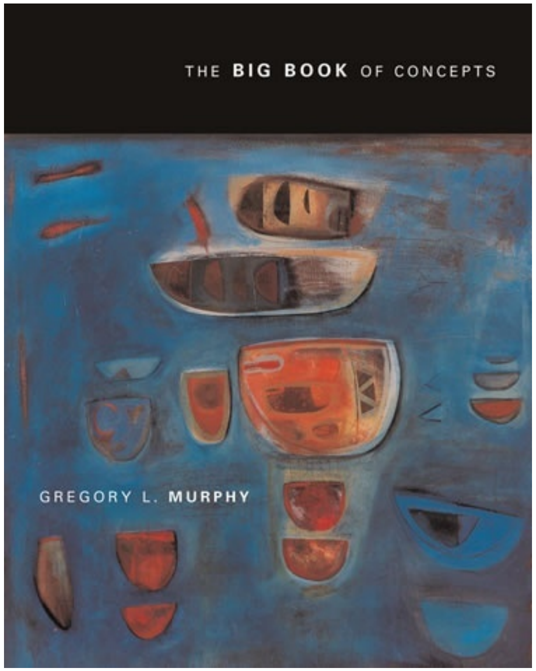
4. Ditch Categories – keep all instances and connect them

Down with Categories!!!

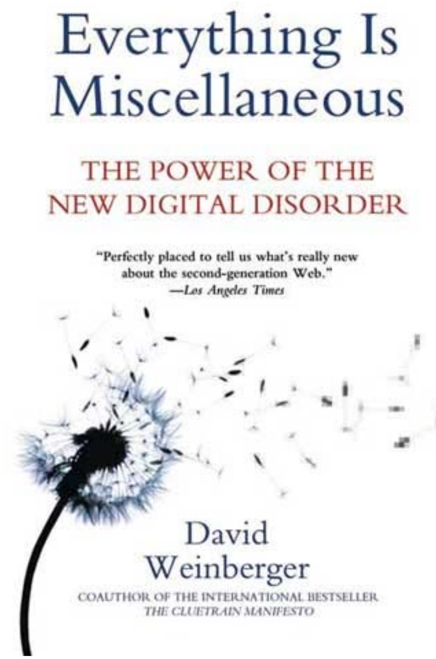
Alexei (Alyosha) Efros
CMU

Joint work with
Tomasz Malisiewicz

Acknowledgements



Murphy
Big Book of Concepts



Weinberger
Everything is Miscellaneous

Talks by Moshe Bar; writings of Shimon Edelman

Many great discussions with many colleagues, especially
Tomasz Malisiewicz, James Hays, and Derek Hoiem

Understanding an Image

slide by Fei Fei, Fergus & Torralba

Object naming -> Object categorization

slide by Fei Fei, Fergus & Torralba

Object categorization

sky

building

flag

face

banner

wall

street lamp

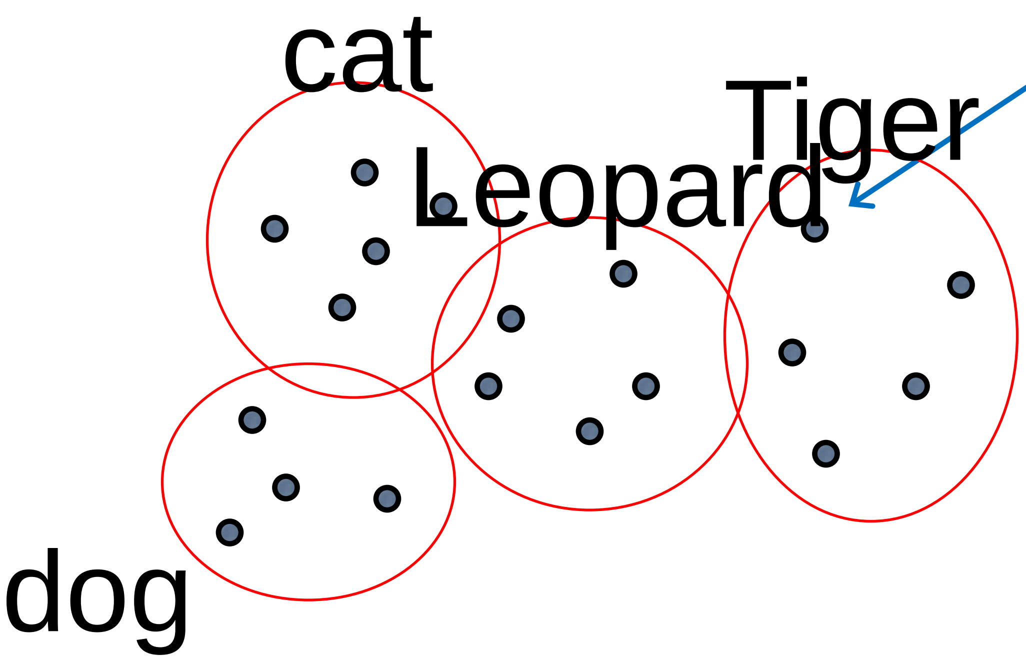
bus

bus

cars

Why Categorize?

1. Knowledge Transfer
2. Communication



Classical View of Categories

- Dates back to Plato & Aristotle

  1. Categories are defined by a list of properties shared by all elements in a category
  2. Category membership is binary
  3. Every member in the category is equal

Problems with Classical View

- Humans don’t do this!
  - People don’t rely on abstract definitions / lists of shared properties (Wittgenstein 1953, Rosch 1973)
    - e.g. define the properties shared by all “games”
    - e.g. are curtains furniture? Are olives fruit?
  - Typicality
    - e.g. Chicken -> bird, but bird -> eagle, pigeon, etc.
  - Language-dependent
    - e.g. “Women, Fire, and Dangerous Things” category is Australian aboriginal language (Lakoff 1987)
  - Doesn’t work even in human-defined domains
    - e.g. Is Pluto a planet?

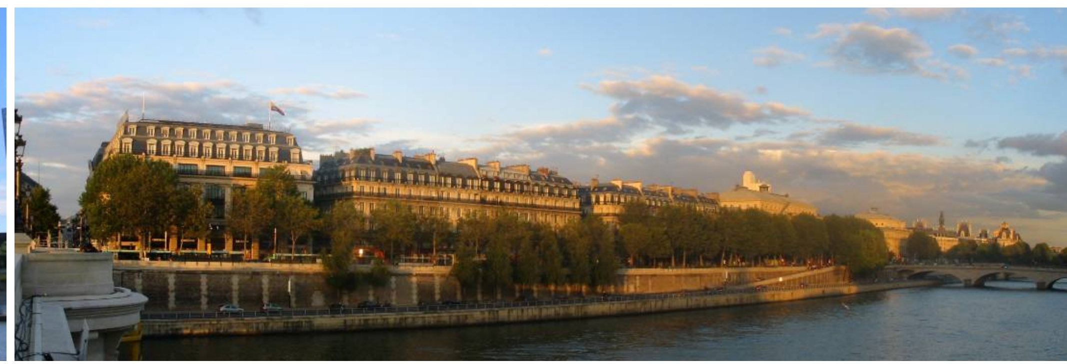
Problems with <u>Visual</u> Categories

- A lot of categories are functional

Chair



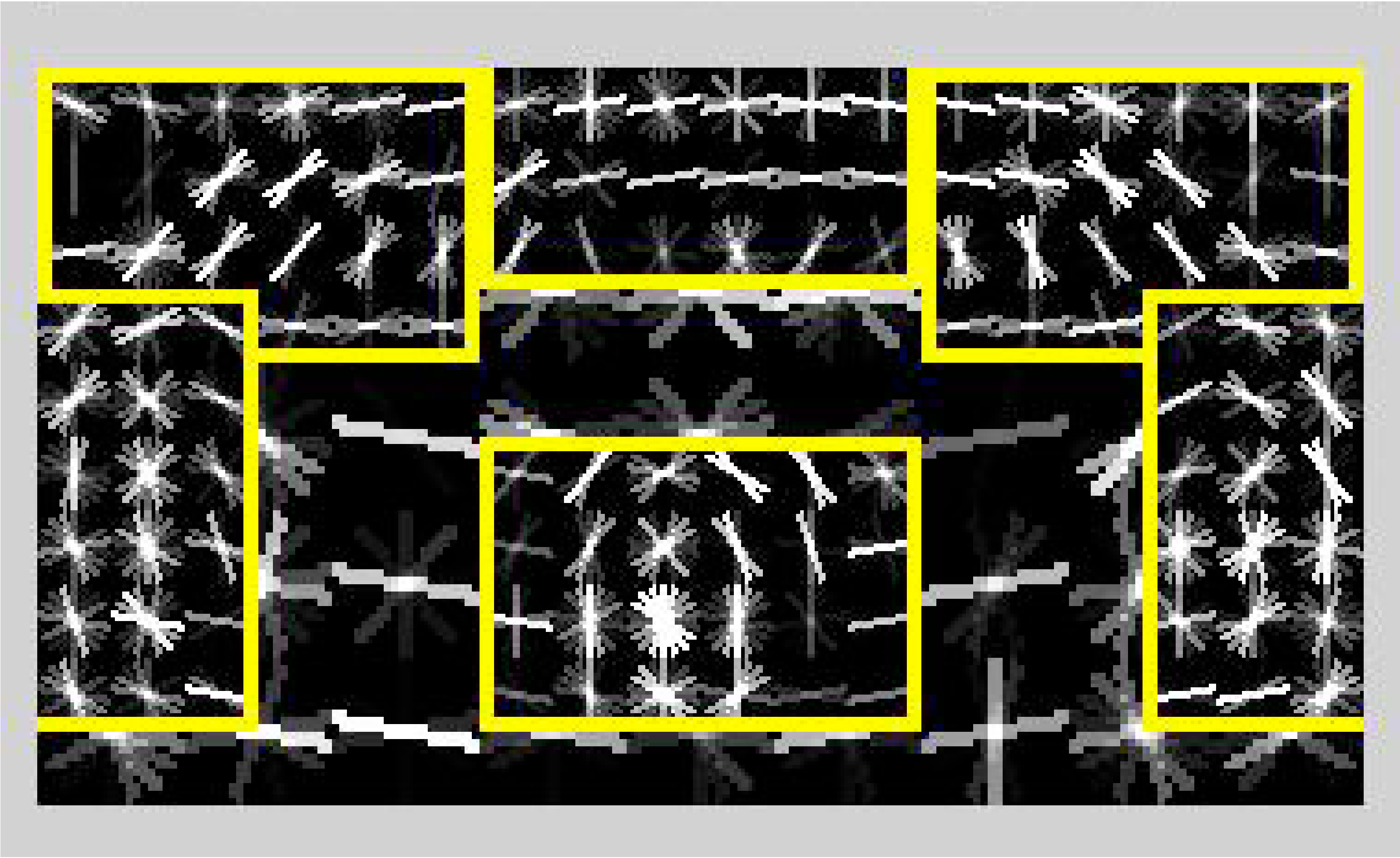
- World is too varied

- Categories are 3D, but images are 2D

car


Typical HOG car detector



Felzenszwalb et al, PASCAL 2

Why not?

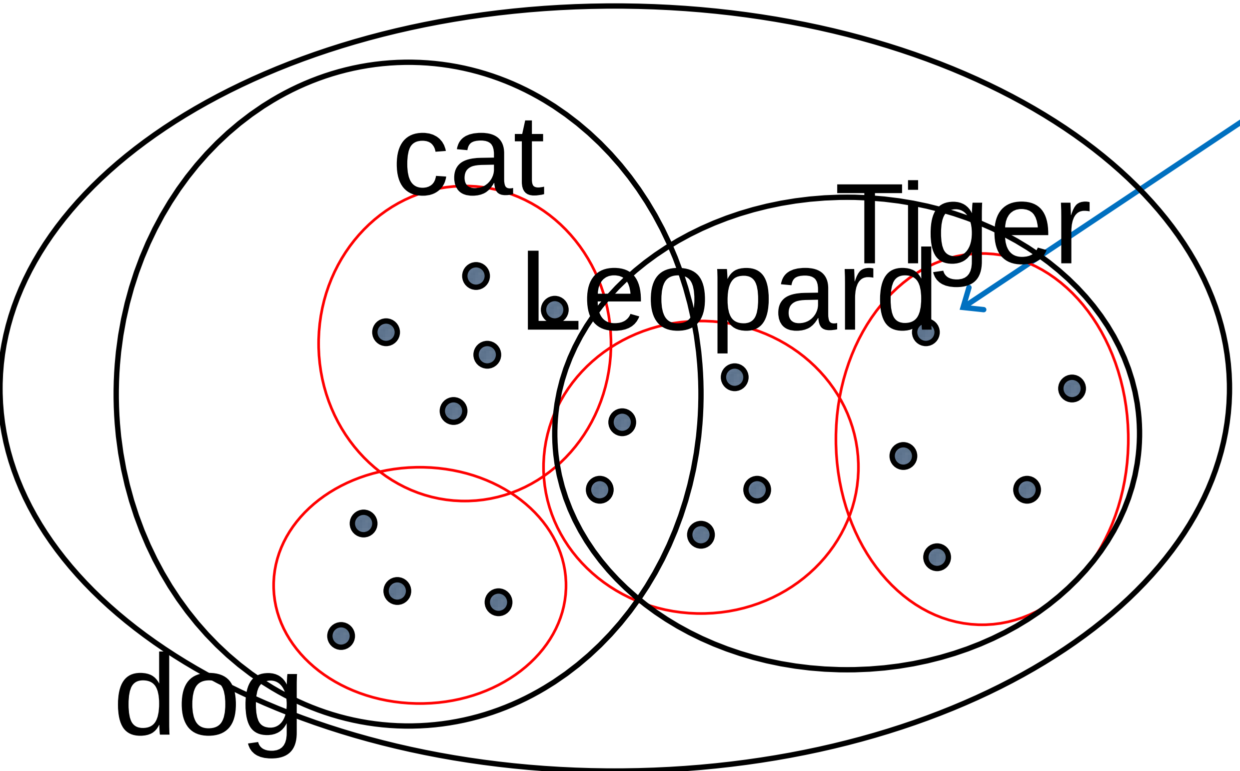

ABSOLUT
Country of Sweden
VODKA
This superb vodka
was distilled from grain grown
in the rich fields of southern Sweden
It has been produced at the famous
old distilleries near Åhus
in accordance with more than
400 years of Swedish tradition
Vodka has been sold under the name
Absolut since 1879.
40% ALC./VOL. (80 PROOF) 1 LITER
IMPORTED
PRODUCED AND BOTTLED IN AHUS, SWEDEN
BY THE ABSOLUT COMPANY
A DIVISION OF V&S VIN&SPRIT AB

+

Solution: hierarchy?

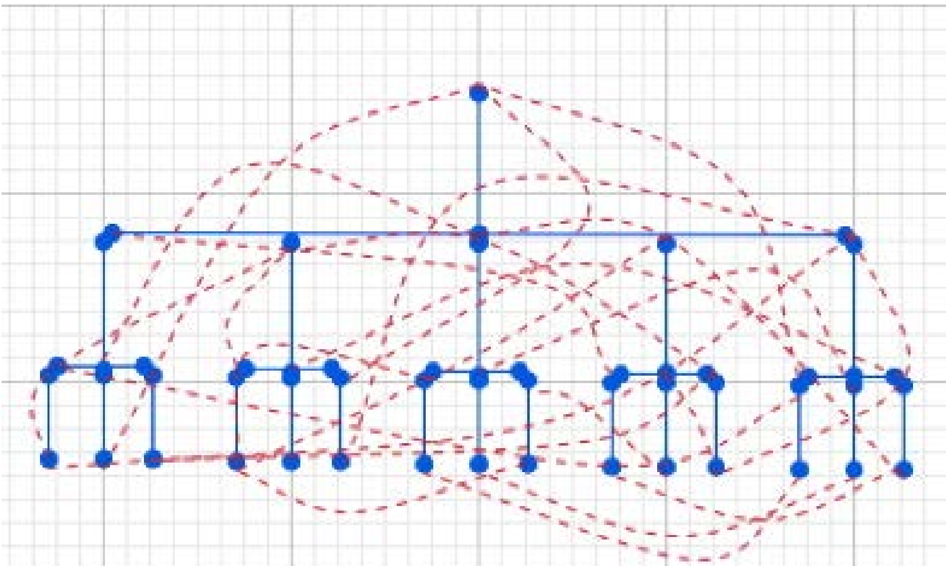
Ontologies, hierarchies, levels of categories (Rosch), etc.

WordNet, ImageNet, etc etc



Still Problematic!

- Intransitivity
  - e.g. car seat is chair, chair is furniture, but …
- Multiple category membership
  - it’s not a tree, it’s a forest!



Clay Shirky, “Ontologies are Overrated”

Fundamental Problem with Categorization

Making decisions too early!

We should only categorize at run-time, once we know the task!

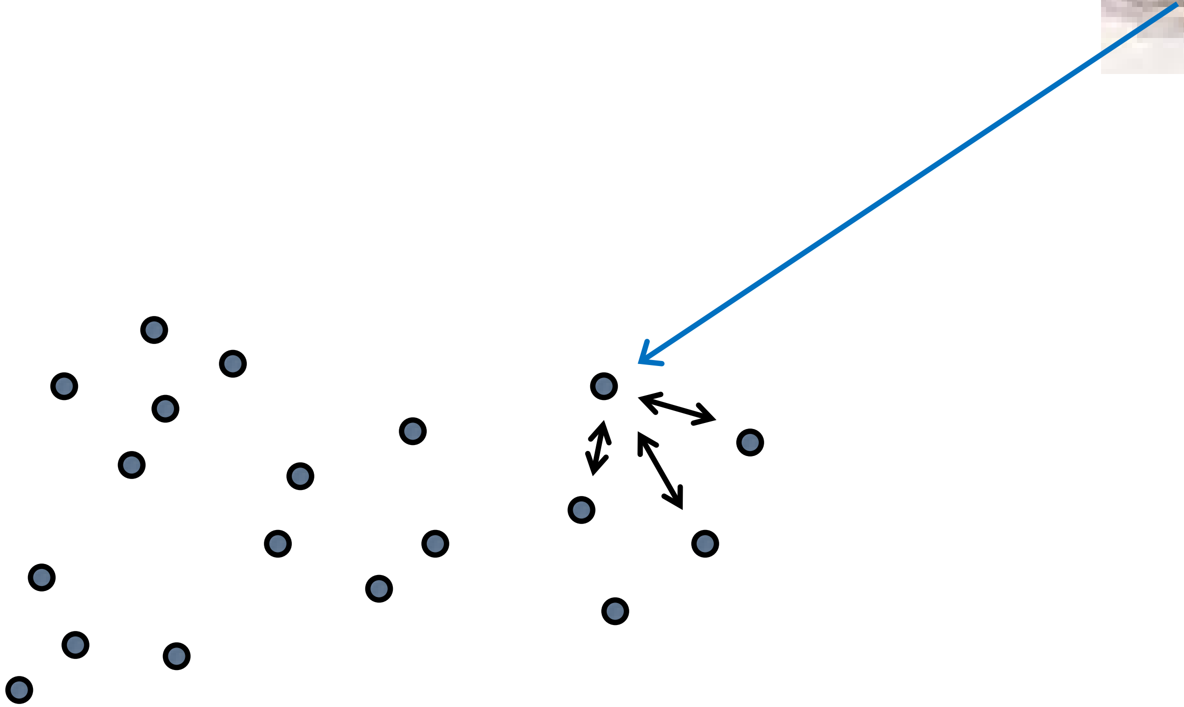
The Dictatorship of Librarians

categories are losing…

YAHOO! vs. Google

On-the-fly Categorization?

1. Knowledge Transfer
2. ~~Communication~~



Association instead of categorization

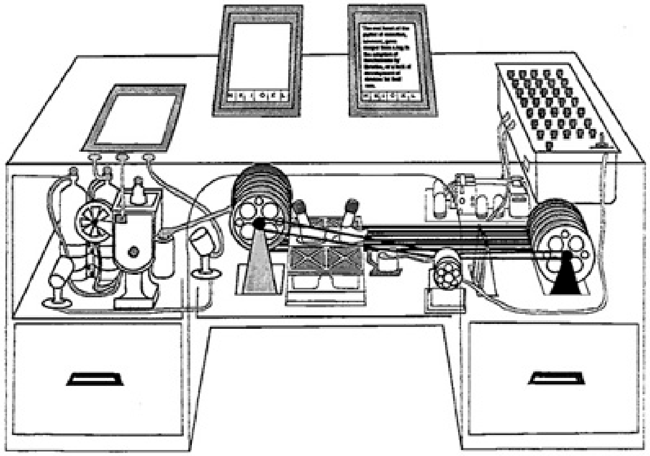
Ask not "what is this?", ask "what is this <u>like</u>"

– Moshe Bar

- Exemplar Theory (Medin & Schaffer 1978, Nosofsky 1986, Krushke 1992)
  - categories represented in terms of remembered objects (exemplars)
  - Similarity is measured between input and all exemplars
  - *think* non-parametric density estimation
- Vanevar Bush (1945), <u>Memex</u> (MEMory EXtender)
  - Inspired hypertext, WWW, Google…

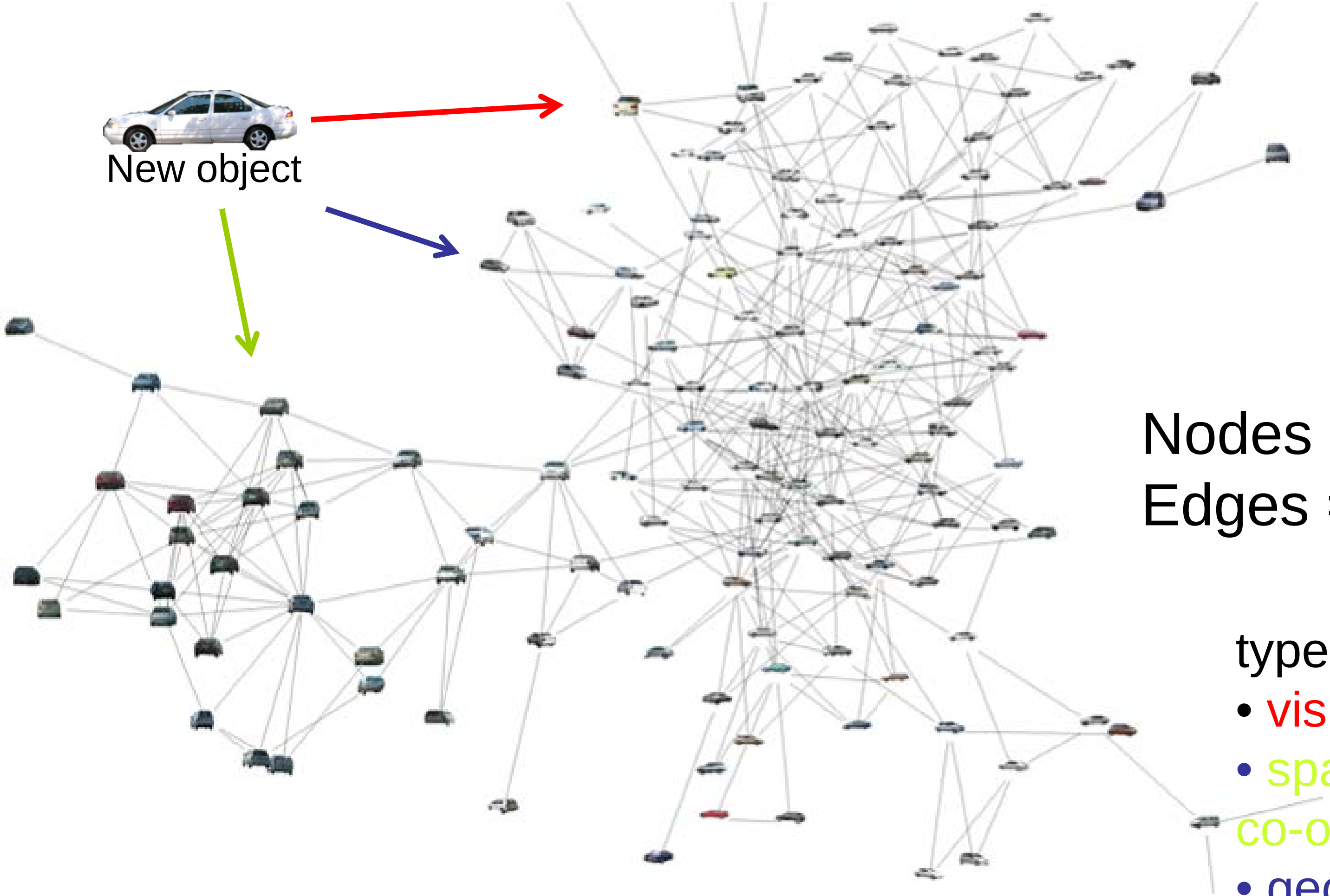
Bush’s Memex (1945)

- Store publications, correspondence, personal work, on microfilm
- Items retrieved rapidly using index codes
  - Builds on “rapid selector”
- Can annotate text with margin notes, comments
- Can construct a *trail* through the material and save it
  - Roots of hypertext
- Acts as an external memory



Visual Memex, a proposal

[Malisiewicz & Efros]

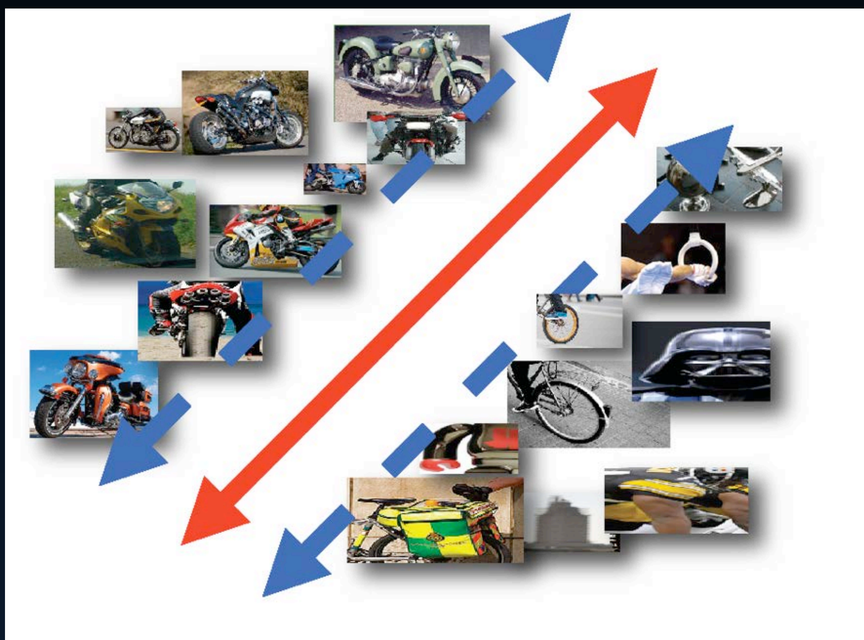


Nodes = instances
Edges = associations

types of edges:

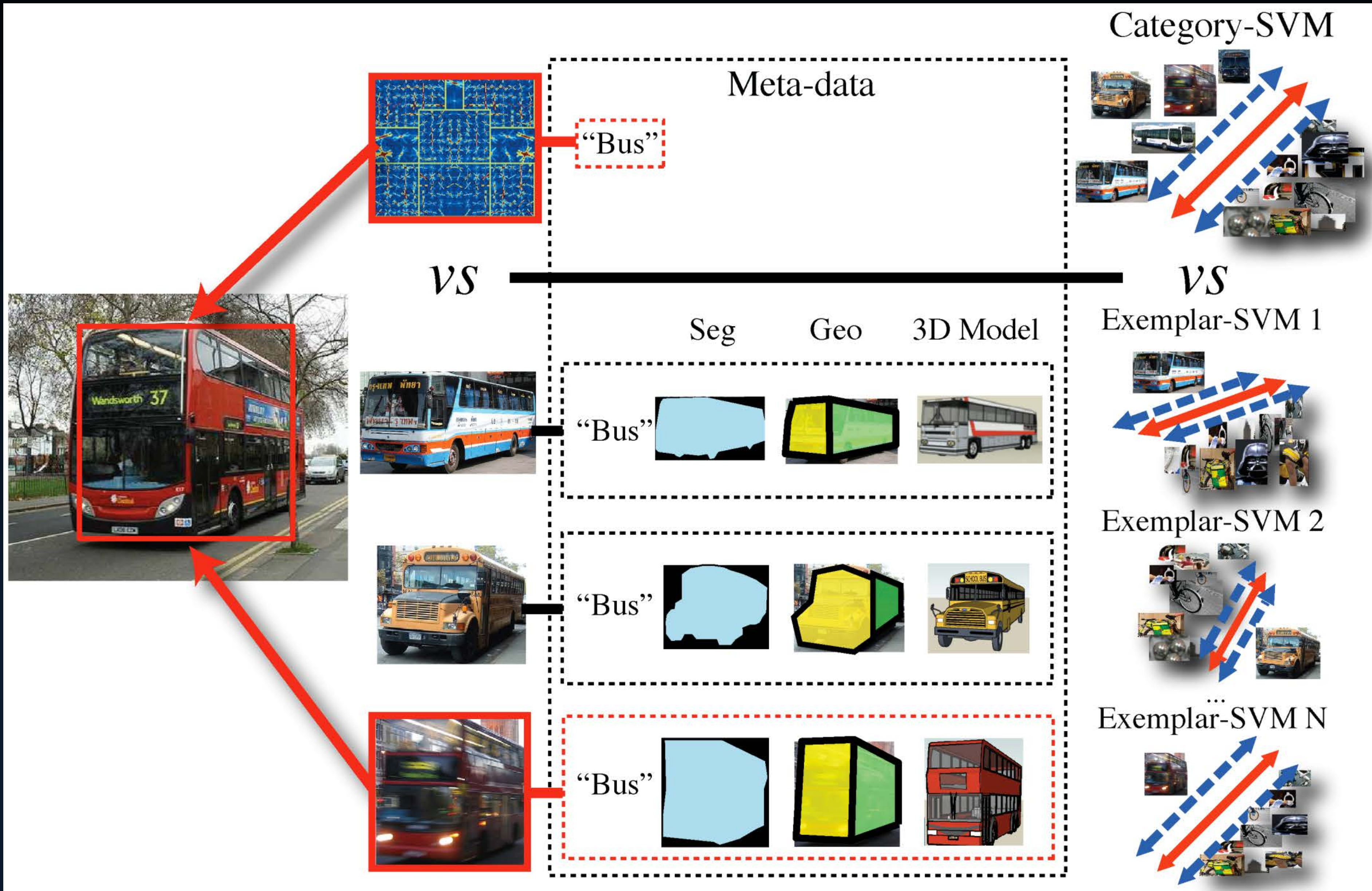
- visual similarity
- spatial, temporal co-occurrence
- geometric structure
- language
- geography
- ..

Object Detection

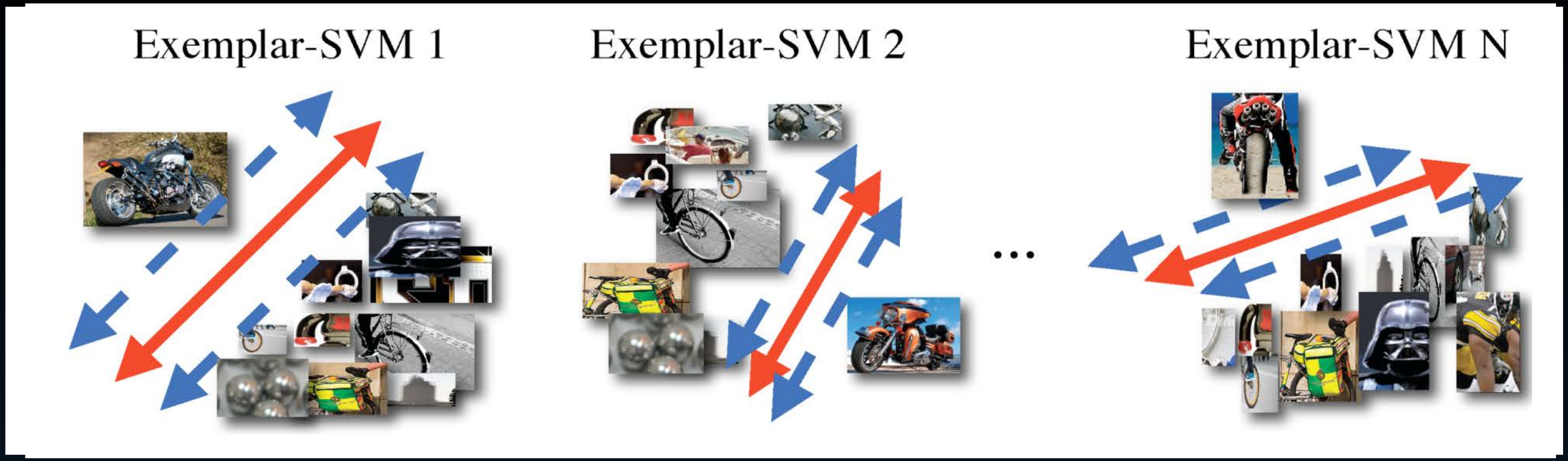


“bus”

Exemplar Object Detection

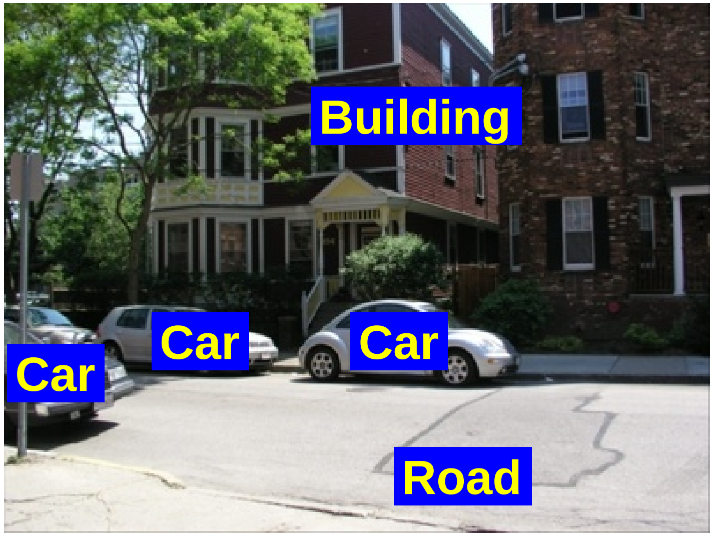


Ensemble of Exemplar-SVMs



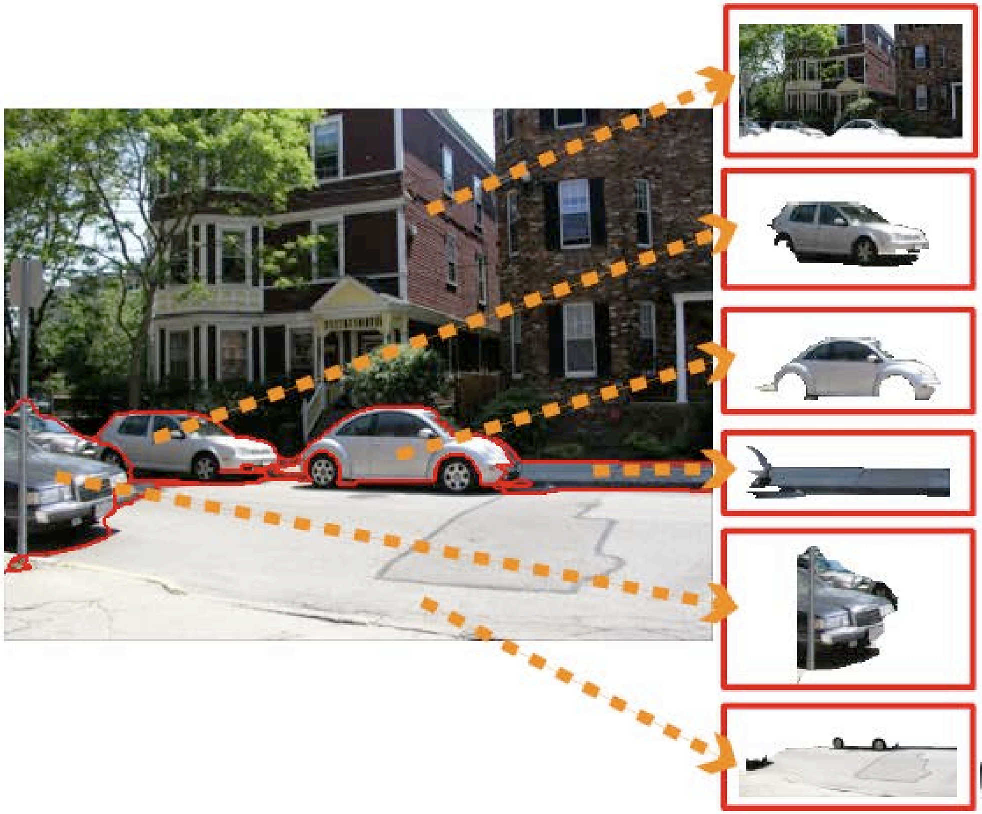
“What is this?”

Input Image



He 2004, Tu 2004, Shotton 2006, Galleguillos 2008, Fei-Fei 2009, Gould 2009, etc.

“What is this *like*?”



Malisiewicz & Efros, CVPR’08

Image Parsing with Context

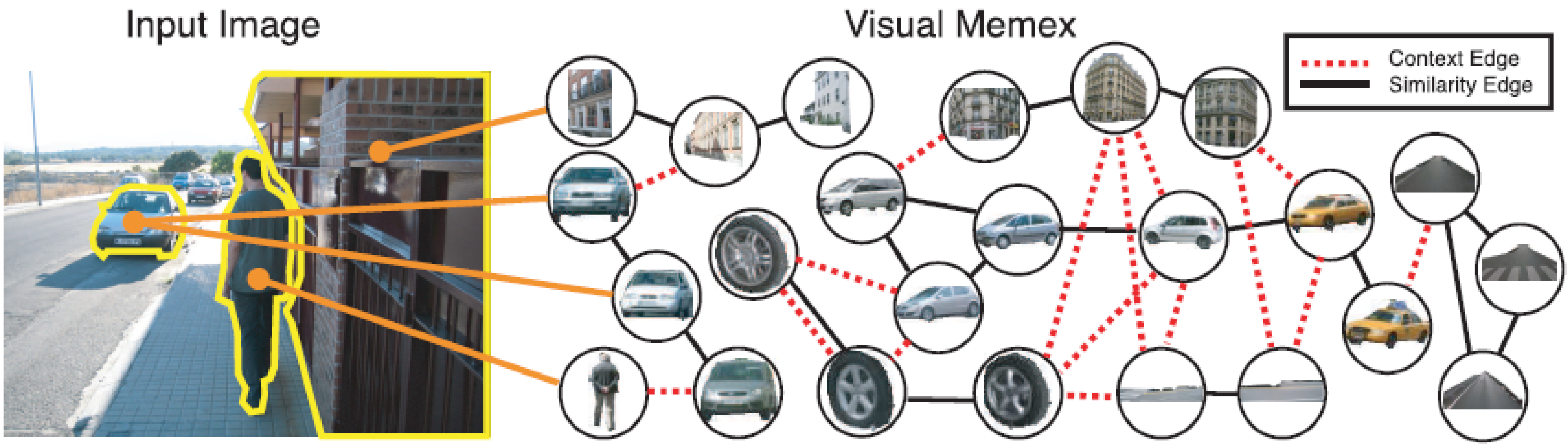


Figure 1: The **Visual Memex** graph encodes object similarity (solid black edge) and spatial context (dotted red edge) between pairs of object exemplars. A spatial context feature is stored for each context edge. The Memex graph can be used to interpret a new image (left) by associating image segments with exemplars in the graph (orange edges) and propagating the information.

Torralba’s Context Challenge

Torralba's Context Challenge

Slide by Antonio Torralba

Torralba's Context Challenge

Slide by Antonio Torralba

Our Challenge Setup

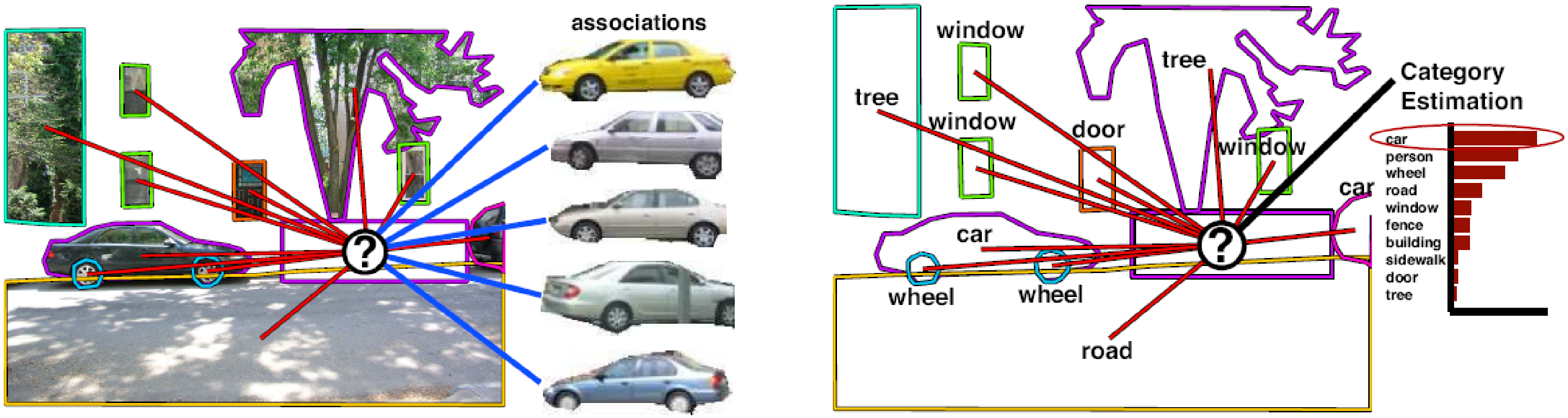


Figure 2: Torralba's Context Challenge: "How far can you go without running a local object detector?" The task is to reason about the identity of the hidden object (denoted by a "?") without local information. In our category-free Visual Memex model, object predictions are generated in the form of exemplar associations for the hidden object. In a category-based model, the category of the hidden object is directly estimated.

Malisiewicz & Efros, NIPS'09

3 models

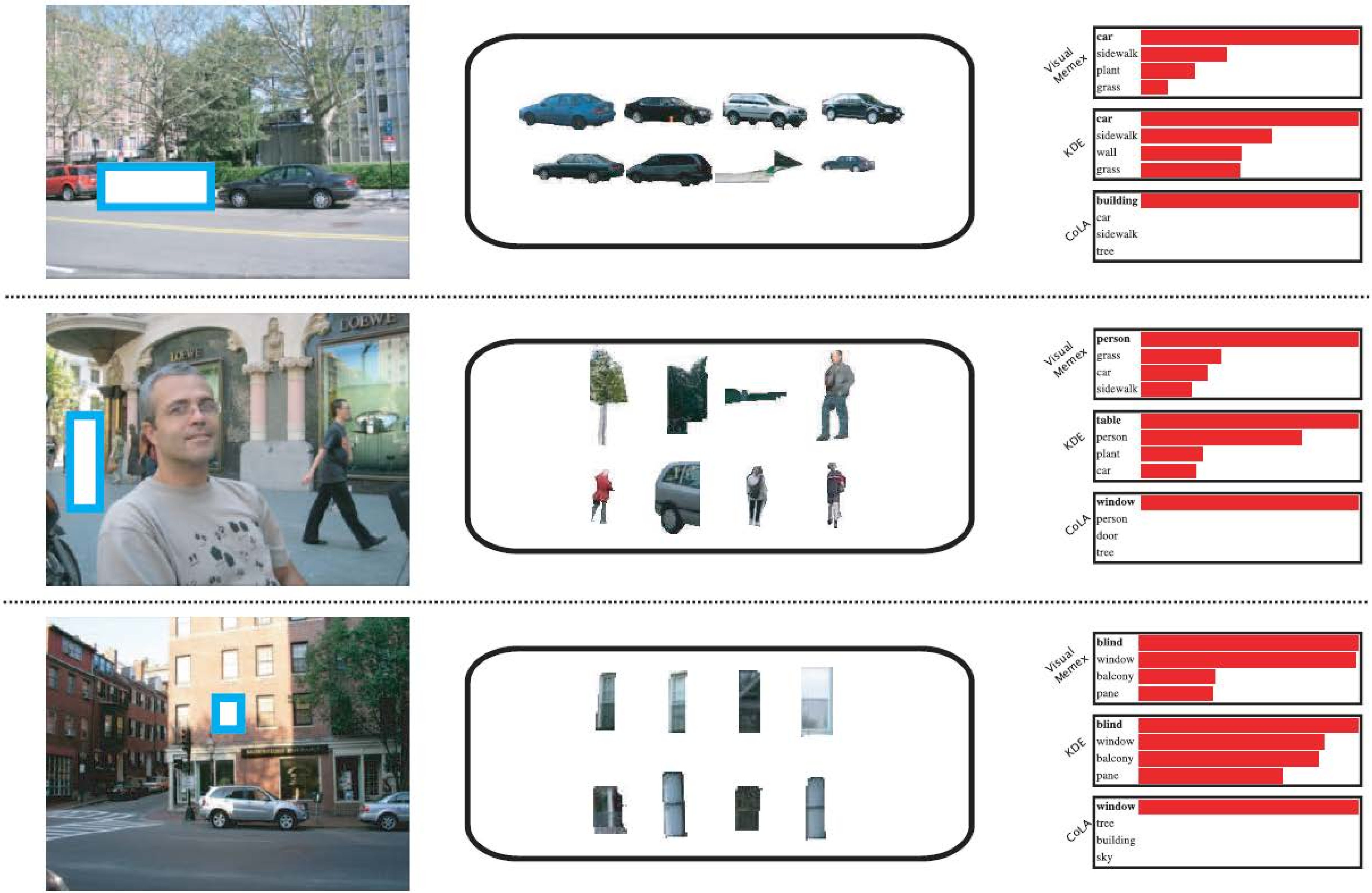
Visual Memex: exemplars, non-parametric object-object relationships

- Recurse through the graph

Baseline: CoLA: categories, parametric object-object relationships

Reduced Memex: categories, non-parametric relationships

Qual. results


LOEWE
Visual Memex
car
sidewalk
plant
grass
KDE
car
sidewalk
wall
grass
CoLA
building
car
sidewalk
tree
Visual Memex
person
grass
car
sidewalk
KDE
table
person
plant
car
CoLA
window
person
door
tree
Visual Memex
blind
window
balcony
pane
KDE
blind
window
balcony
pane
CoLA
window
tree
building
sky

Quant. results

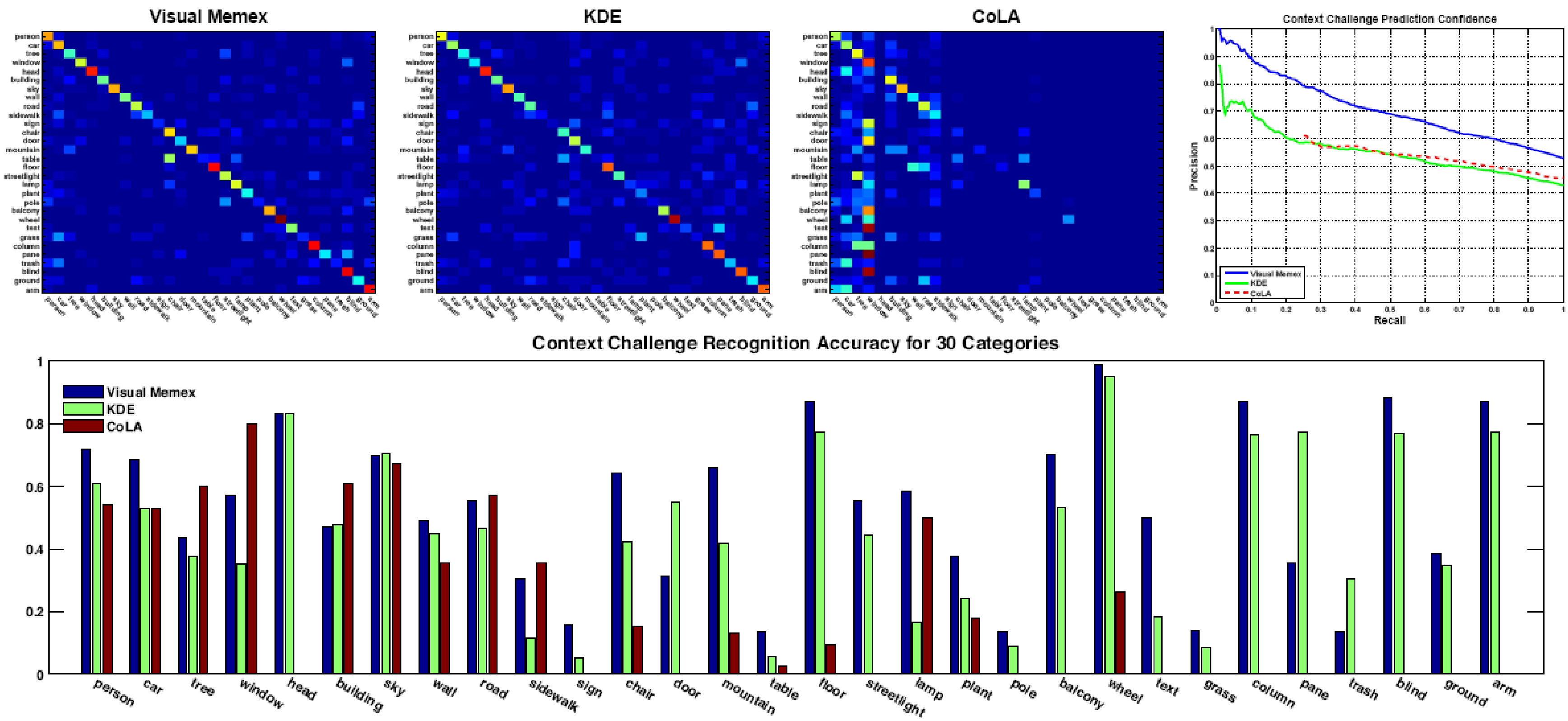


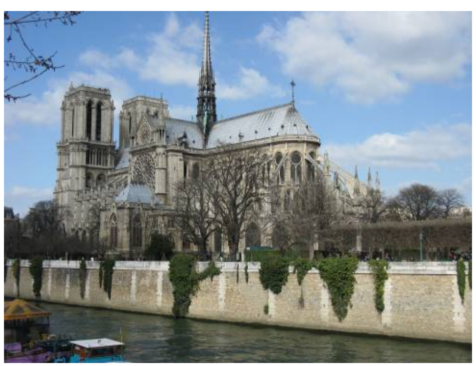
Figure 3: a.) Context Challenge confusion matrices for the 3 methods: Visual Memex, KDE, and CoLA. b.) Recognition Precision versus Recall when thresholding output based on confidence. c) Side by side comparison of the 3 methods' accuracies for 30 categories.

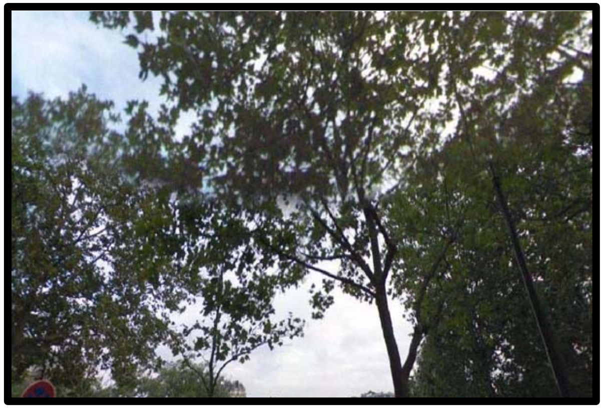
will Big Data solve all your problems?

1. Data is Biased

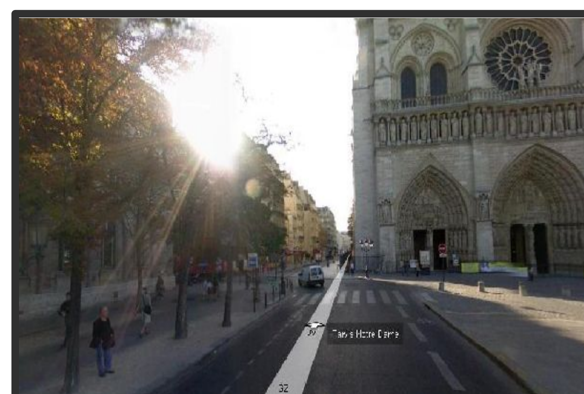
- Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
- But it's <u>not</u> random samples of visual world

Flickr Paris





Google StreetView Paris

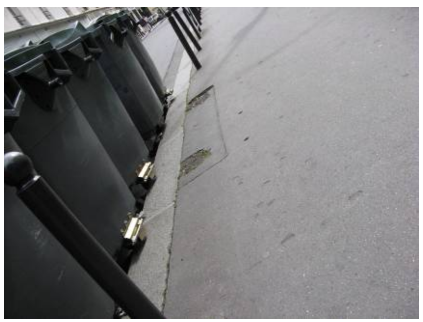


Knopp, Sivic, Pajdla, ECCV 2010

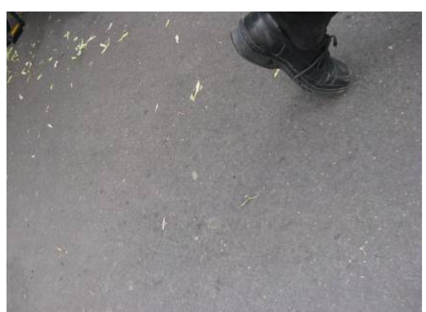
My Paris



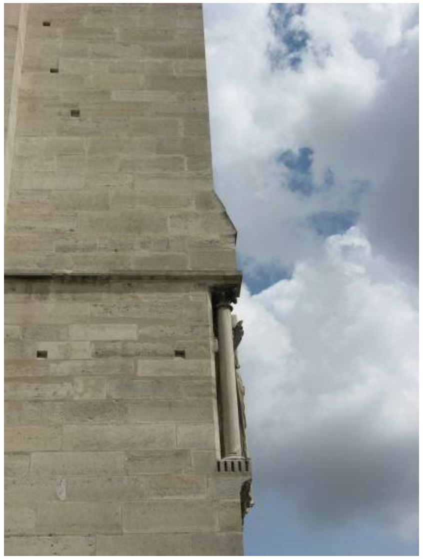


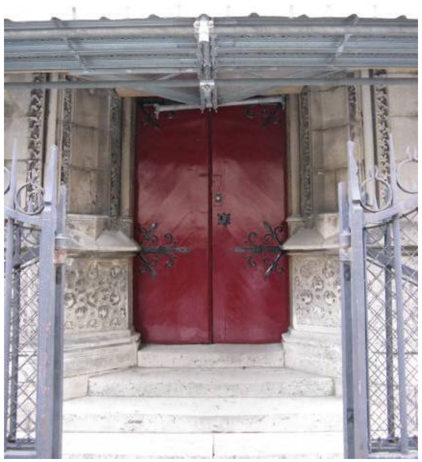




Real Notre Dame



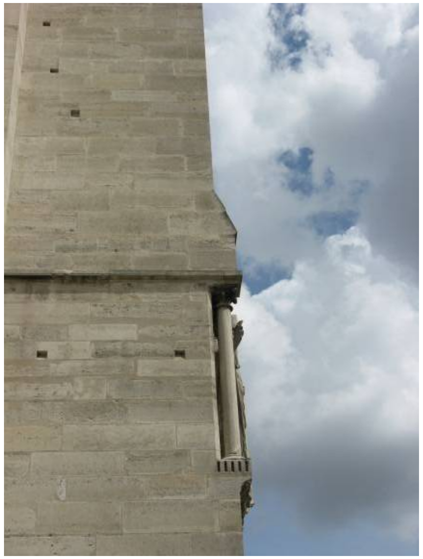


Sampling Bias

- People like to take pictures on vacation

Photographer Bias

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



vs.

Social Bias

Little Leaguer

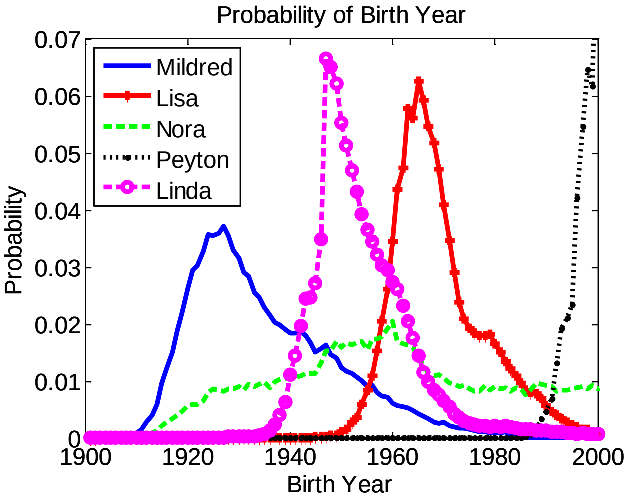
Kids with Santa

The Graduate

Newlyweds

“100 Special Moments” by Jason Salavon

Social Bias



Mildred and Lisa

Source: U.S. Social Security Administration

Gallagher et al CVPR 2008

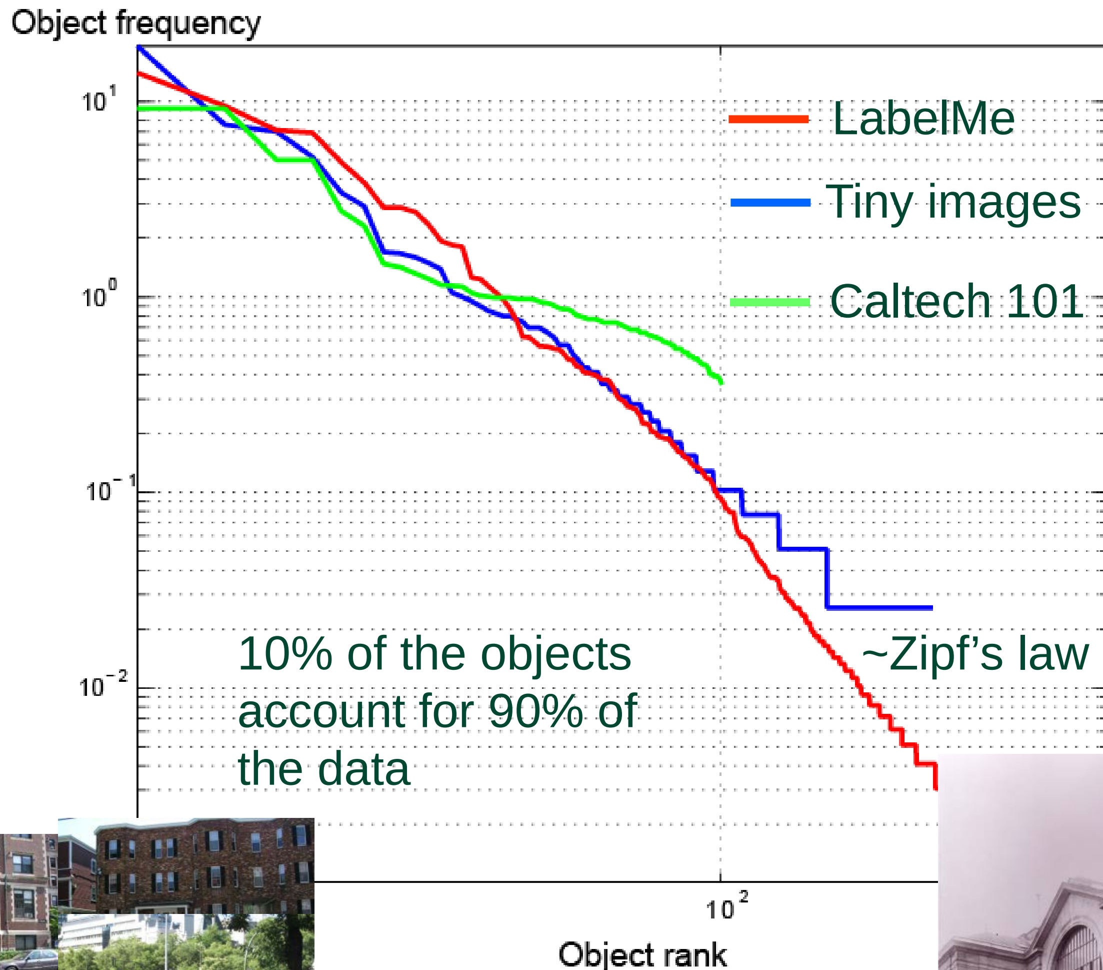
Social Bias

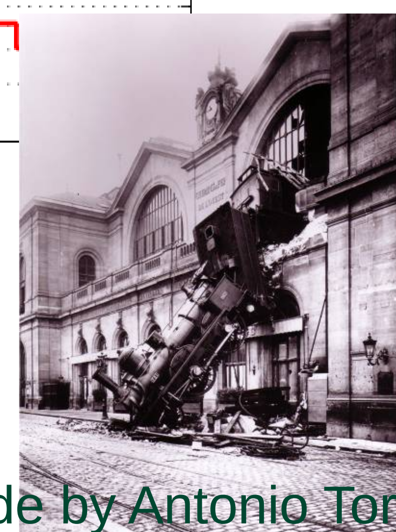
Gallagher et al CVPR 2008

Gallagher et al, CVPR 2009

2. We will never have enough data

Long Tails -- Unfamiliar is Common





Slide by Antonio Torralba

Dealing with sparse data (rare scenes)

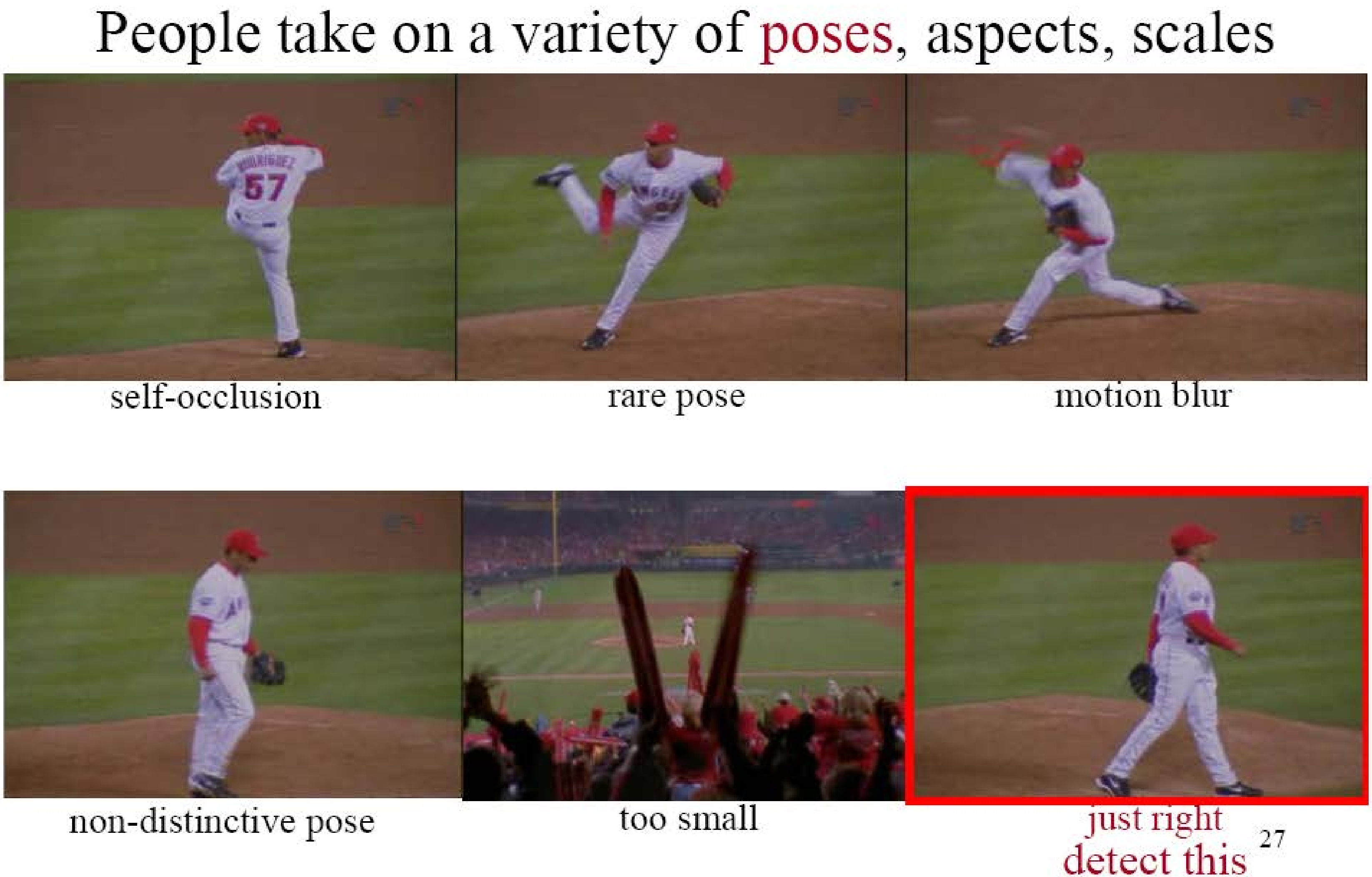
Quick Fixes:

better alignment

- e.g. reduce resolution, sifting, warping, etc.

Understand the simple stuff first

Recognize when it's easy!



Ramanan, Forsyth, Zisserman, 2004

“Poping out” foreground objects

Hoiem et al, ICCV 2007

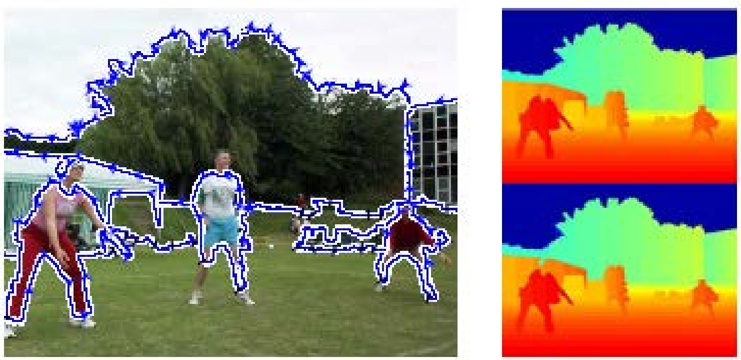
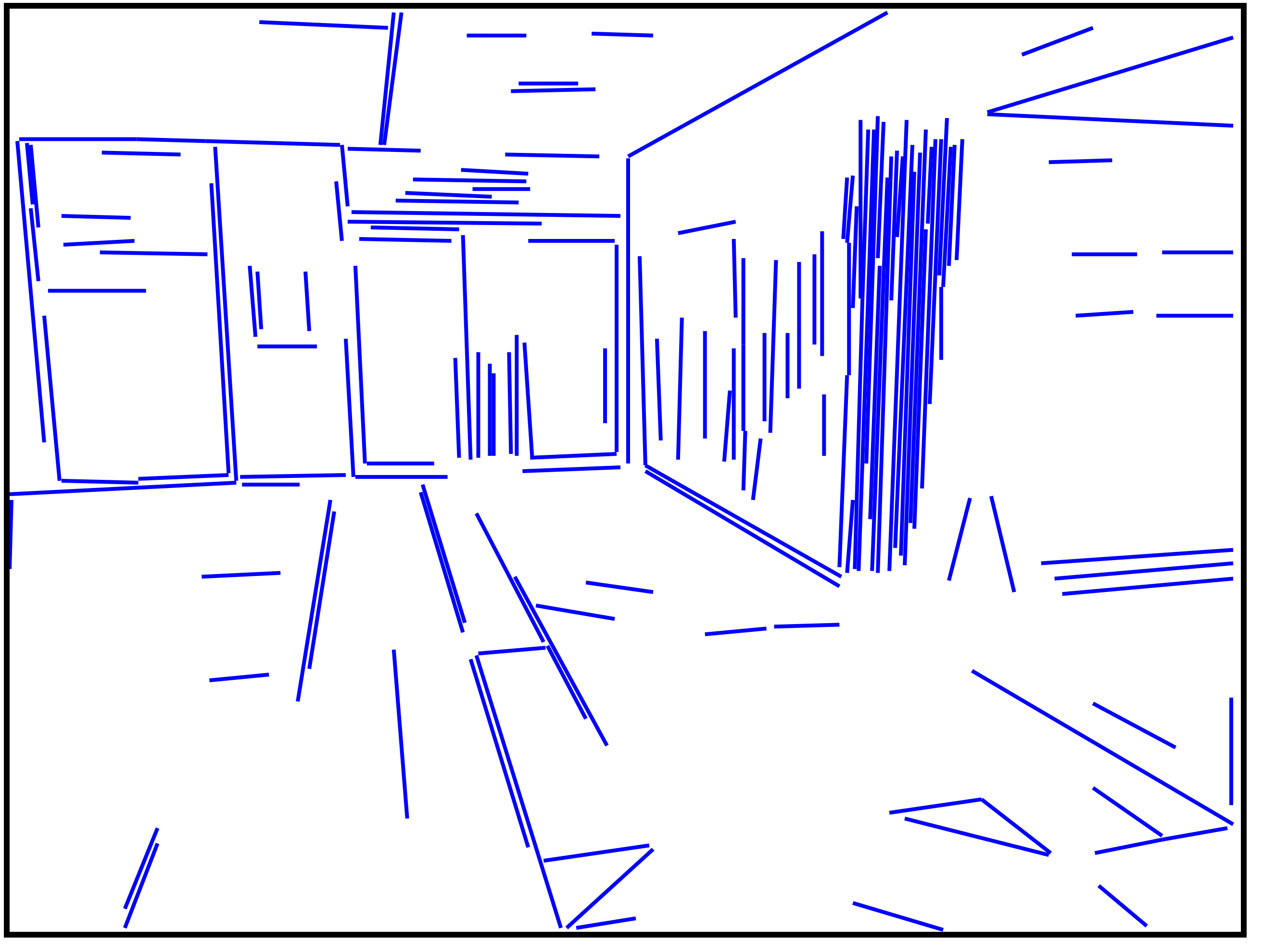


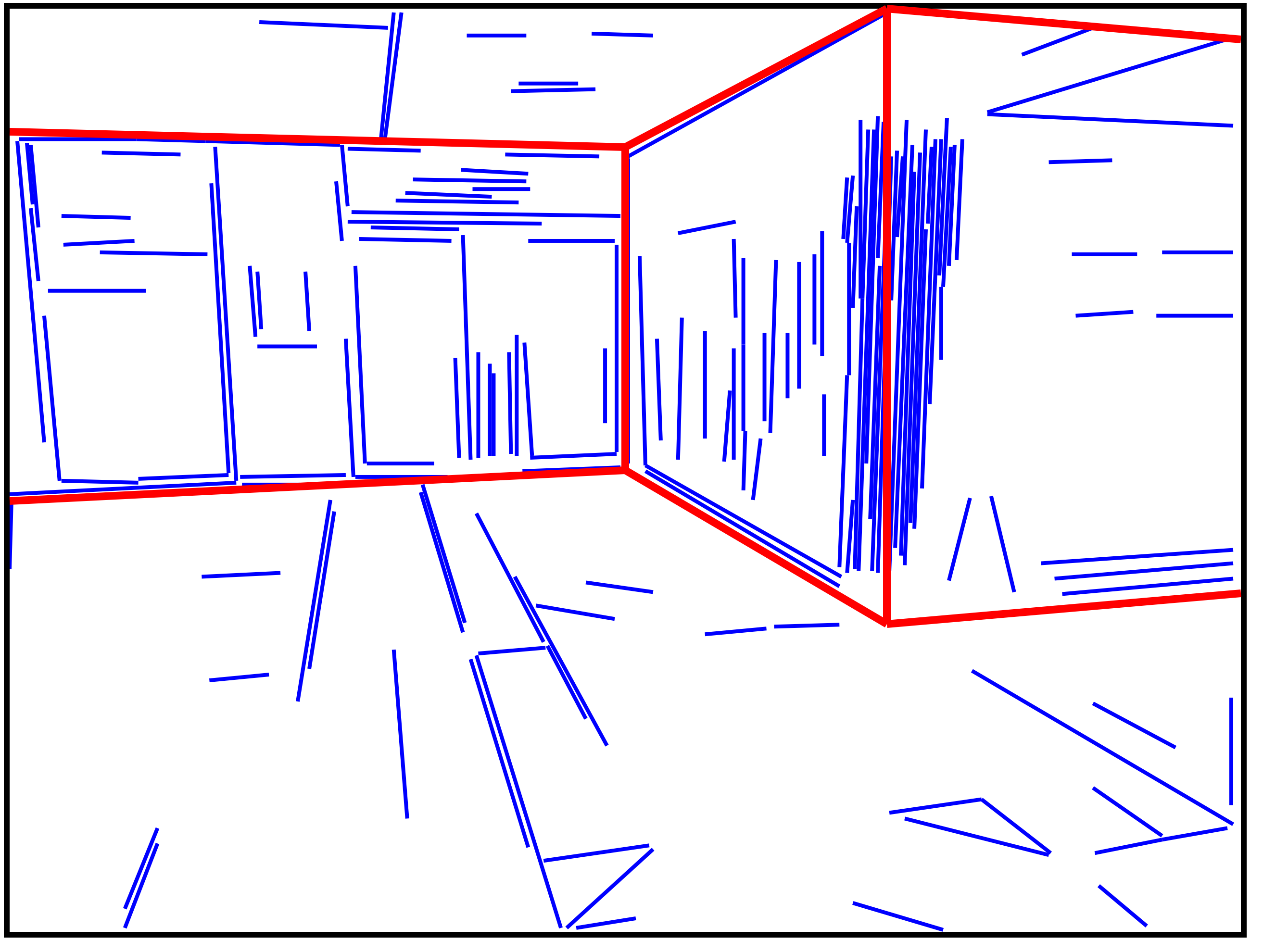
Figure 10. Object popout. We show five out of the fifteen most “solid” regions in the Geometric Context dataset. Our algorithm often finds foreground objects, which would be helpful for unsupervised object discovery [21].

Guess structure



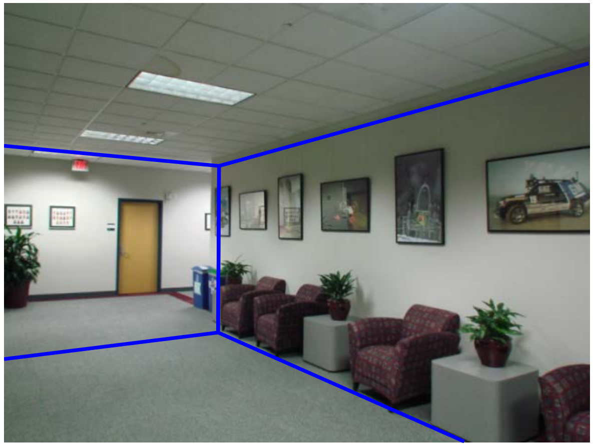
David C. Lee, Martial Hebert, Takeo Kanade, CVPR'09

Guess structure

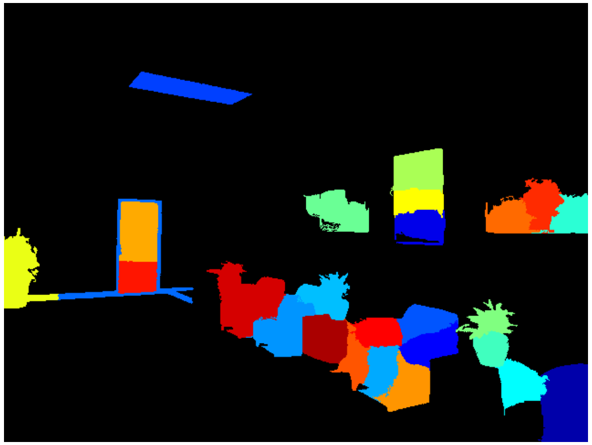


David C. Lee, Martial Hebert, Takeo Kanade, CVPR'09

Subtracting away structure



Structure



Objects

Wall appearance modeling

David C. Lee, Martial Hebert, Takeo Kanade, CVPR'09

Dealing with sparse data (rare scenes)

Long-term Fixes:

segment into chunks

- e.g. segmentation for recognition approaches

Attributes – densifying the labels

From categorization to association

- Ask not “what is this?”, ask “what is this <u>like</u>?”

Conclusion…

“If you torture data long enough,
it might confess”

- Ronald Coase