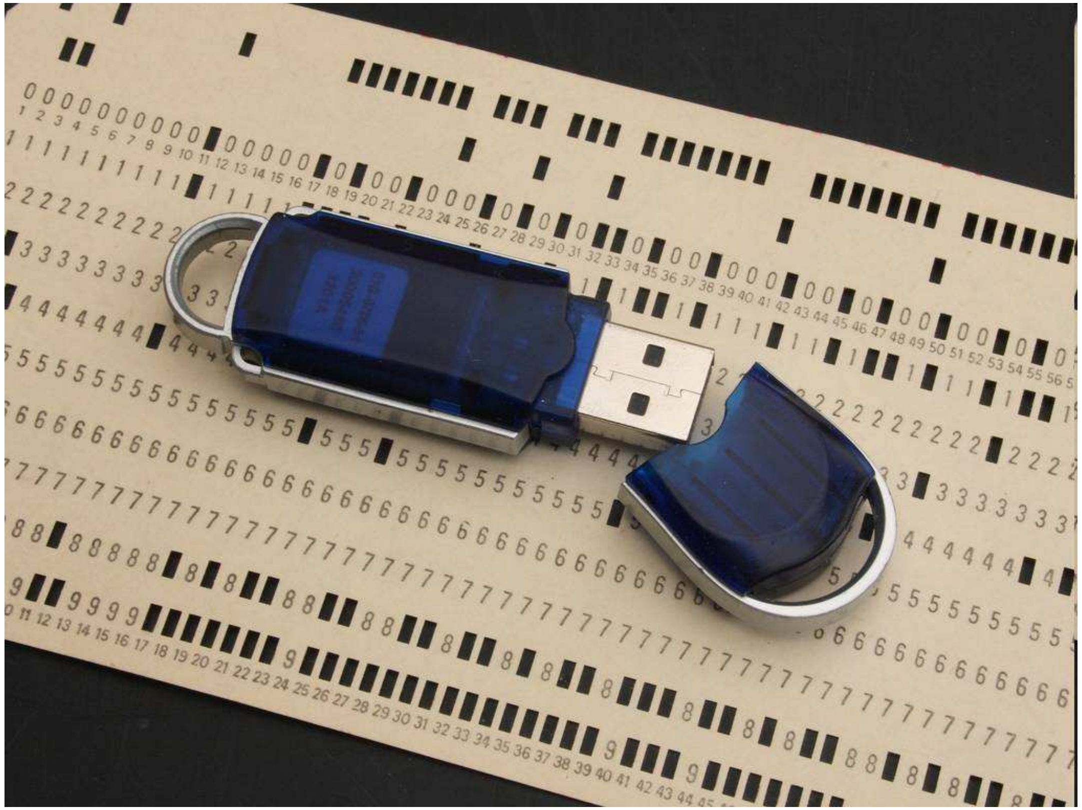
The Promise and Peril of
Big Data



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

Outline

- Why we need lots of data?
- The Promise of Big Data
- Perils of Big Data
  - Bias
  - Long tails -- we will never have enough data. “Unfamiliar is common”
  - Categorization in the modern world: "Everything is Miscelanious“

Texture

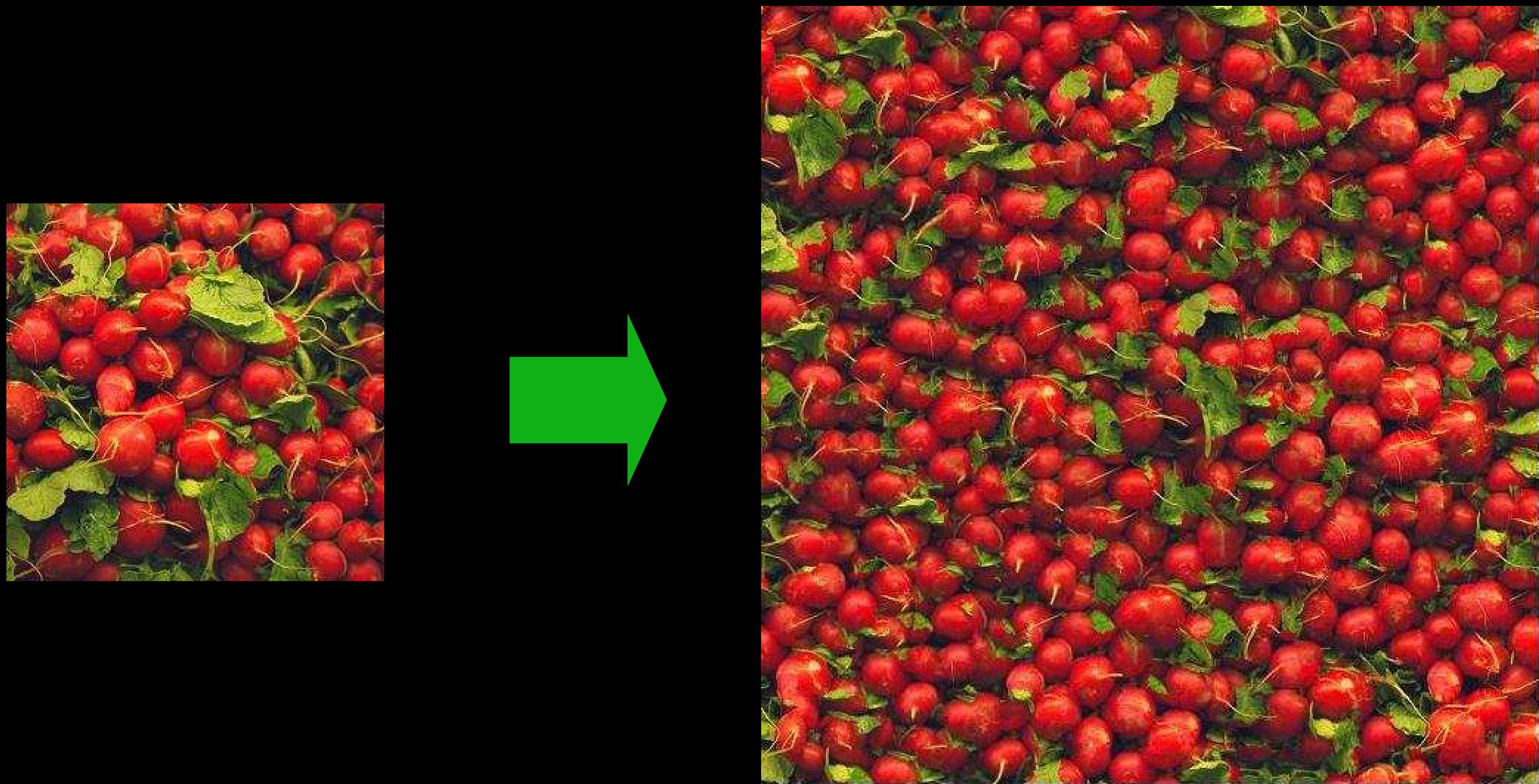
- Texture depicts spatially repeating patterns
- Many natural phenomena are textures

radishes

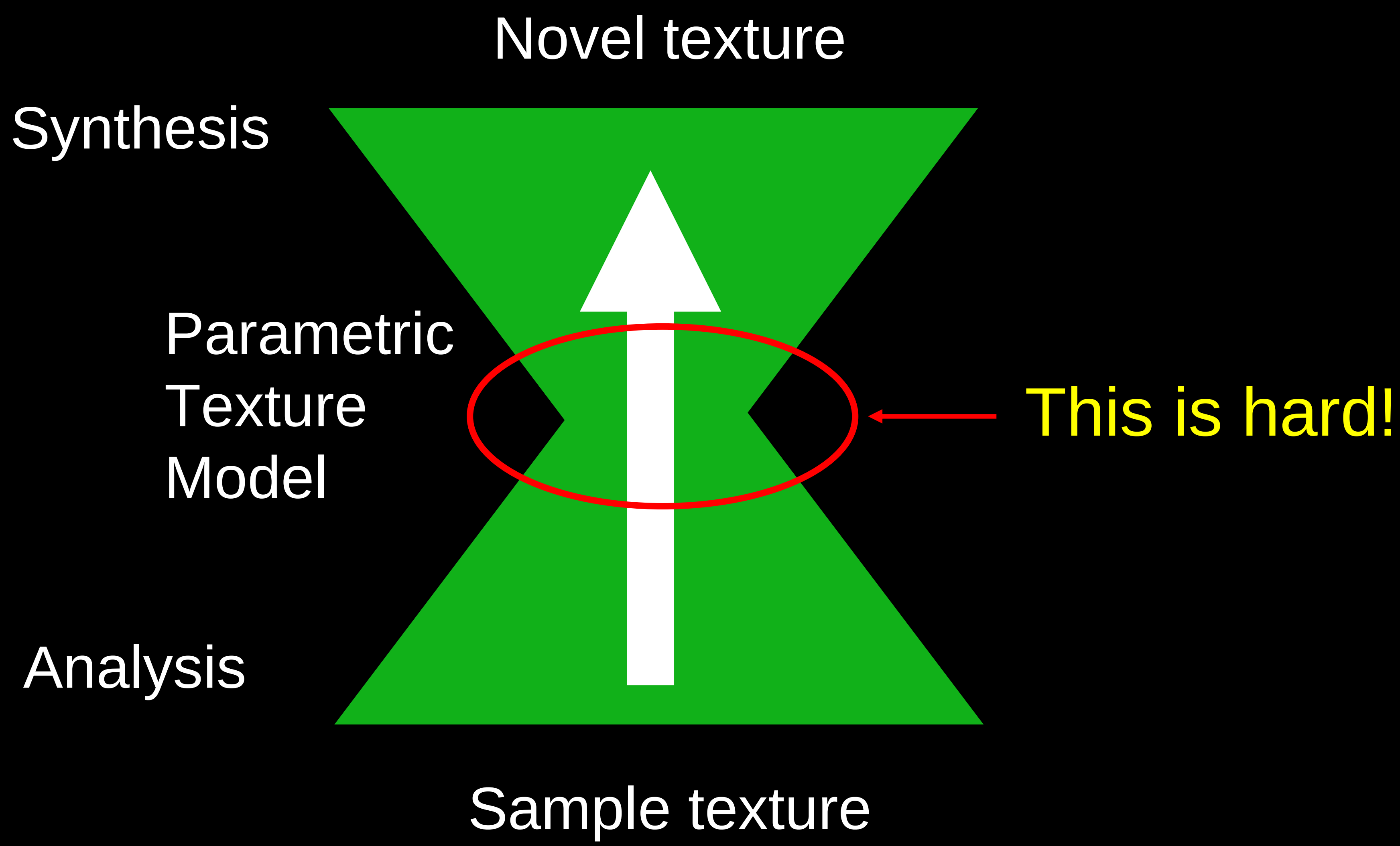
rocks

yogurt

Texture Synthesis



Classical Texture Synthesis



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

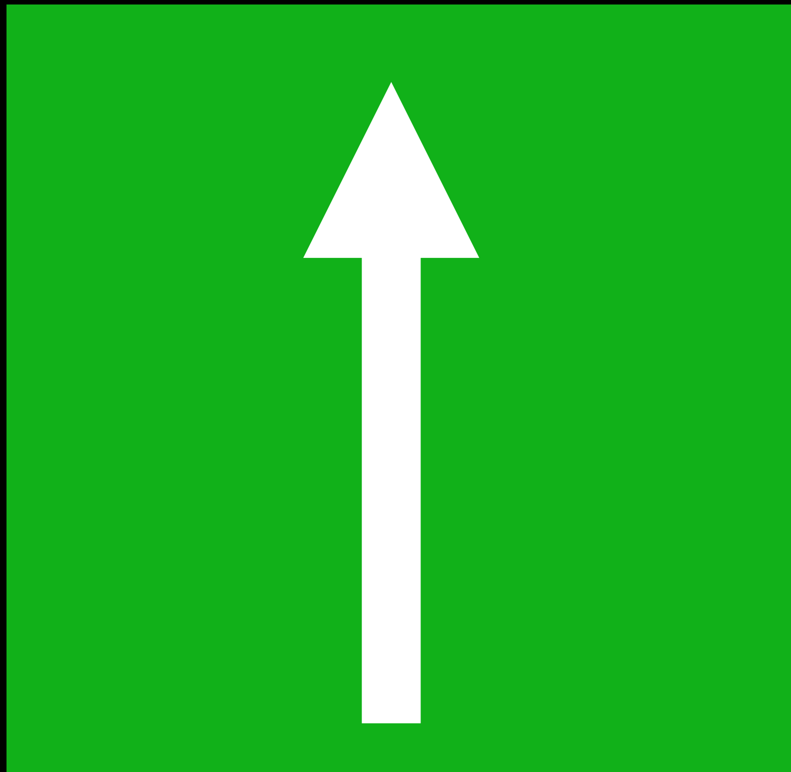
Mark V. Shaney (Bell Labs)

- 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…

Non-parametric Approach

Novel texture

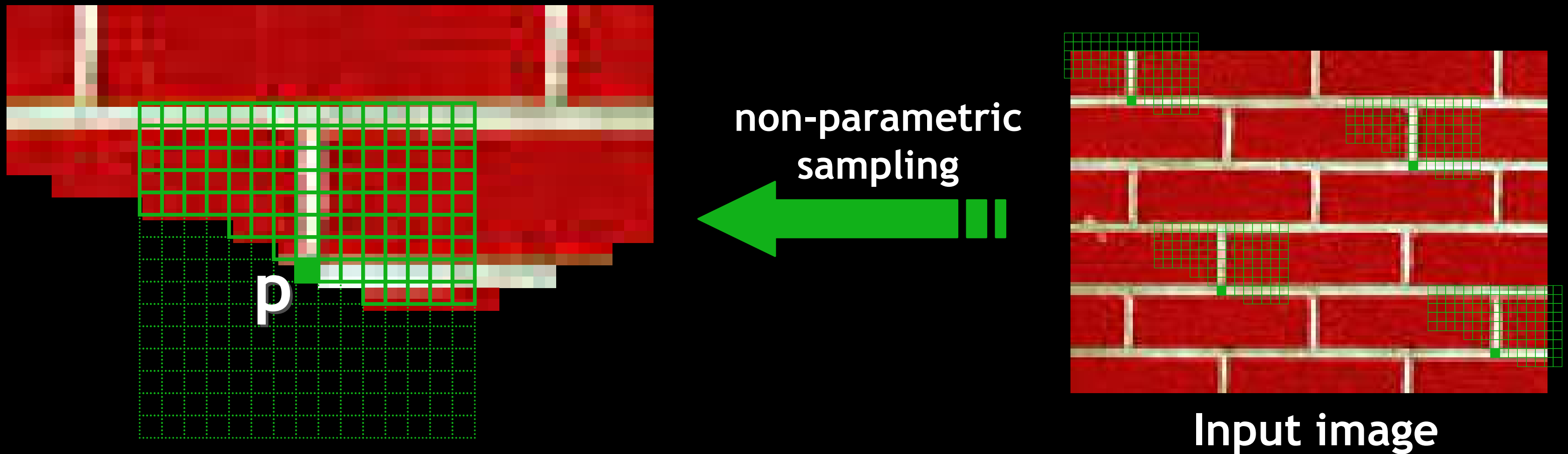
Synthesis



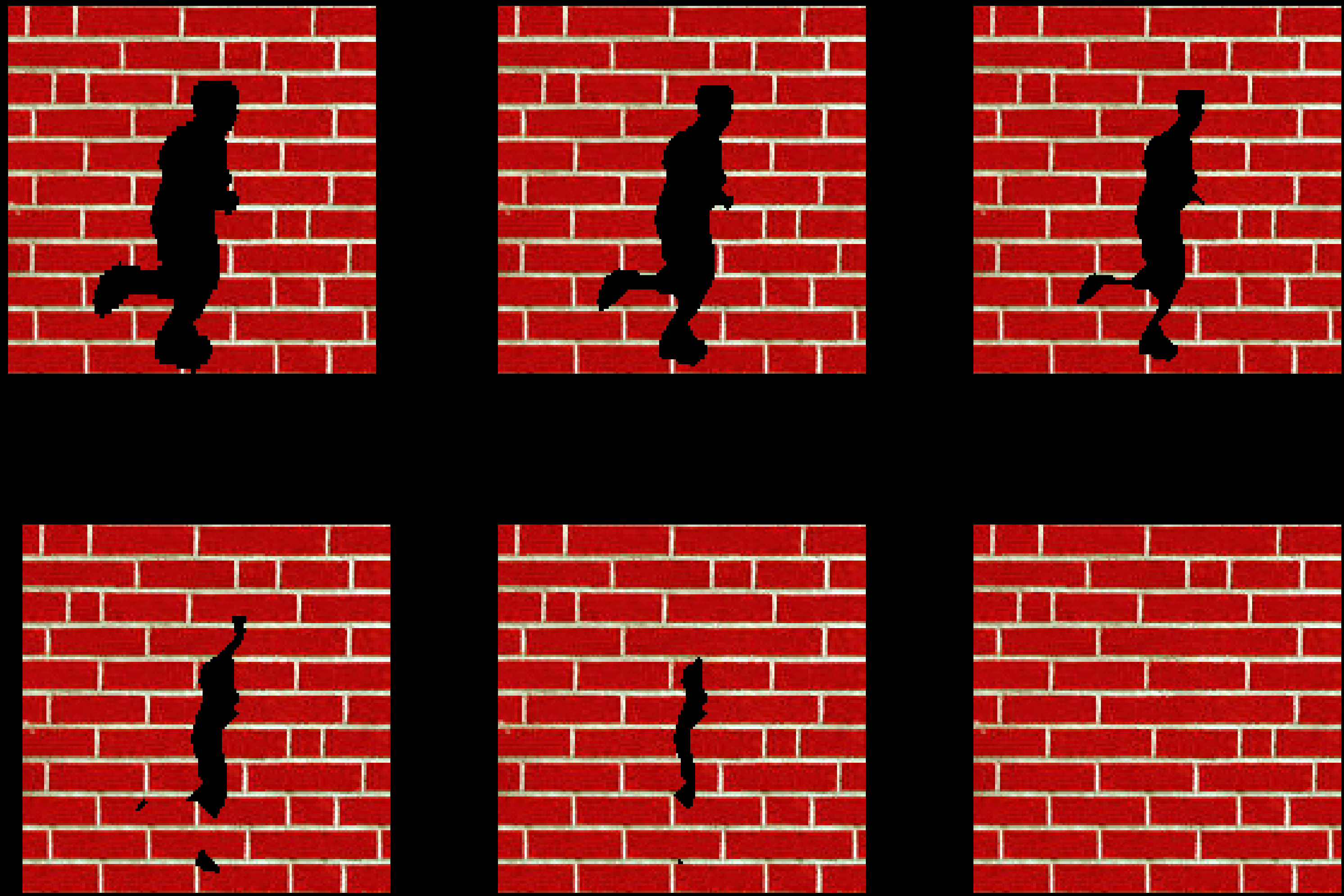
Analysis

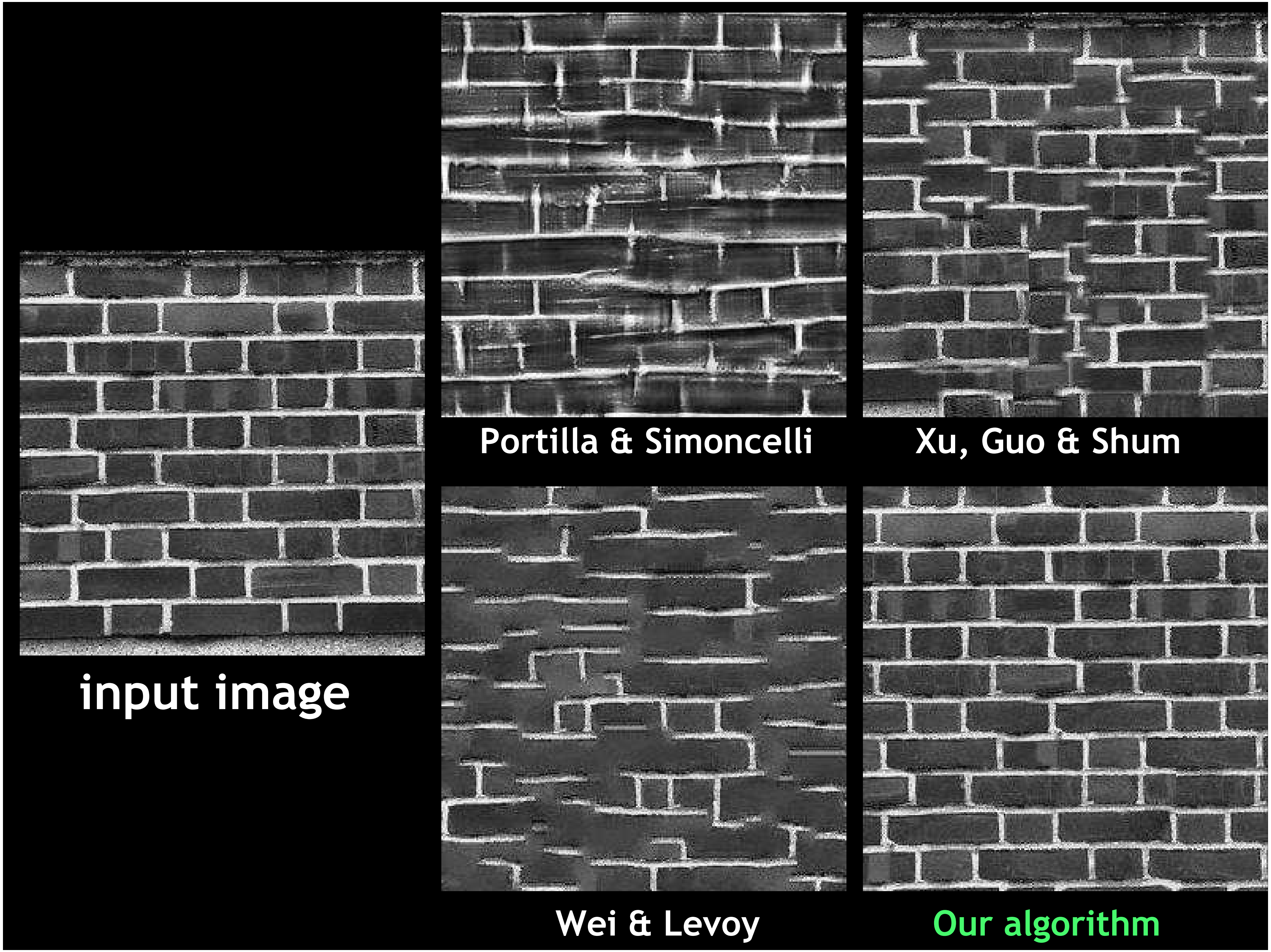
Sample texture

[Efros & Leung, '99]



Texture Growing



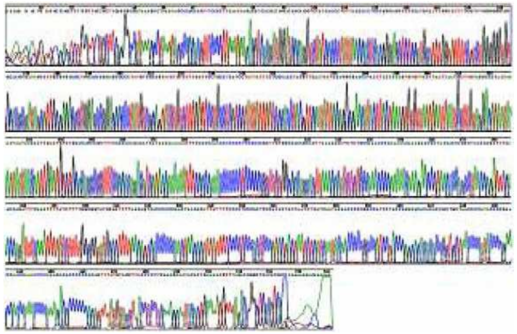


input image
Portilla & Simoncelli
Xu, Guo & Shum
Wei & Levoy
Our algorithm

Two Kinds of Things in the World

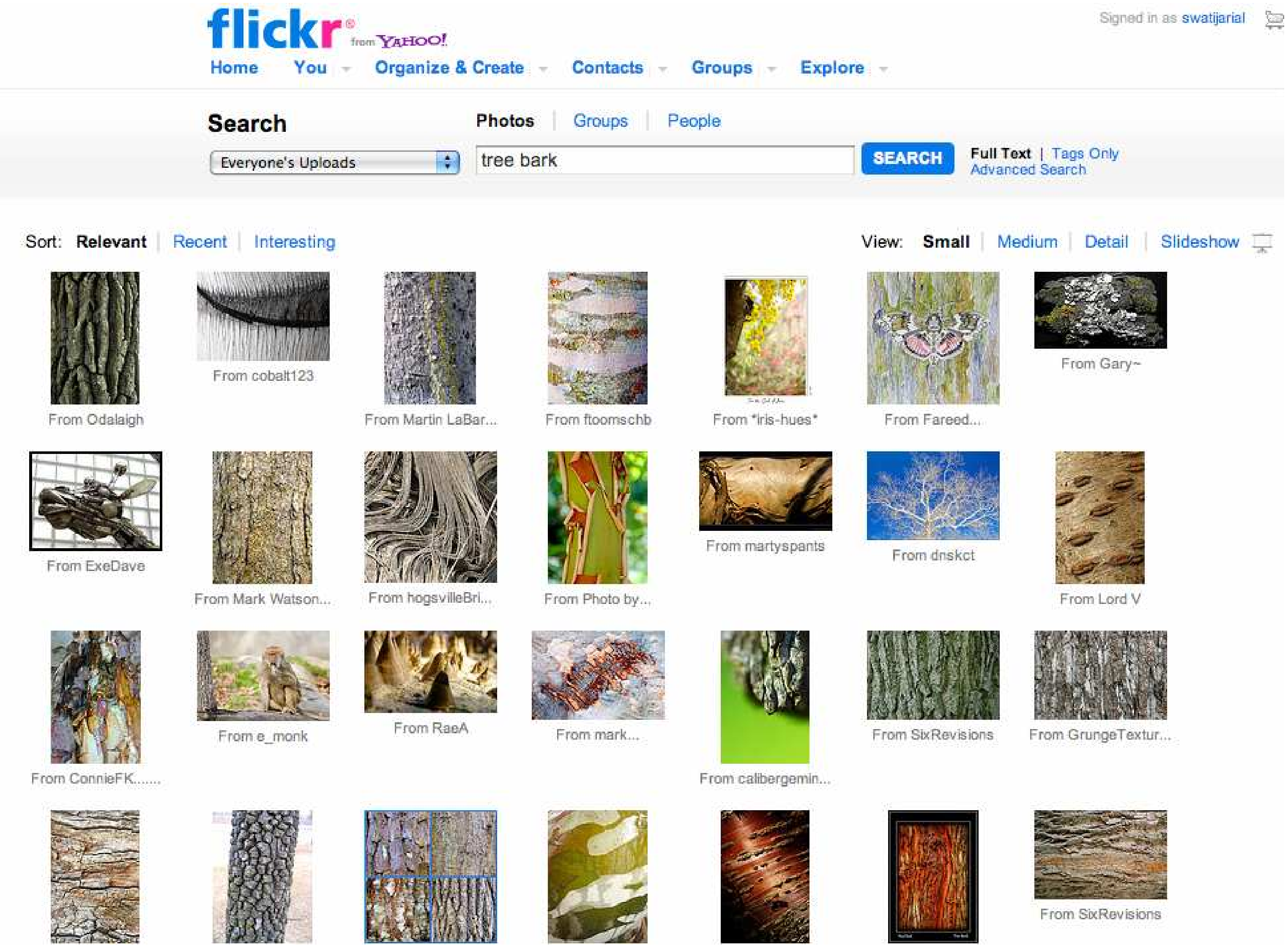
Navier-Stokes Equation

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



+ 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 <u>in the data</u>

Computer Vision

Two disciplines which happen to share the same name:

- Vision as Measurement:
  - e.g. stereo, structure-from-motion, illumination estimation
  - output: depth (meters), visual angle (radians), brightness (cd/m^2), etc.
- Vision as Understanding:
  - recognition
  - output: human concepts

"chair"

Recognition Learning Spectrum

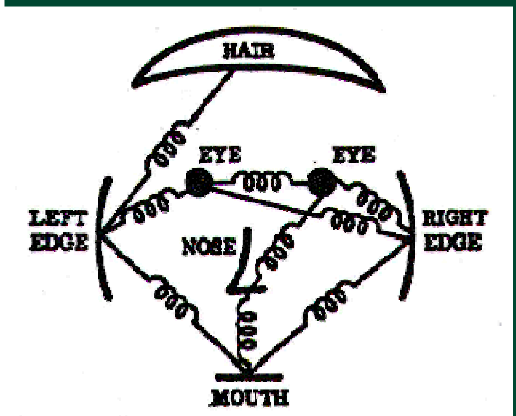
Extrapolation problem
Generalization

Interpolation problem
Correspondence

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

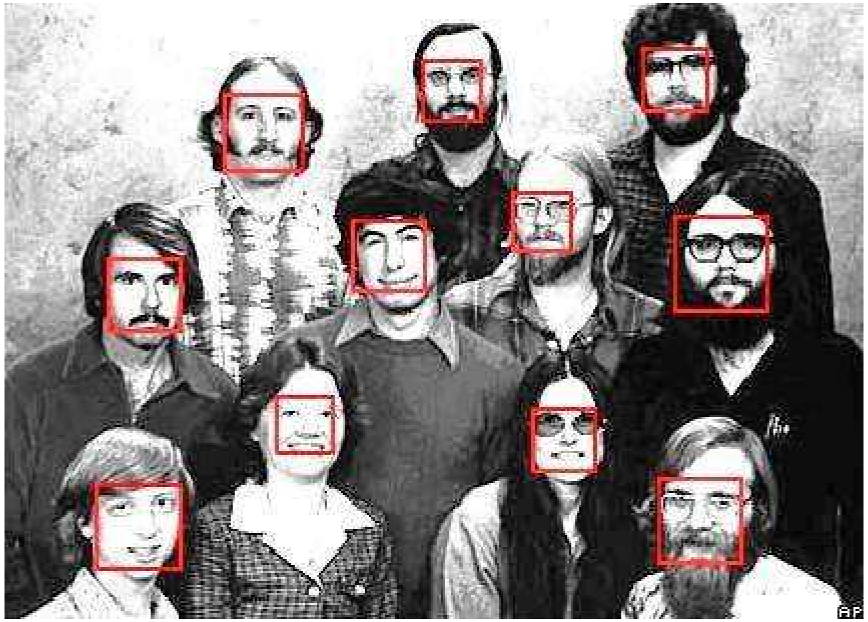
Number of training samples

Traditional datasets



Slide by Antonio Torralba

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: 2000 dim. HOG, only 300 “chair” examples

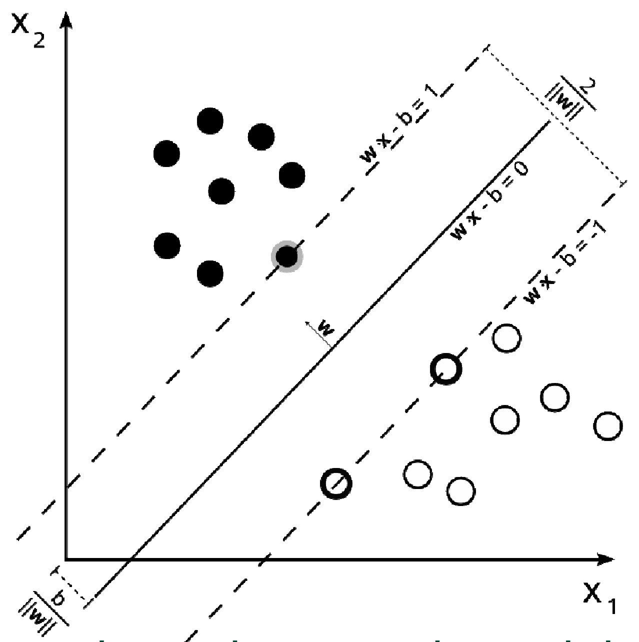


Figure that Francis Bach hates

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?!









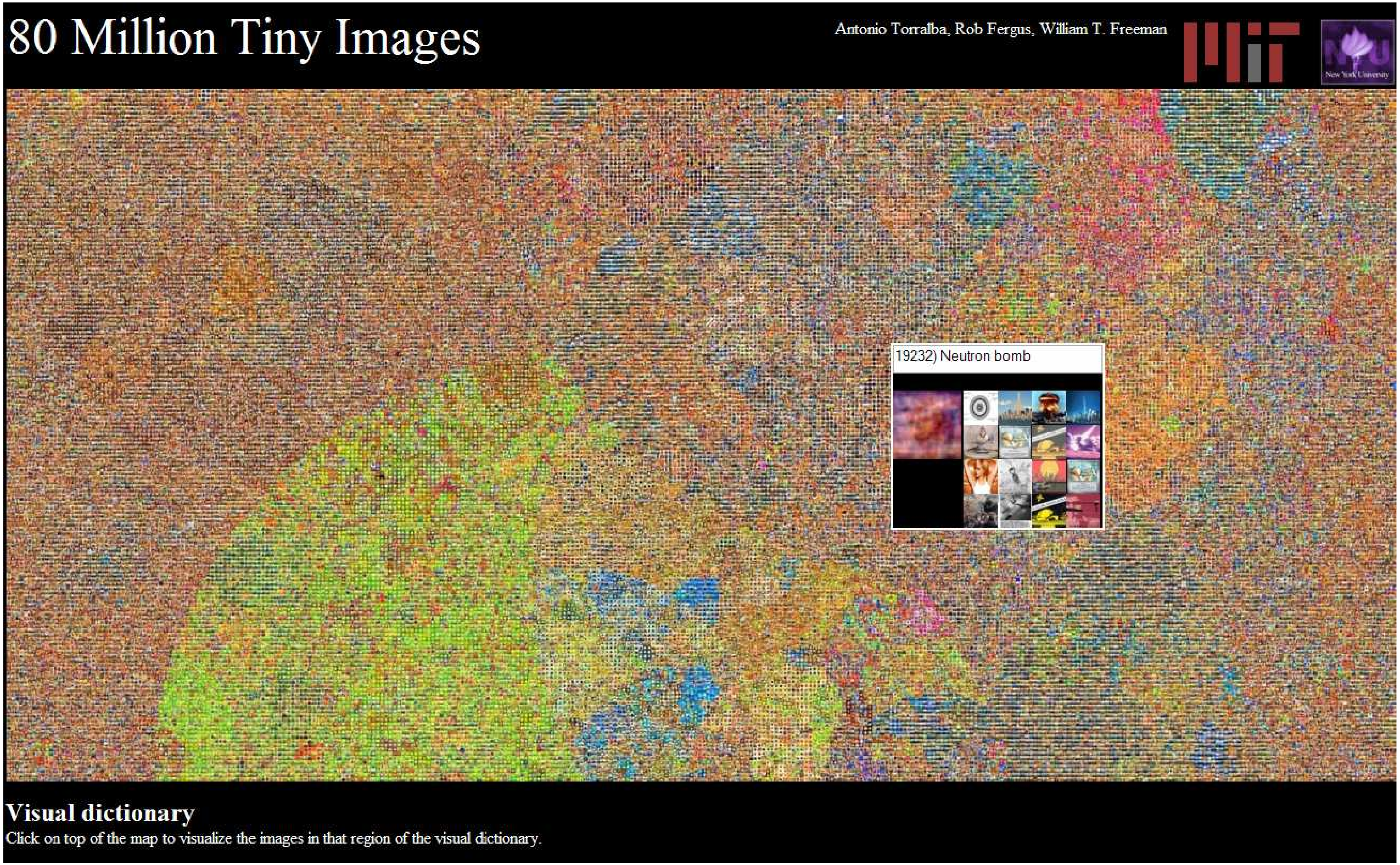





Caltech 101

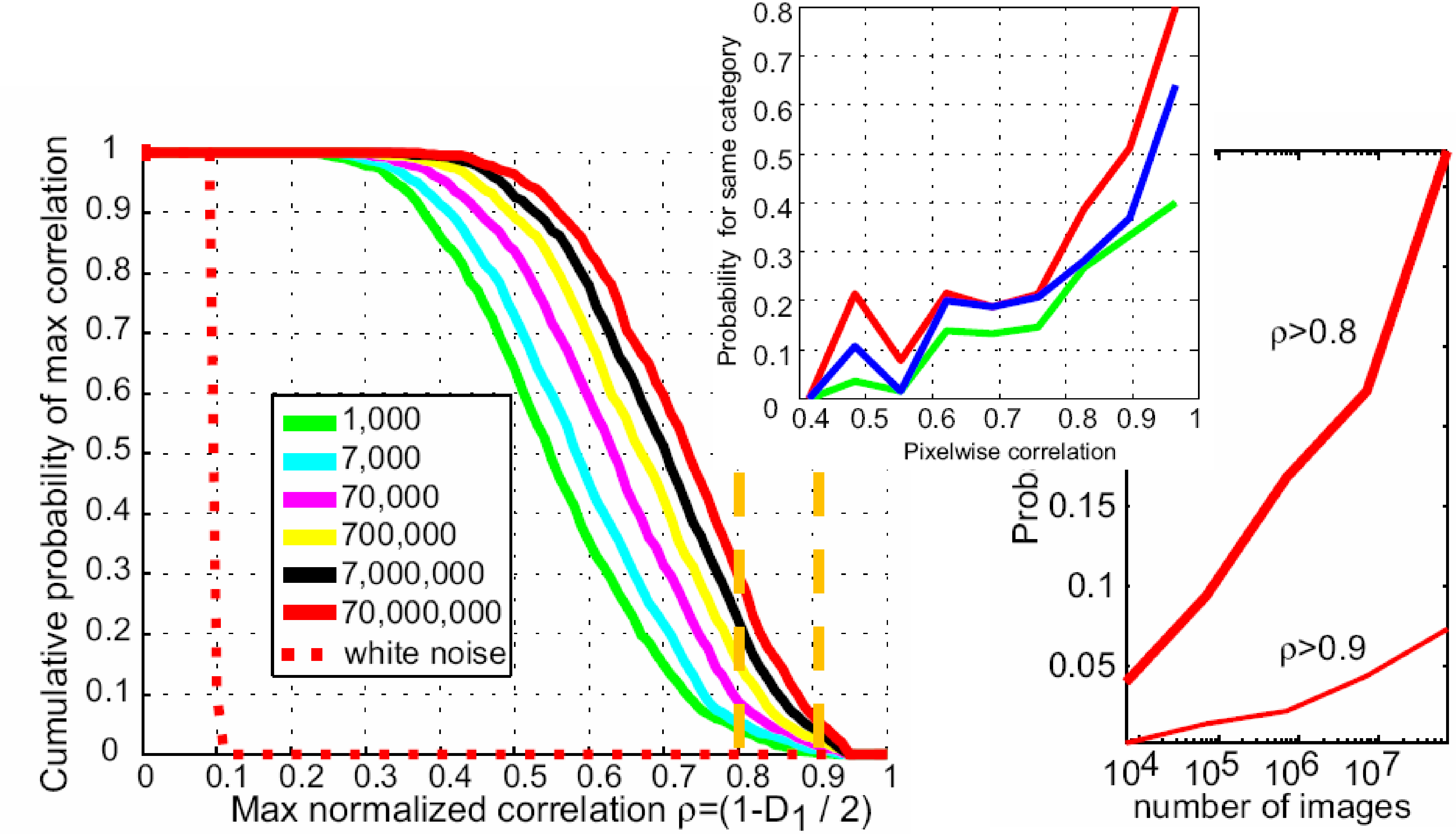
my claim:

Large-scale data is necessary, but certainly not sufficient, to solve recognition



Torralba, Fergus, Freeman. PAMI 2008

How connected is the visual space?



Torralba, Fergus, Freeman. PAMI 2008

Magic

Of

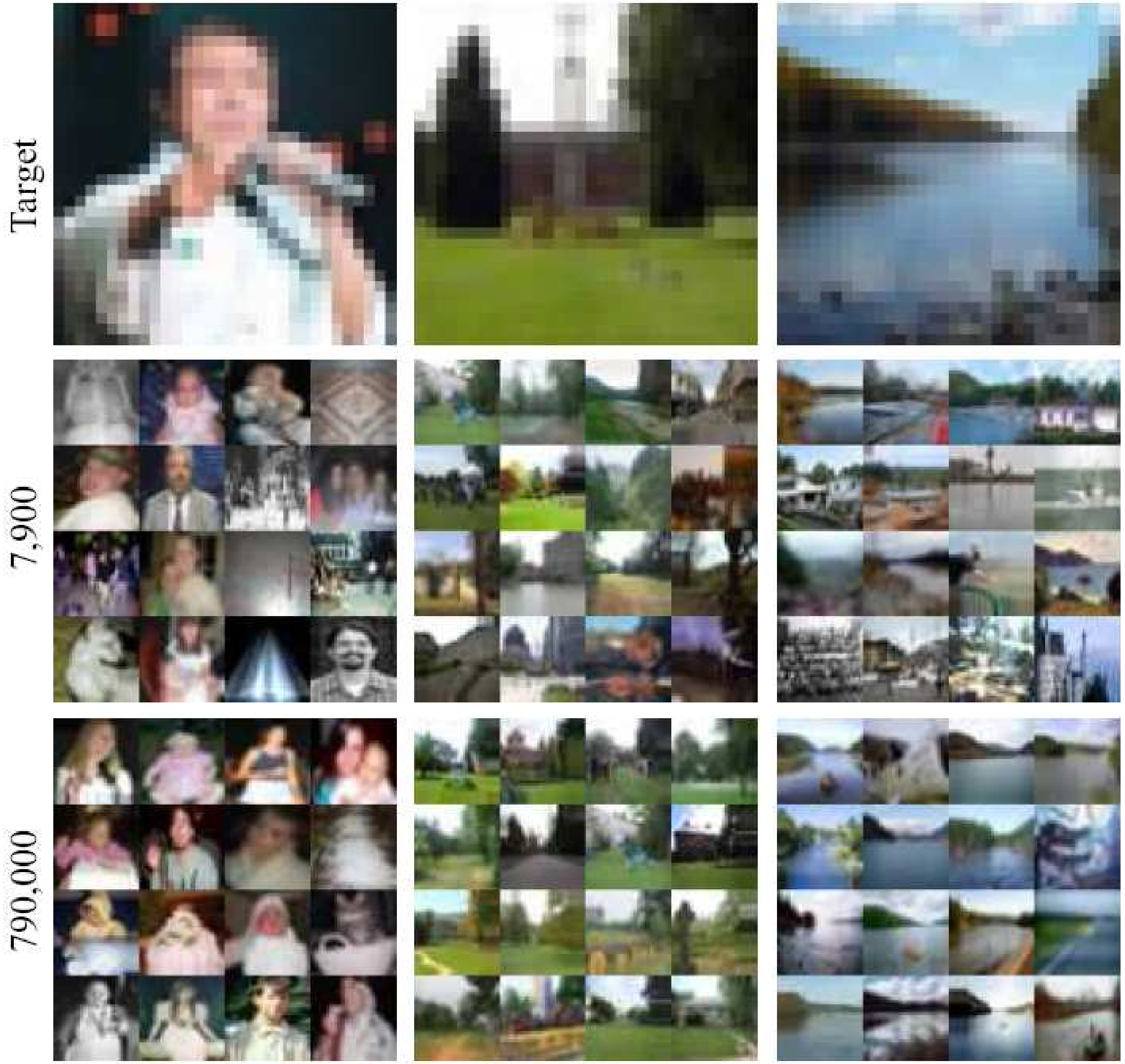
data

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

Magic

Of

data

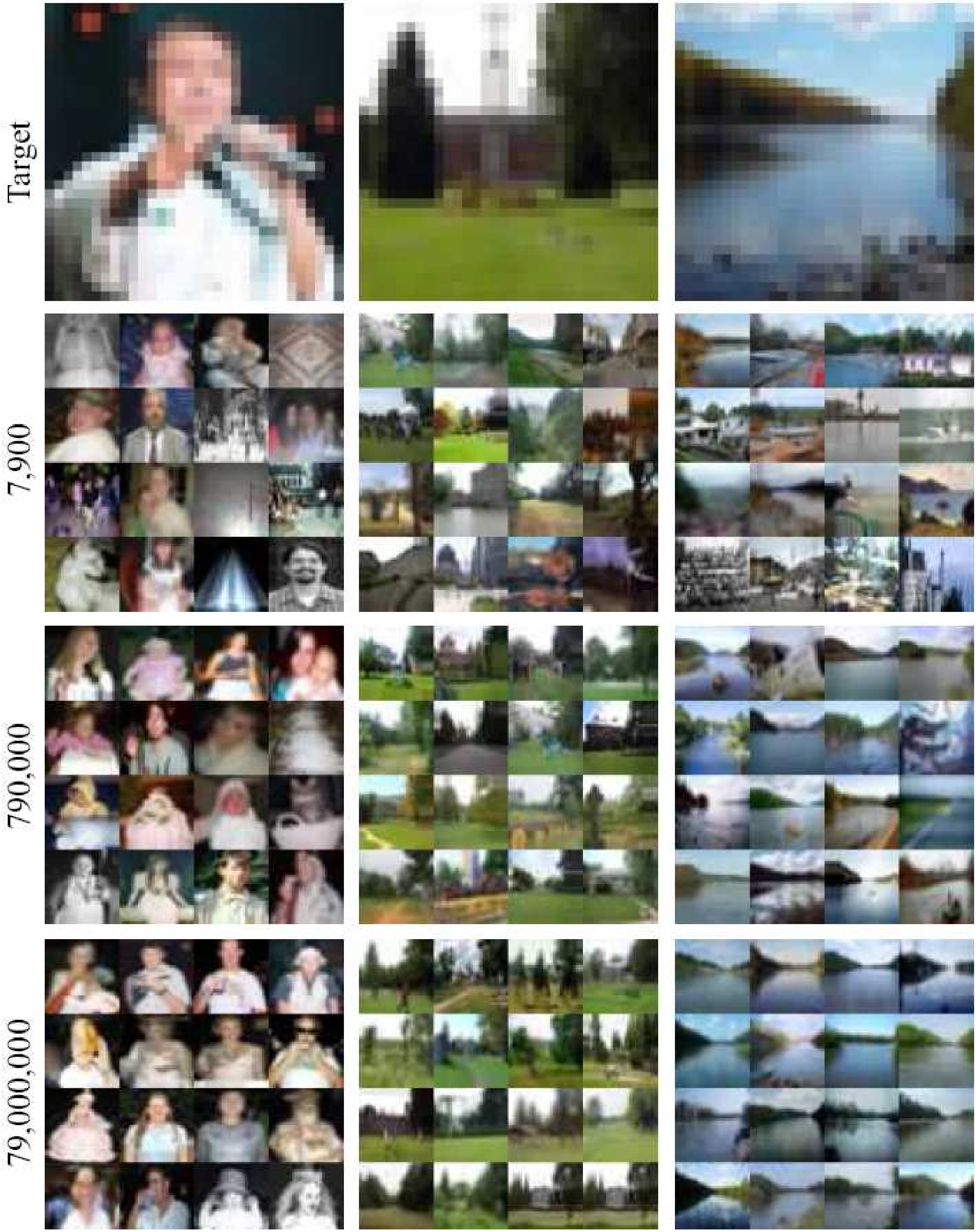


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Magic

Of

data



Is this something humans do at all?

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

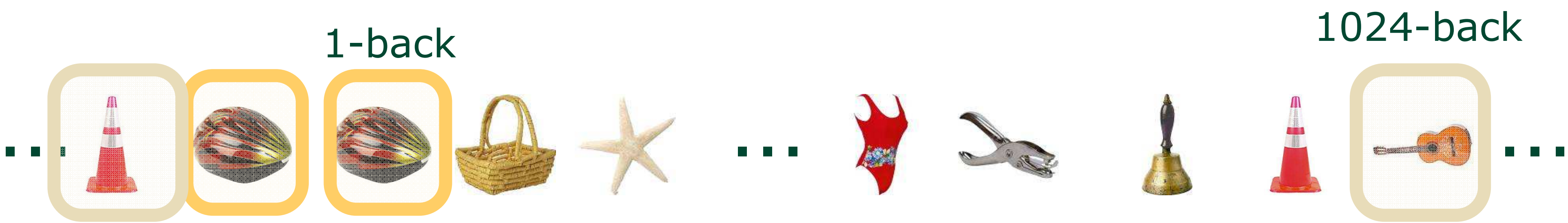
"Gist" Only

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

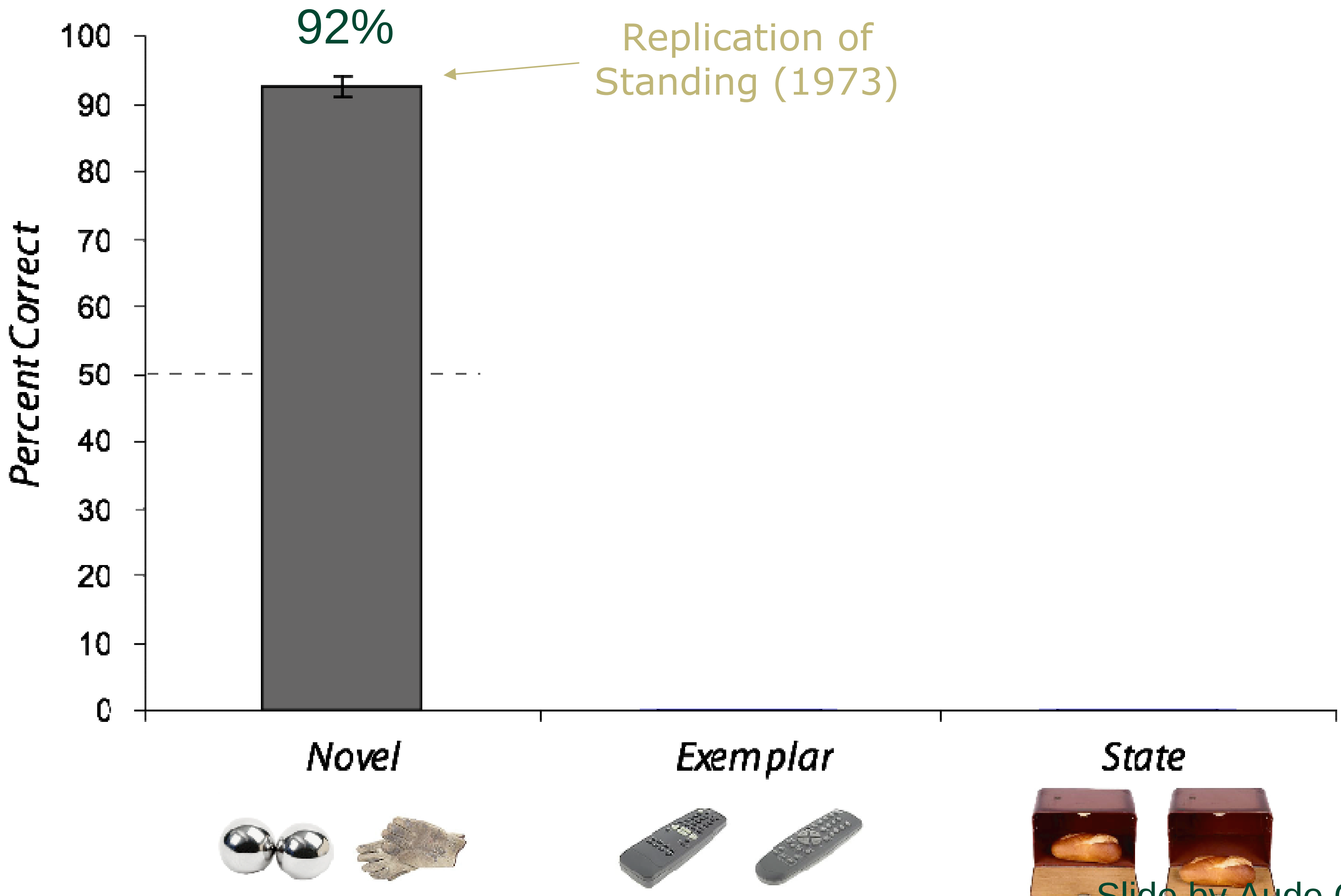
Slide by Aude Oliva

how far can we push the fidelity of visual LTM representation ?

Same object, different states

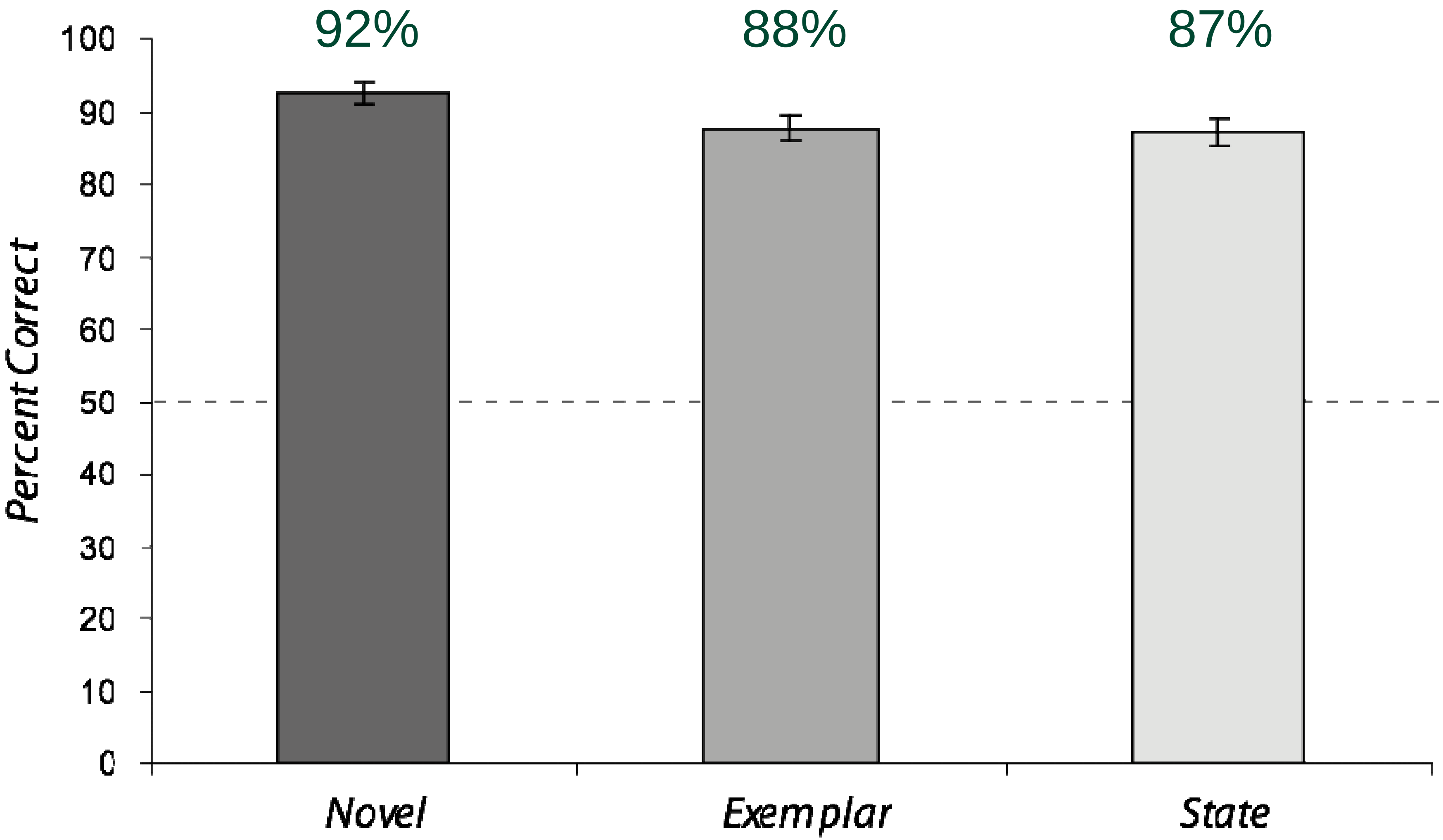
Slide by Aude Oliva

Massive Memory I: Recognition Memory Results



Slide by Aude Oliva

Massive Memory I: Recognition Memory Results



Slide by Aude Oliva

The Good News

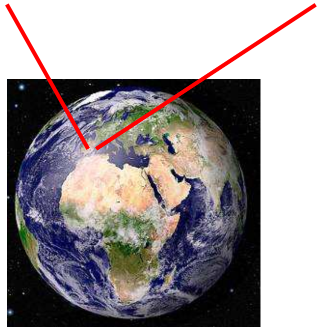
Really stupid algorithms + Lots of Data

= “Unreasonable Effectiveness”

Raw (unabelled) Data

Useful since visual world has structure

actual images << # possible images



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

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

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

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

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

Automatic Colorization Result

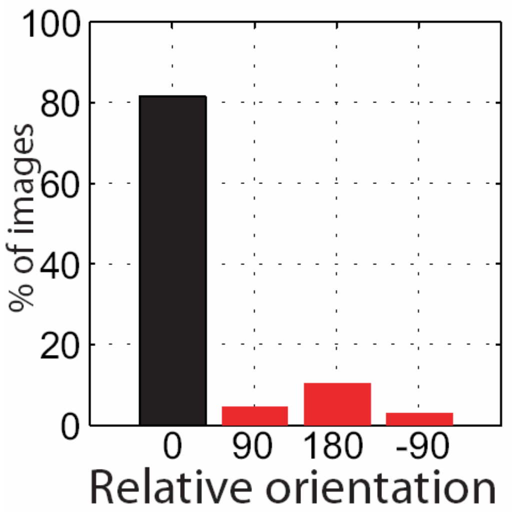
Grayscale input High resolution

Colorization of input using average

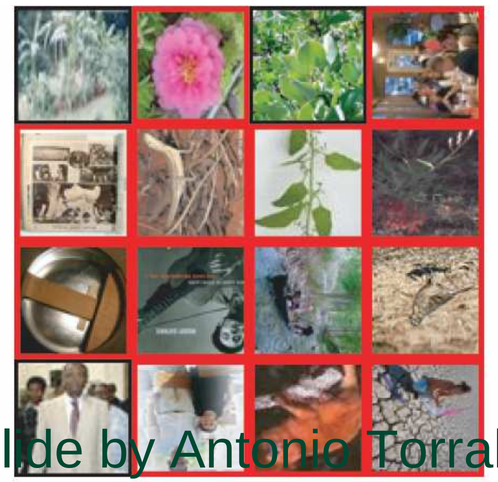
A. Torralba, R. Fergus, W.T.Freeman. 2008

Automatic Orientation

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

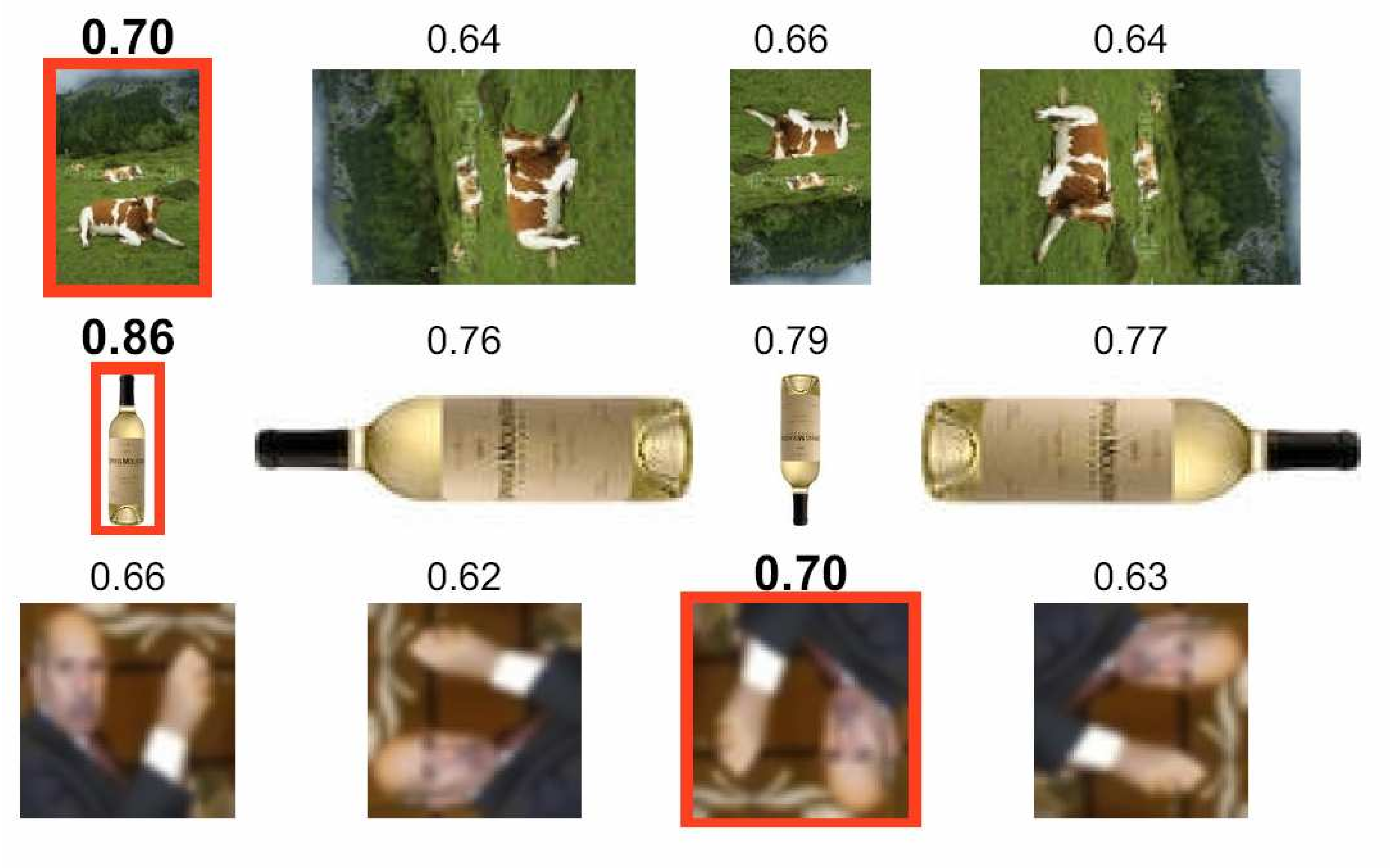


- Examples of high and low confidence images:



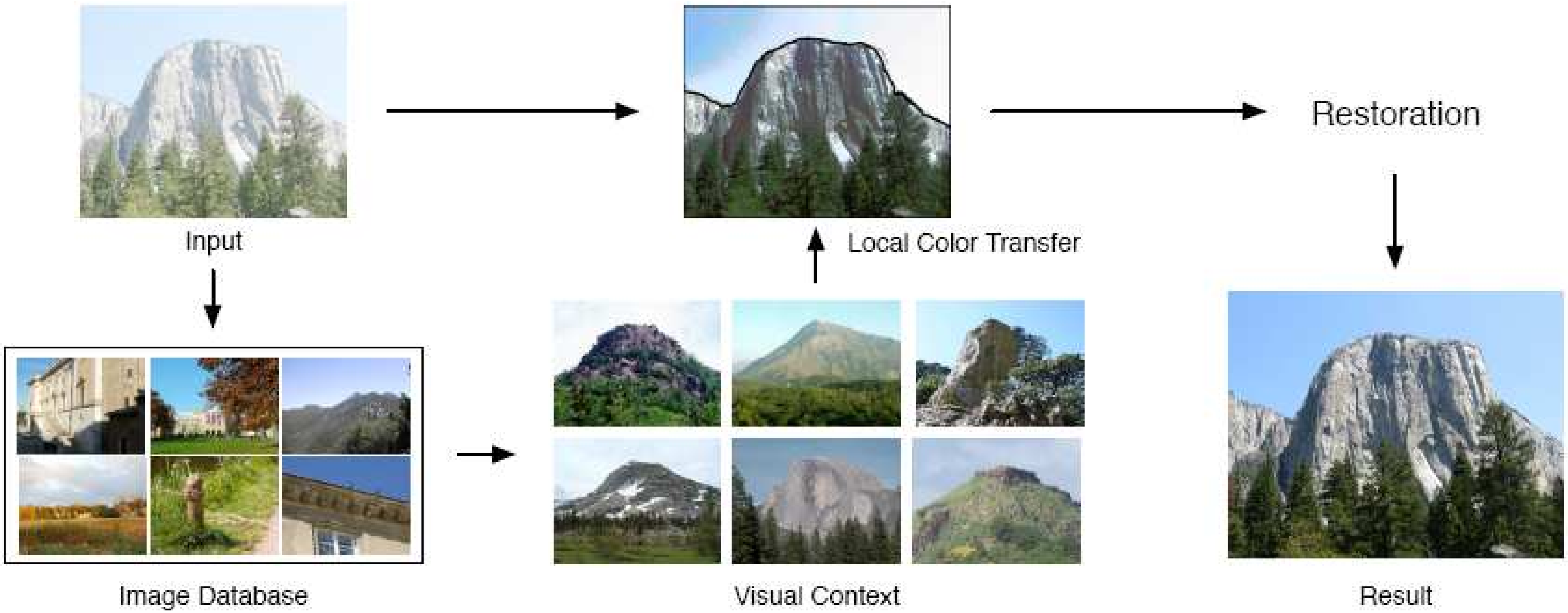
Slide by Antonio Torralba

Automatic Orientation Examples



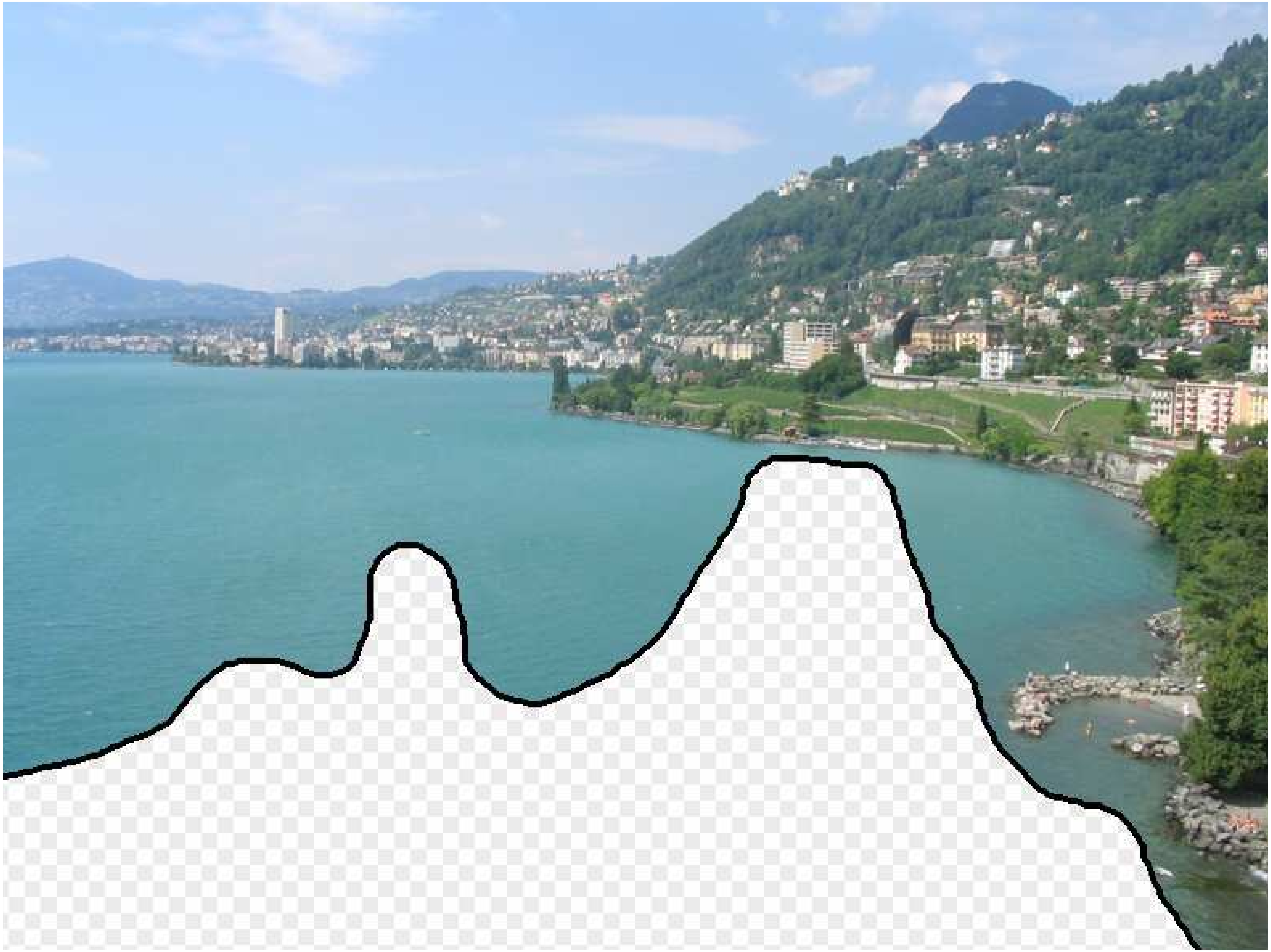
A. Torralba, R. Fergus, W.T.Freeman. 2008

Image Restoration using Online Photo Collections [ICCV'09]

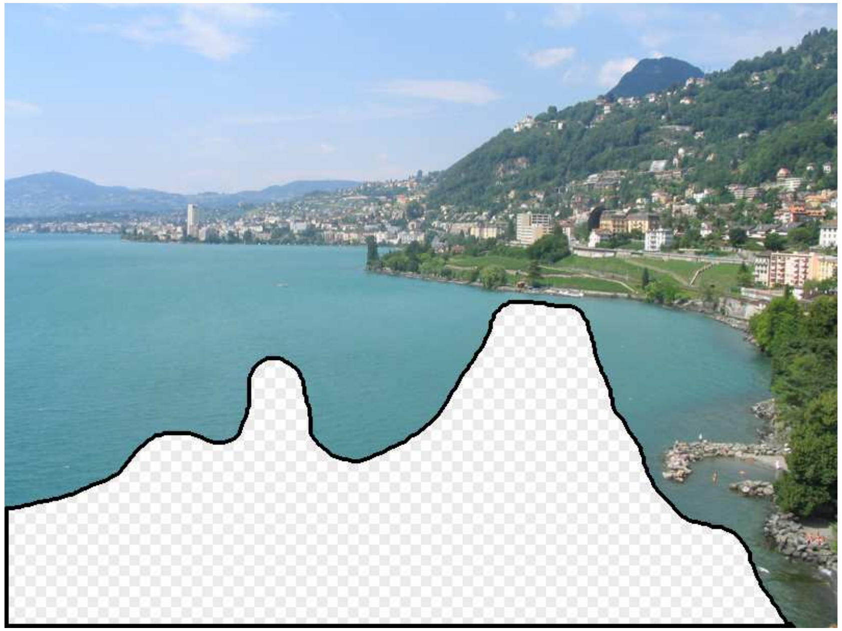


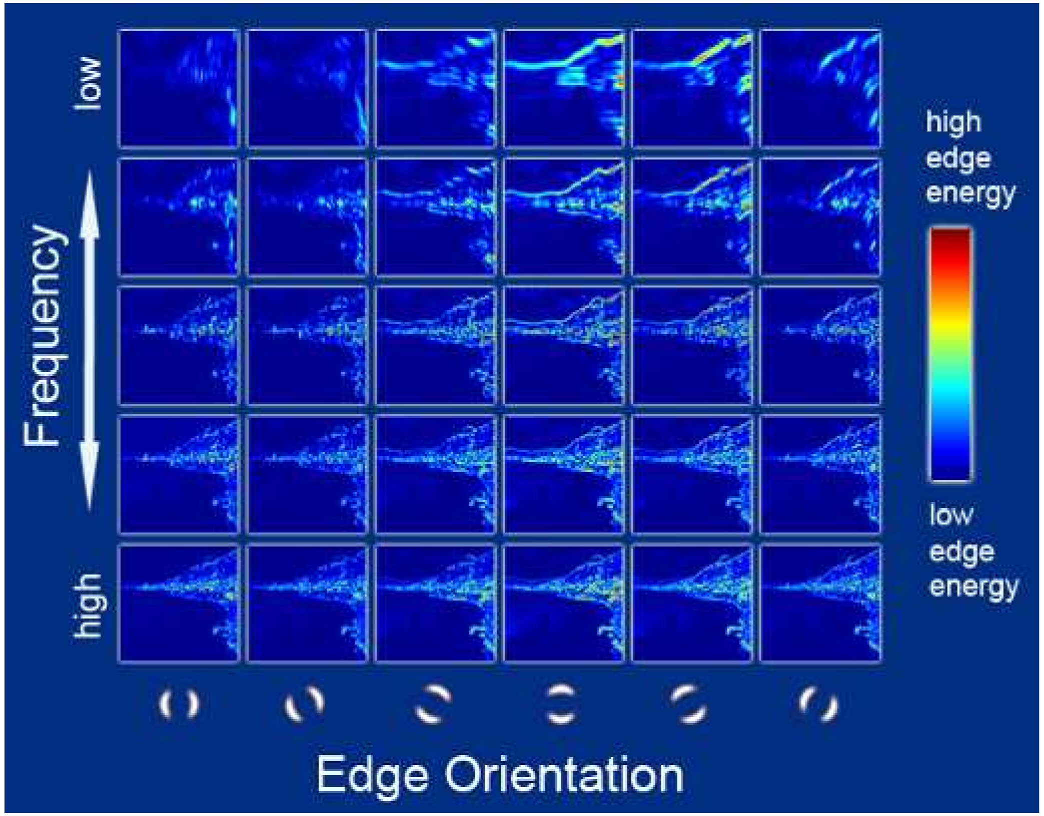
Dale, Johnson, Sunkavalli, Matusik, Pfister, ICCV'09


[Hays & Efros, SIGGRAPH'07]

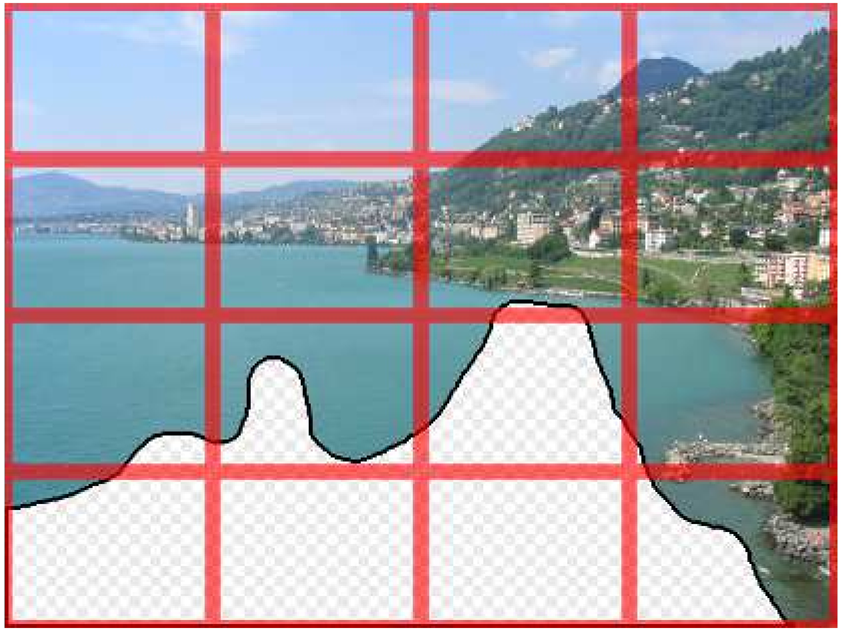


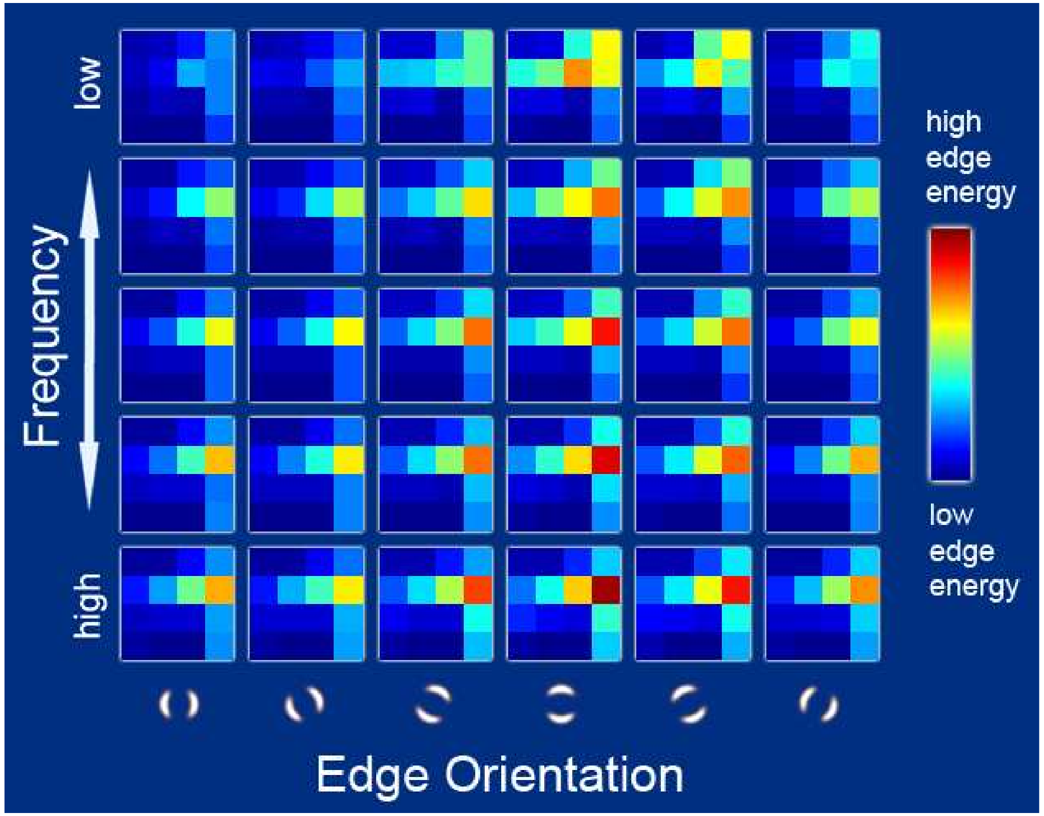
Scene Descriptor





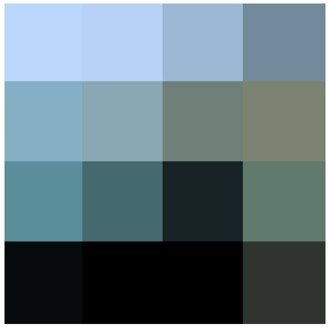
Scene Descriptor



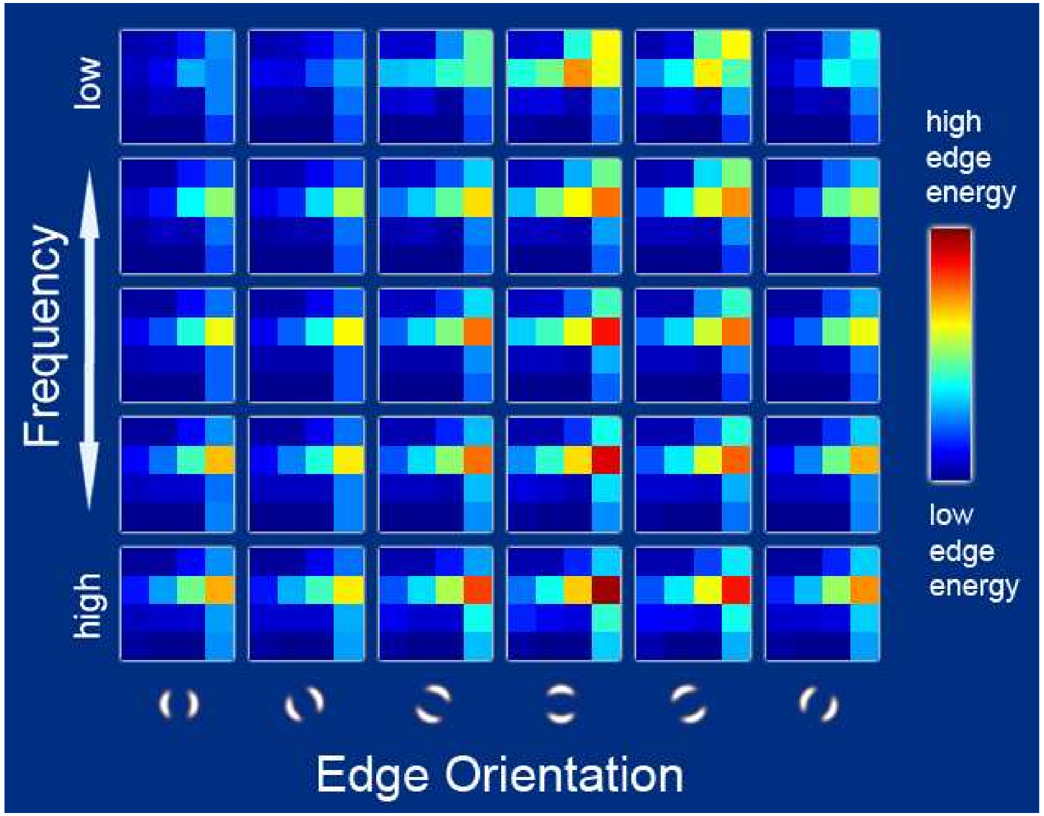


Gist scene descriptor
(Oliva and Torralba 2001)

Scene Descriptor

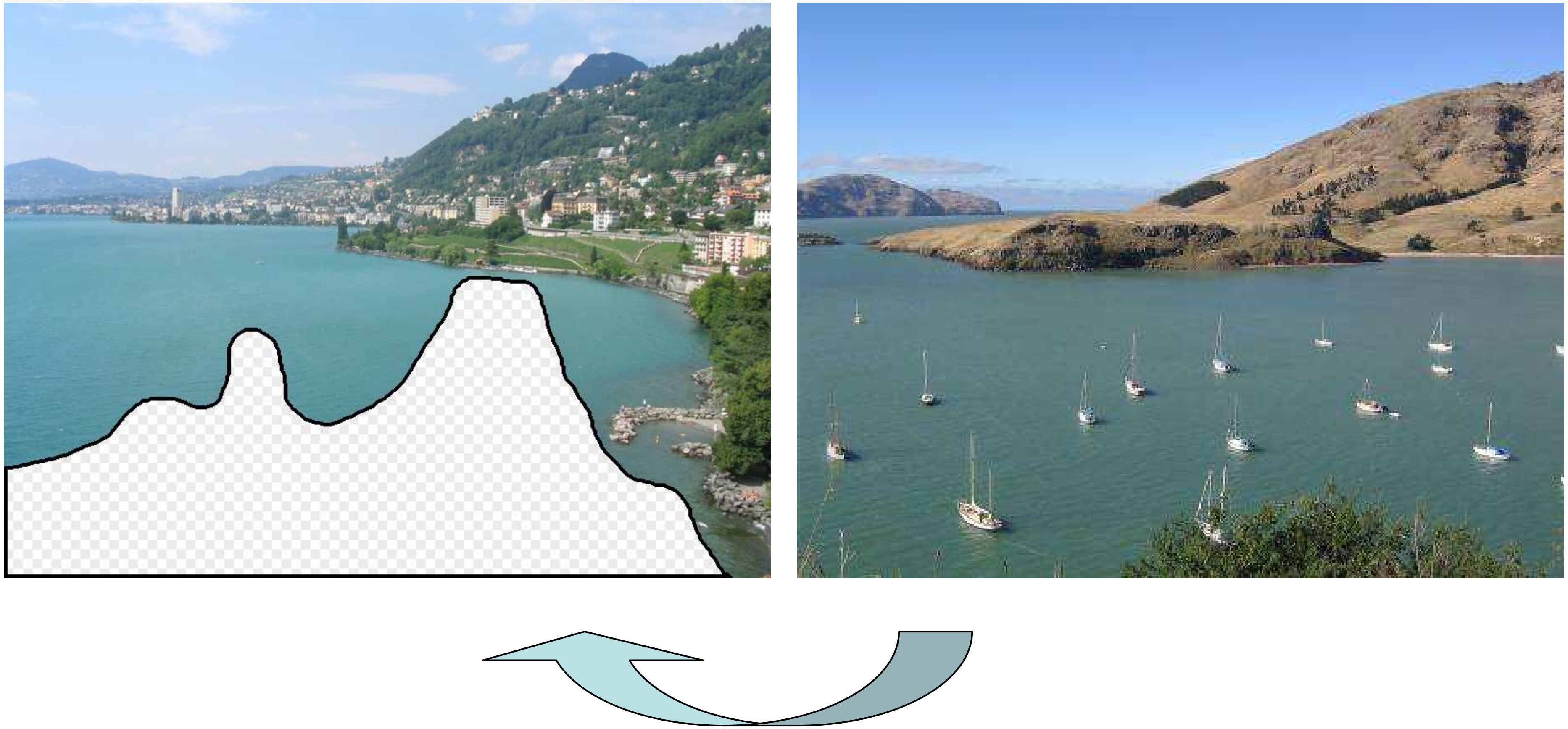


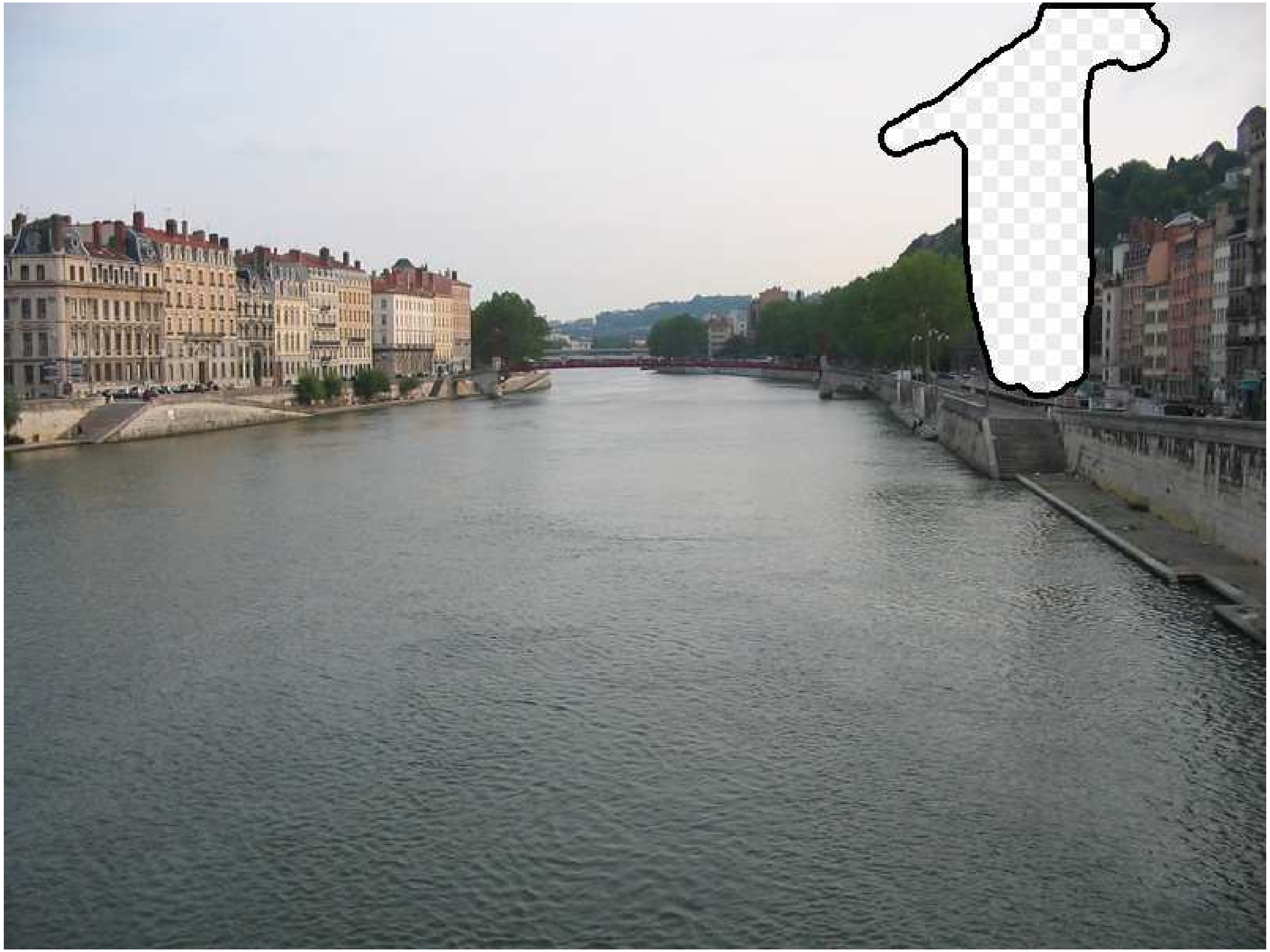
+

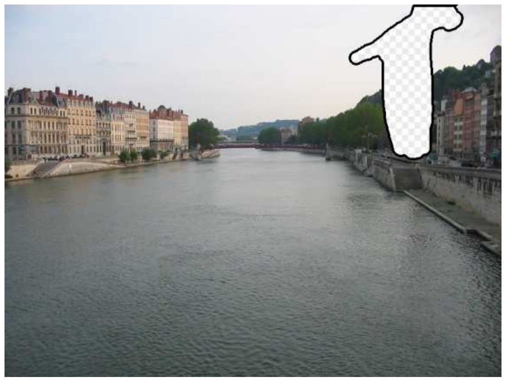


Gist scene descriptor
(Oliva and Torralba 2001)

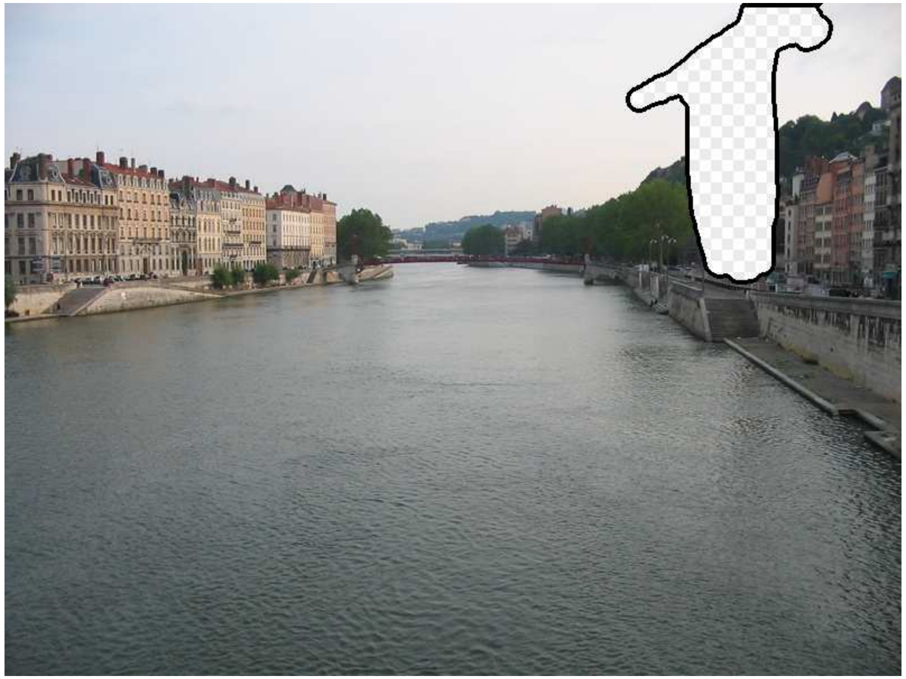
2 Million Flickr Images

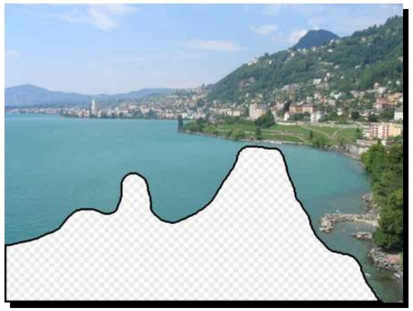


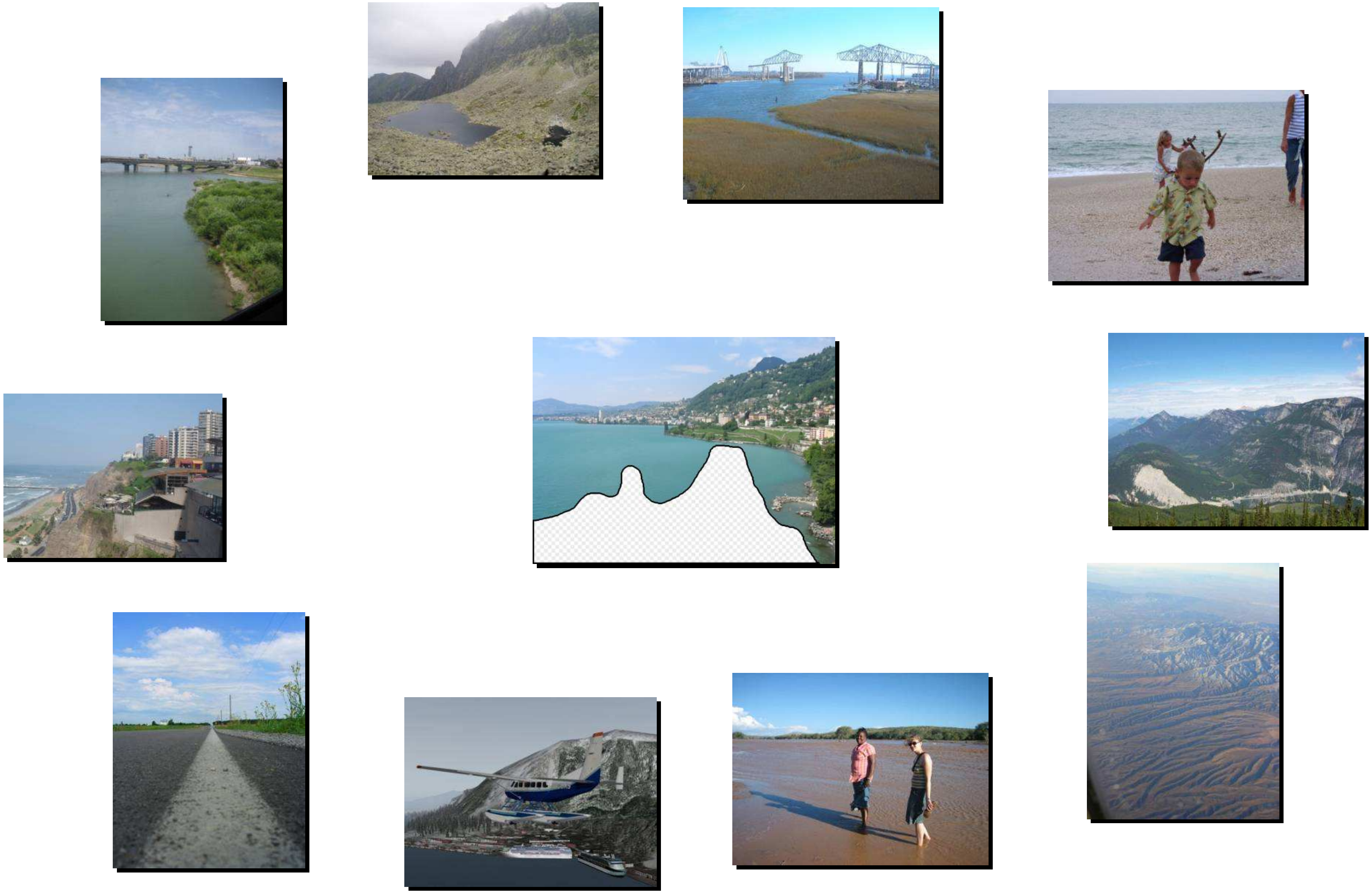




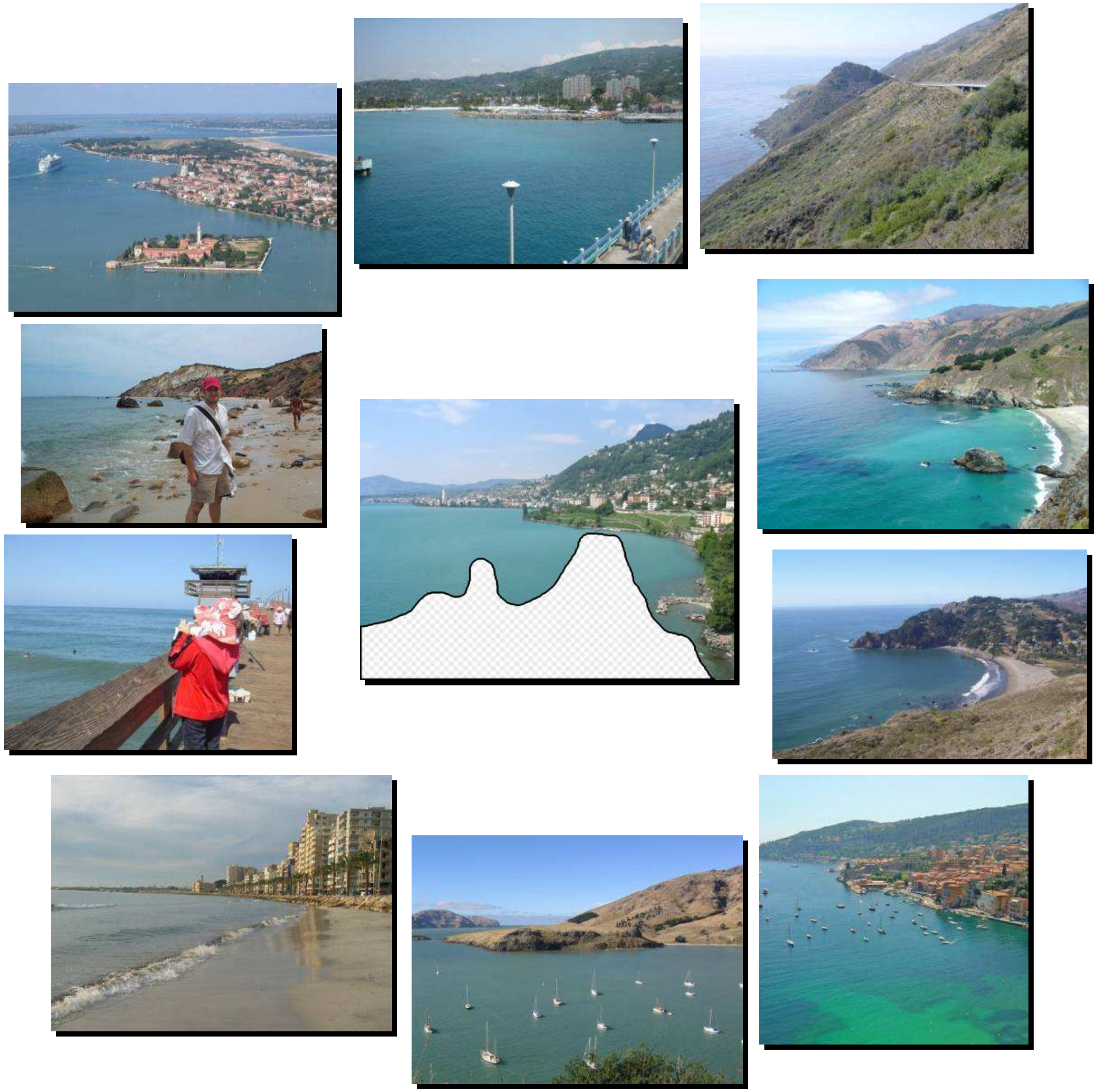
… 200 scene matches





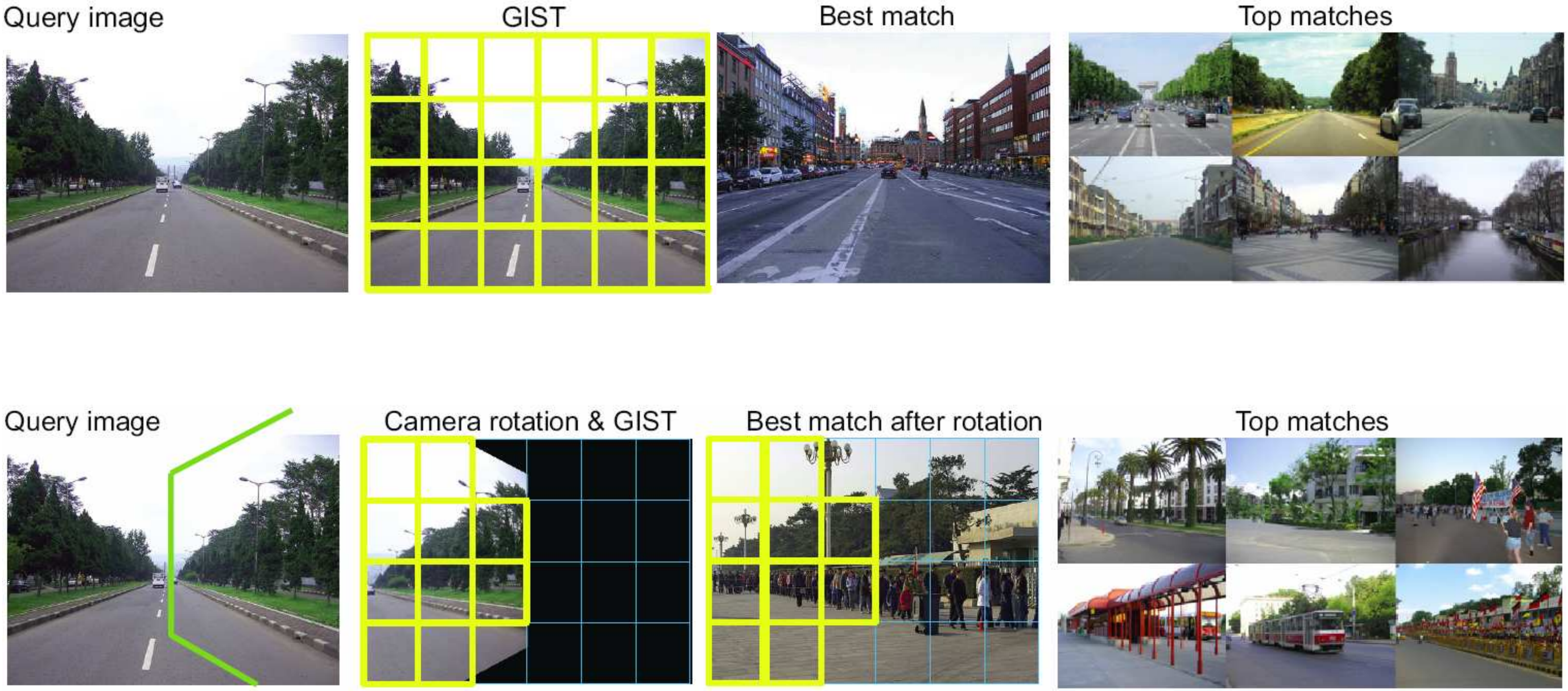


Nearest neighbors from a collection of 20 thousand images



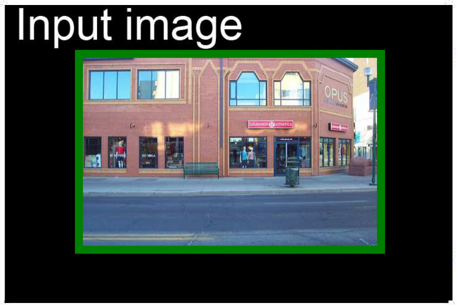
Nearest neighbors from a collection of 2 million images

Scene matching with camera transformations

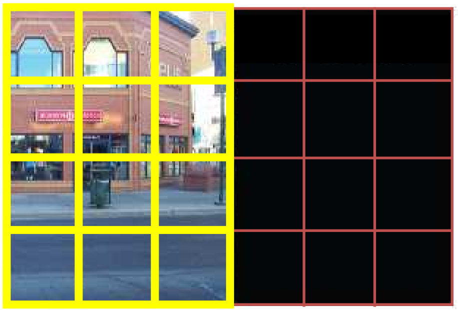


Sivic, Kaneva, Torralba, Avidan, Freeman, Internet Vision Workshop, 2008
updated version to appear in Proceedings of the IEEE (2010)

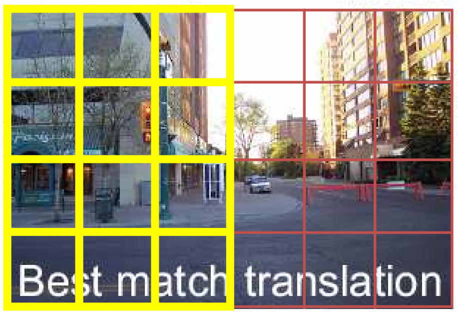
Scene matching with camera view transformations: Translation



1. Move camera



2. View from the virtual camera



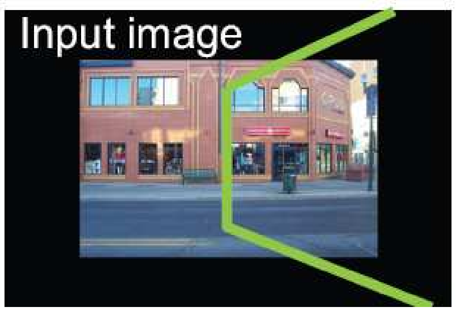
3. Find a match to fill the missing pixels

4. Locally align images

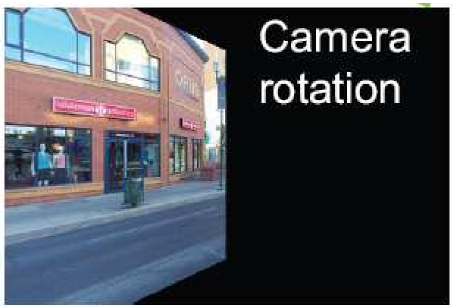
5. Find a seam

6. Blend in the gradient domain

Scene matching with camera view transformations: Camera rotation



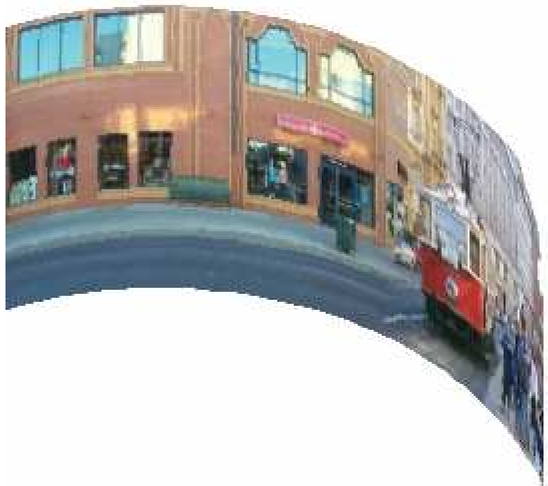
1. Rotate camera



2. View from the virtual camera

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

4. Stitched rotation



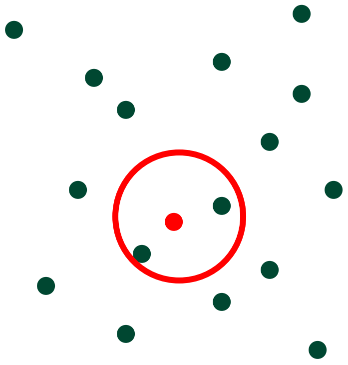
5. Display on a cylinder

Josef’s cool movie…

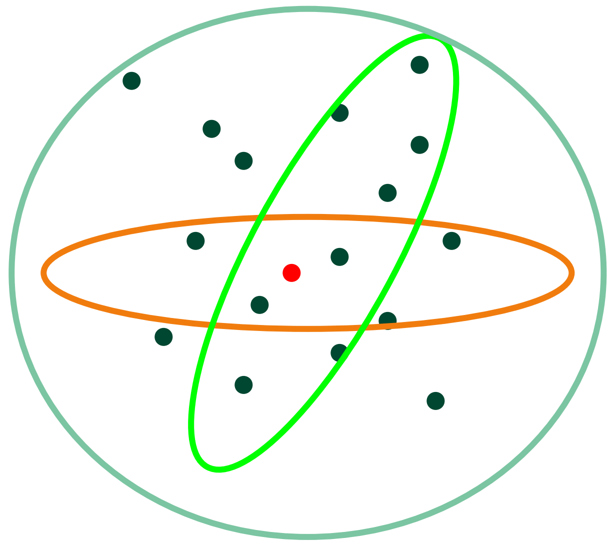
Data with labels (correspondences)

Useful even for really noisy labels!

Two simple ways to use Lots of Data



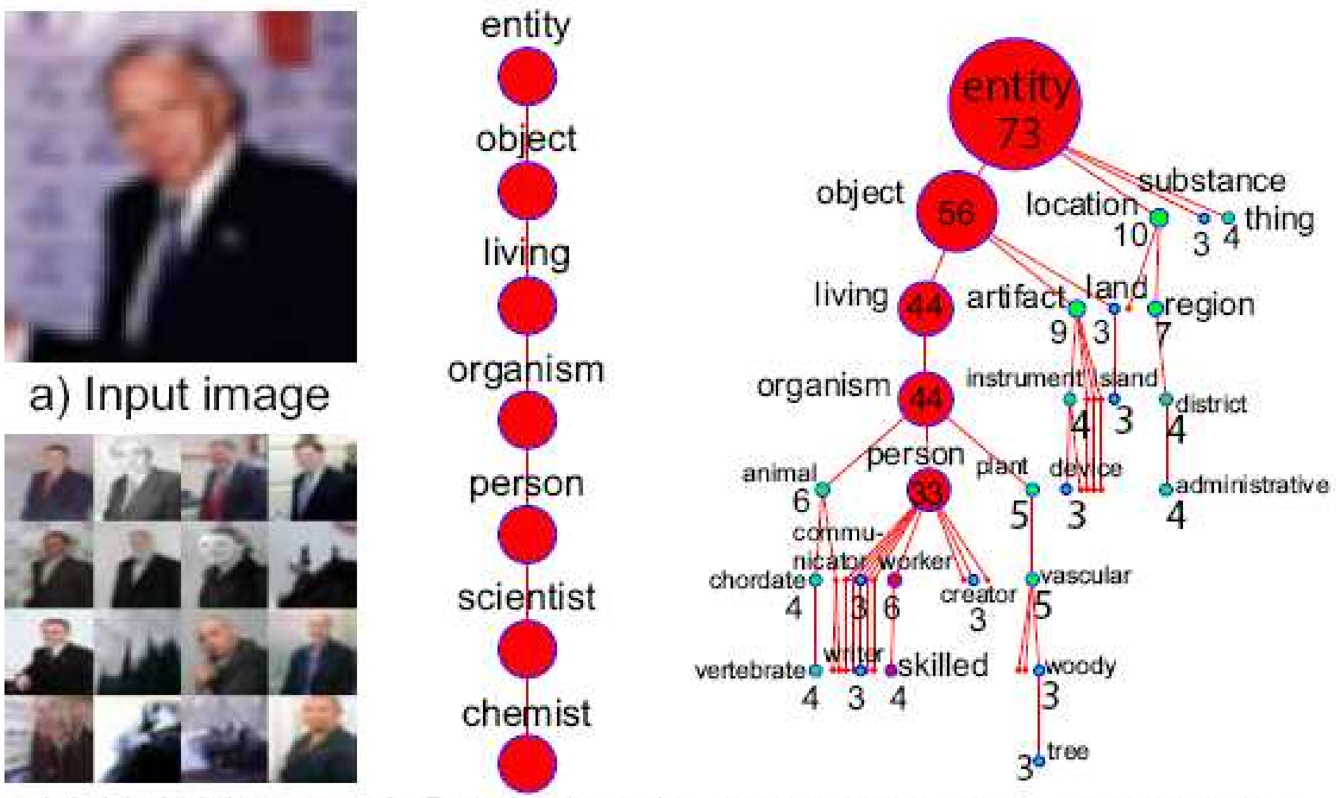
- Find that needle in the haystack and disregard the rest (a.k.a. kNN)



- See what different subsets of data think of you

1. kNN + Label Transfer

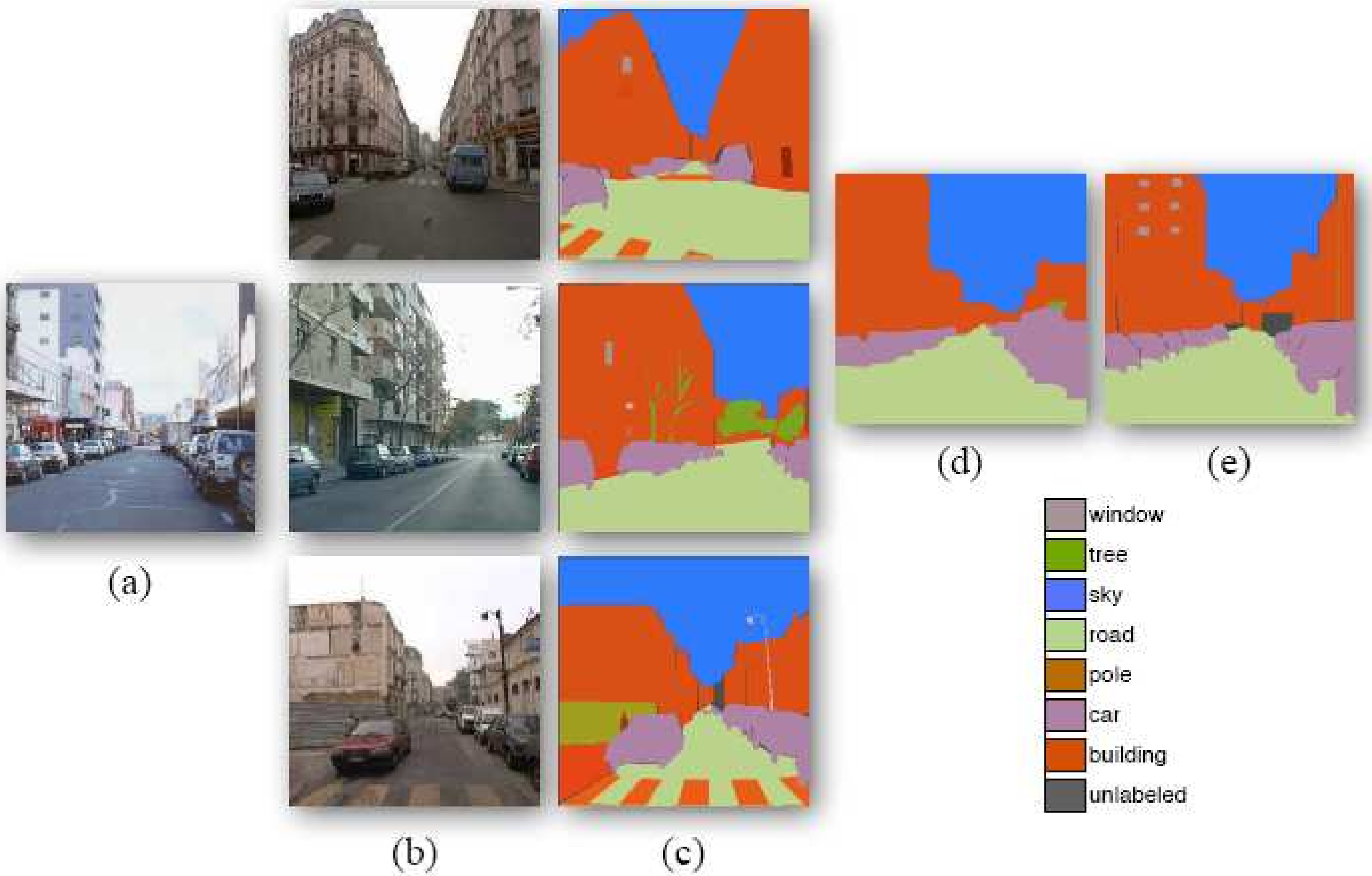
80 Million Tiny Images [PAMI'08]



b) Neighbors c) Ground truth d) Wordnet voted branches

Torralba, Fergus, Freeman, PAMI 2008

Non-parametric Scene Parsing [CVPR'09]



Liu, Yuen, Torralba, CVPR 2009

im2gps [CVPR'08]

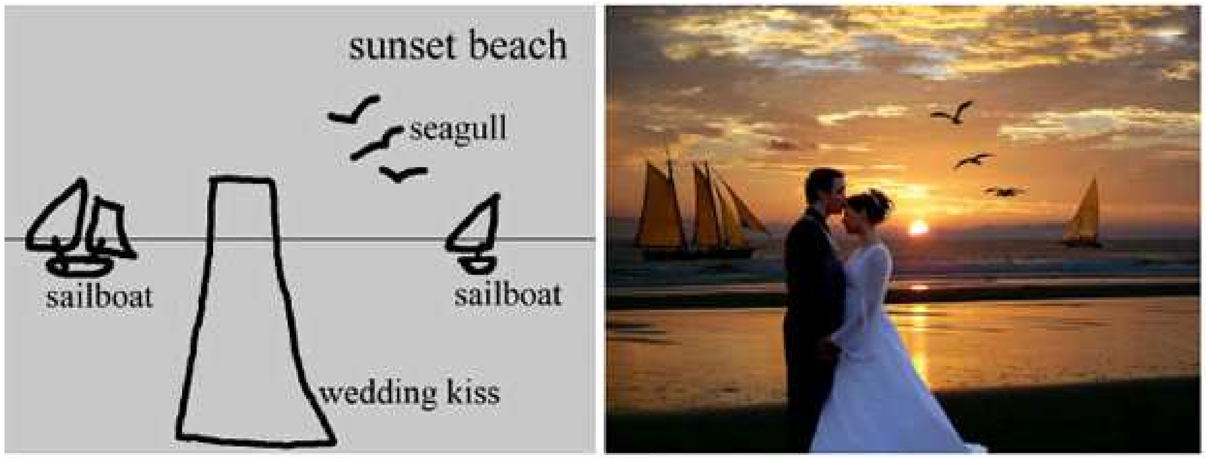
Query Photograph

Hays & Efros, CVPR 2008

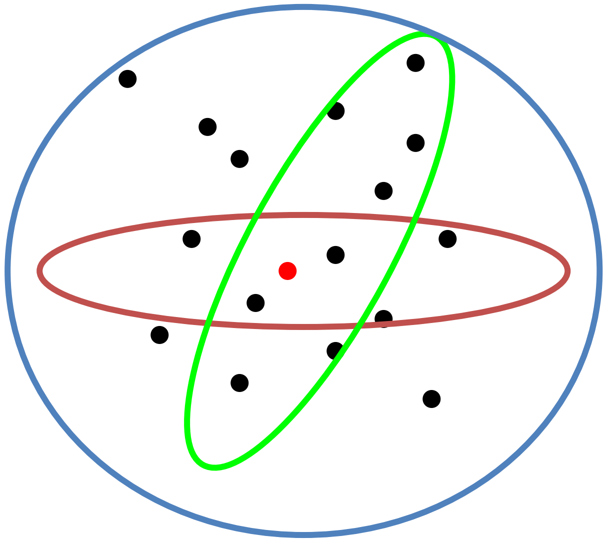
Assembling Visual Content

- Semantic Photo Synthesis [Johnson, ’06]

- Photo Clip Art [Lalonde, ‘07]
- Sketch2Photo [Chen,’09]



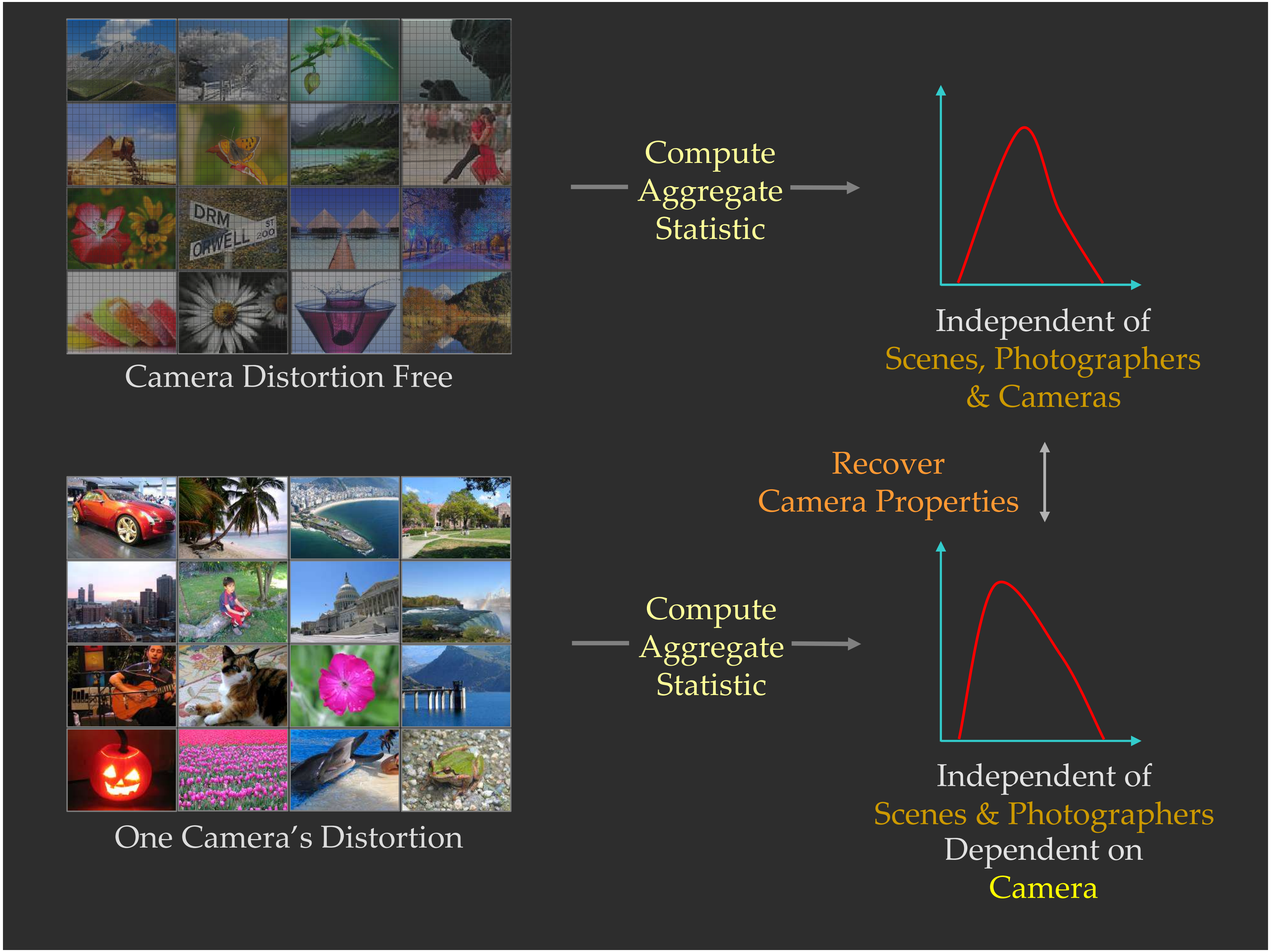
2. Subpopulation Labels



e.g. See Attributes

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

3. Relative (e.g. binary) labels

- Many concepts lack precise definition
  - E.g. beauty
- Or well-established boundaries
  - E.g. are curtains furniture?
- Or definition is context-specific:
  - E.g. "hairy" in "hairy dog" vs. "hairy man"

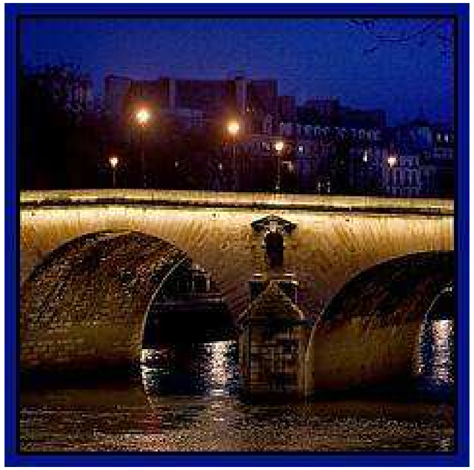
- Relative attributes:
  - Same/different, degree of similarity, etc.
  - Work starting on this, e.g.:
    - O. Tamuz, C. Liu, S. Belongie, O. Shamir and A. Kalai. Adaptively Learning the Crowd Kernel. ICML’11

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

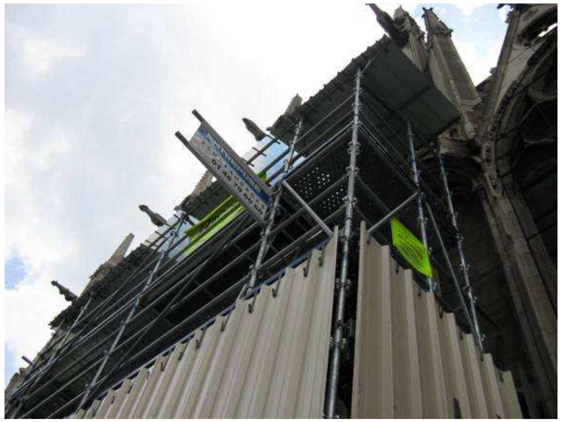









Real Notre Dame



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

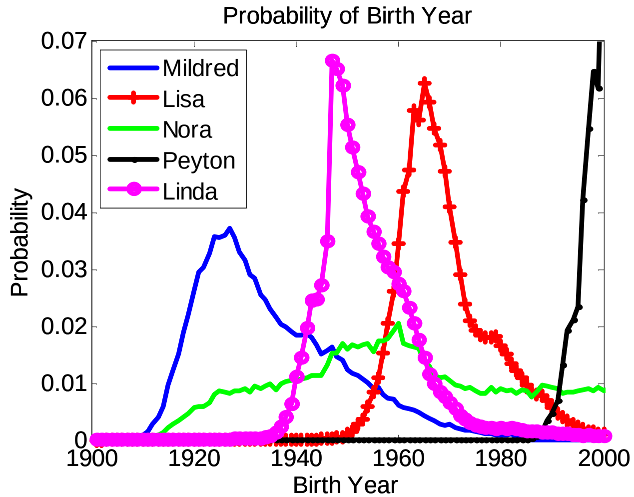
Kids with Santa

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

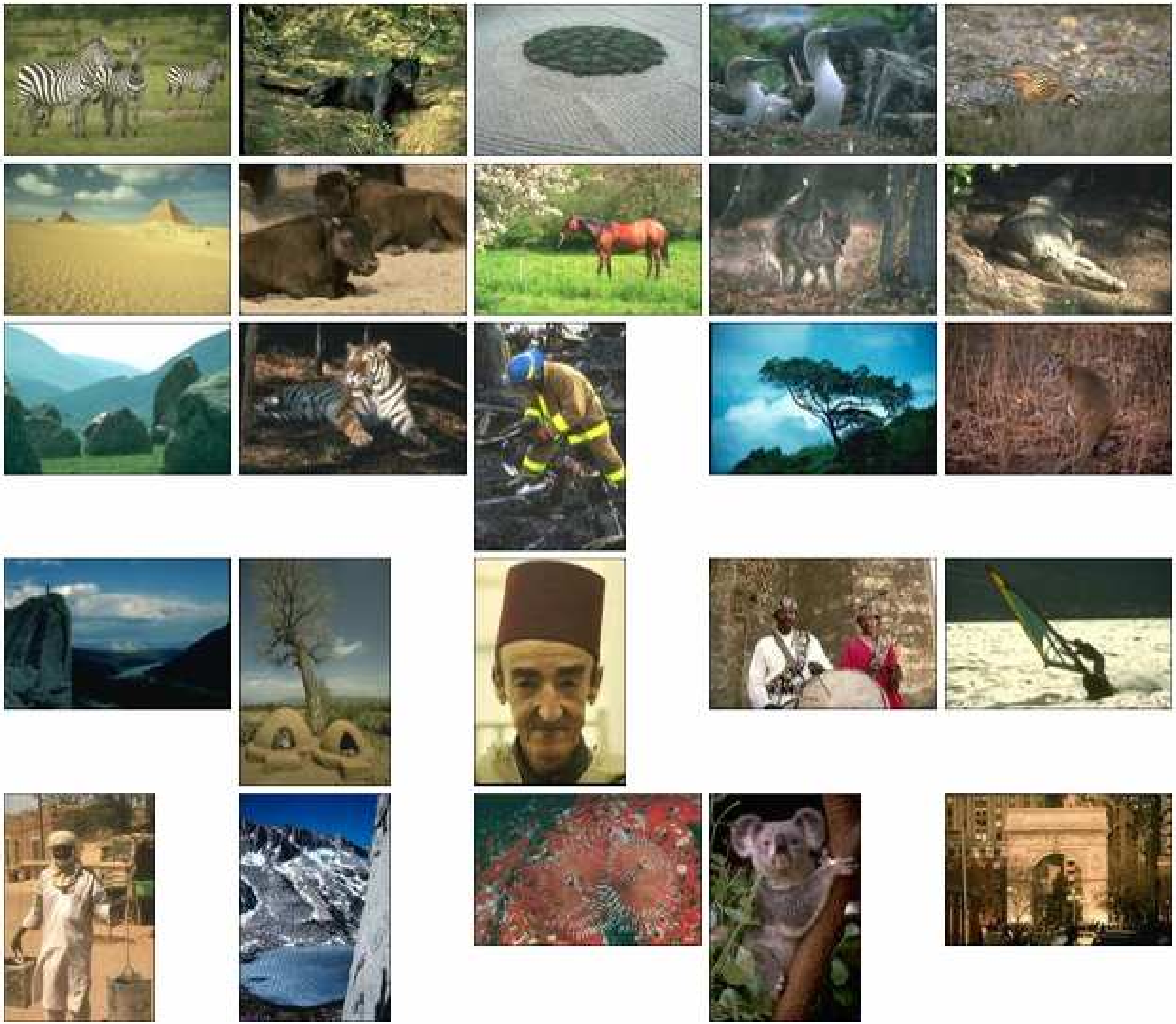
Brief History of Recognition Datasets

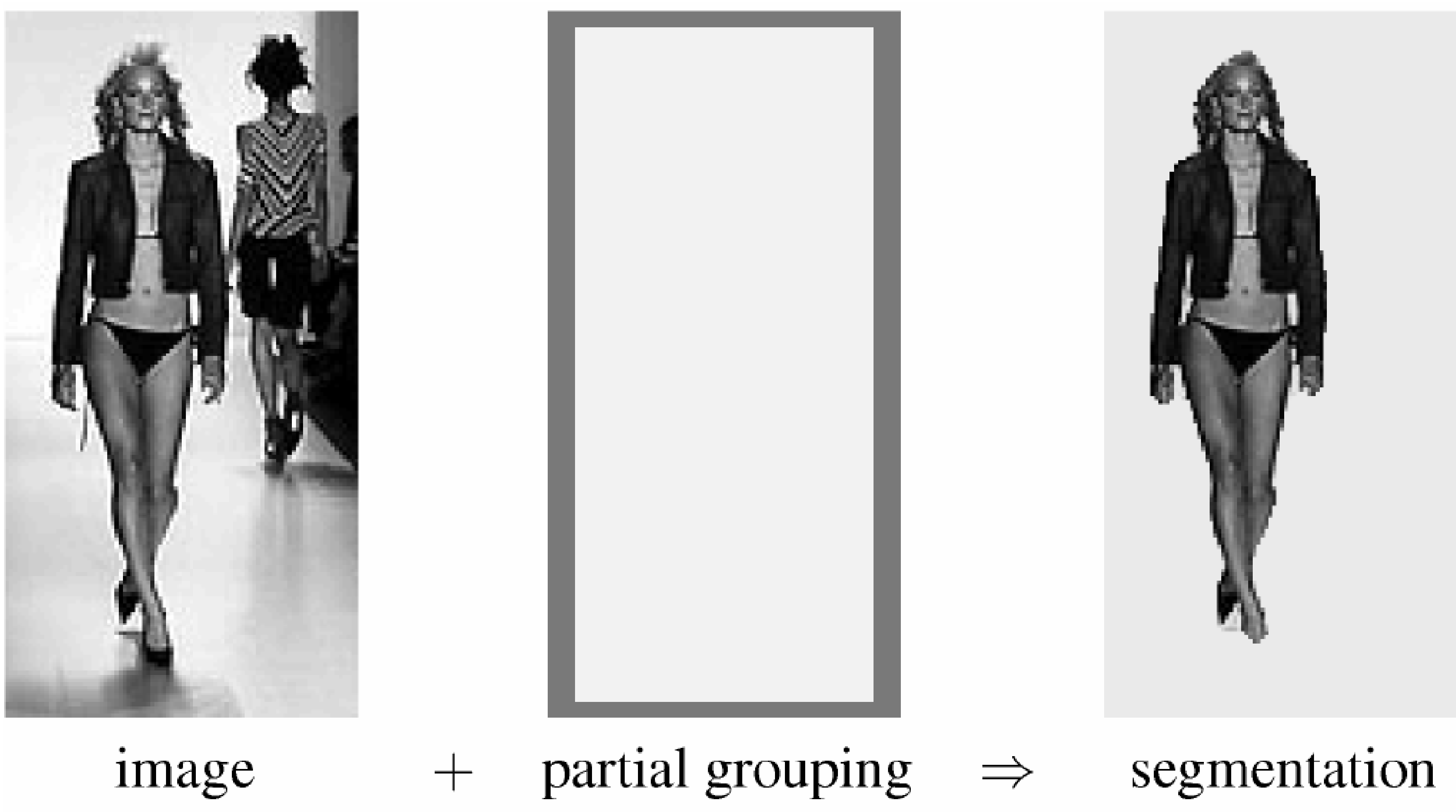
"trying to escape bias"

The first dataset

Columbia Object Image Library (COIL-100) (1996)

Corel Dataset



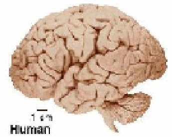


Yu & Shi, 2004

Caltech 101

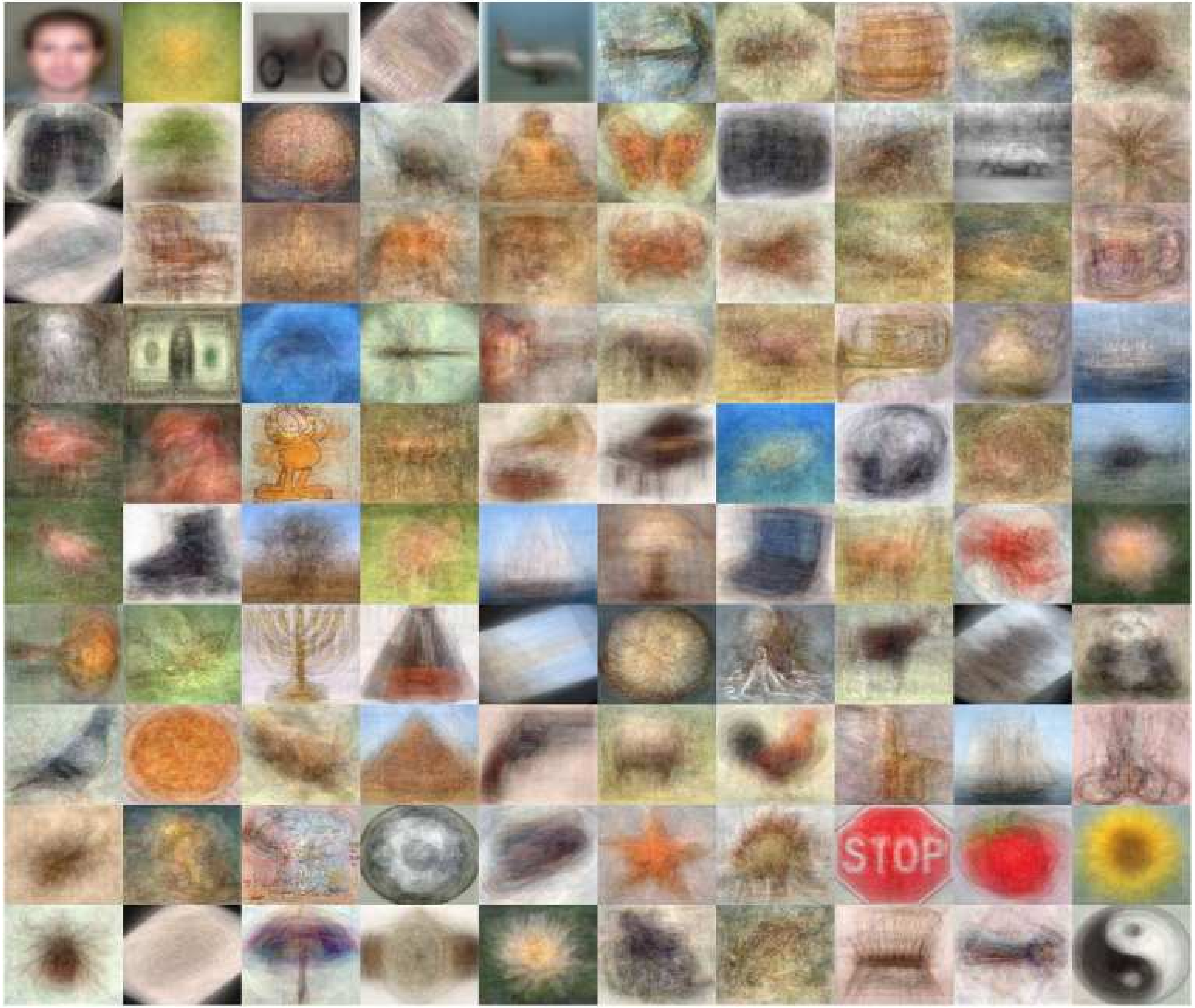
Caltech256

[Description][Download][Discussion [Other Datasets]

Description

Pictures of objects belonging to 101 categories. About 40 to 800 images per category. Most categories have about 50 images. Collected in September 2003 by Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato. The size of each image is roughly 300 x 200 pixels.

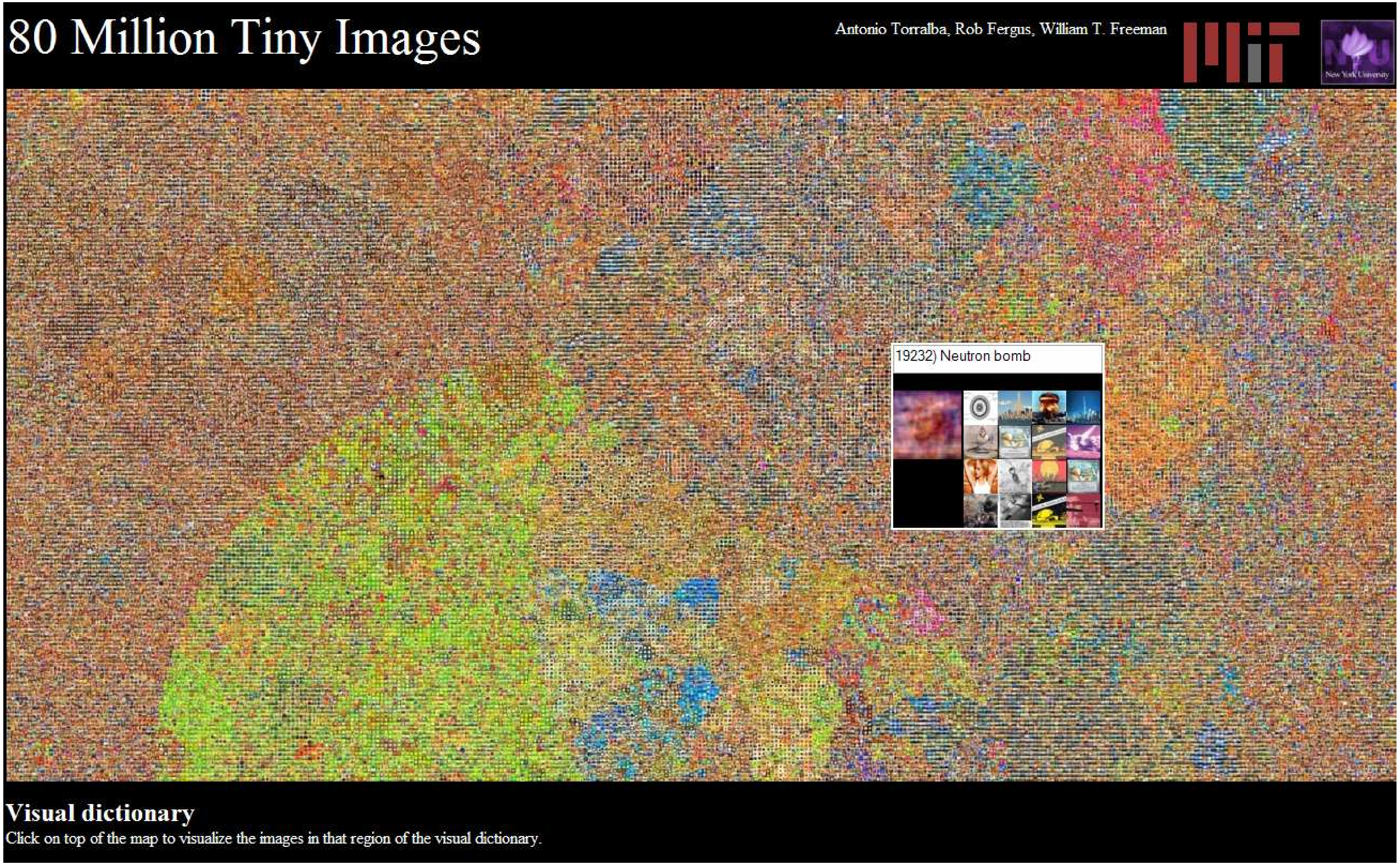


Average Caltech categories (Torralba)

Visual Object Classes Challenge 2011
(VOC2011)

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

TinyImages + ImageNet



Unbiased Look at Dataset Bias

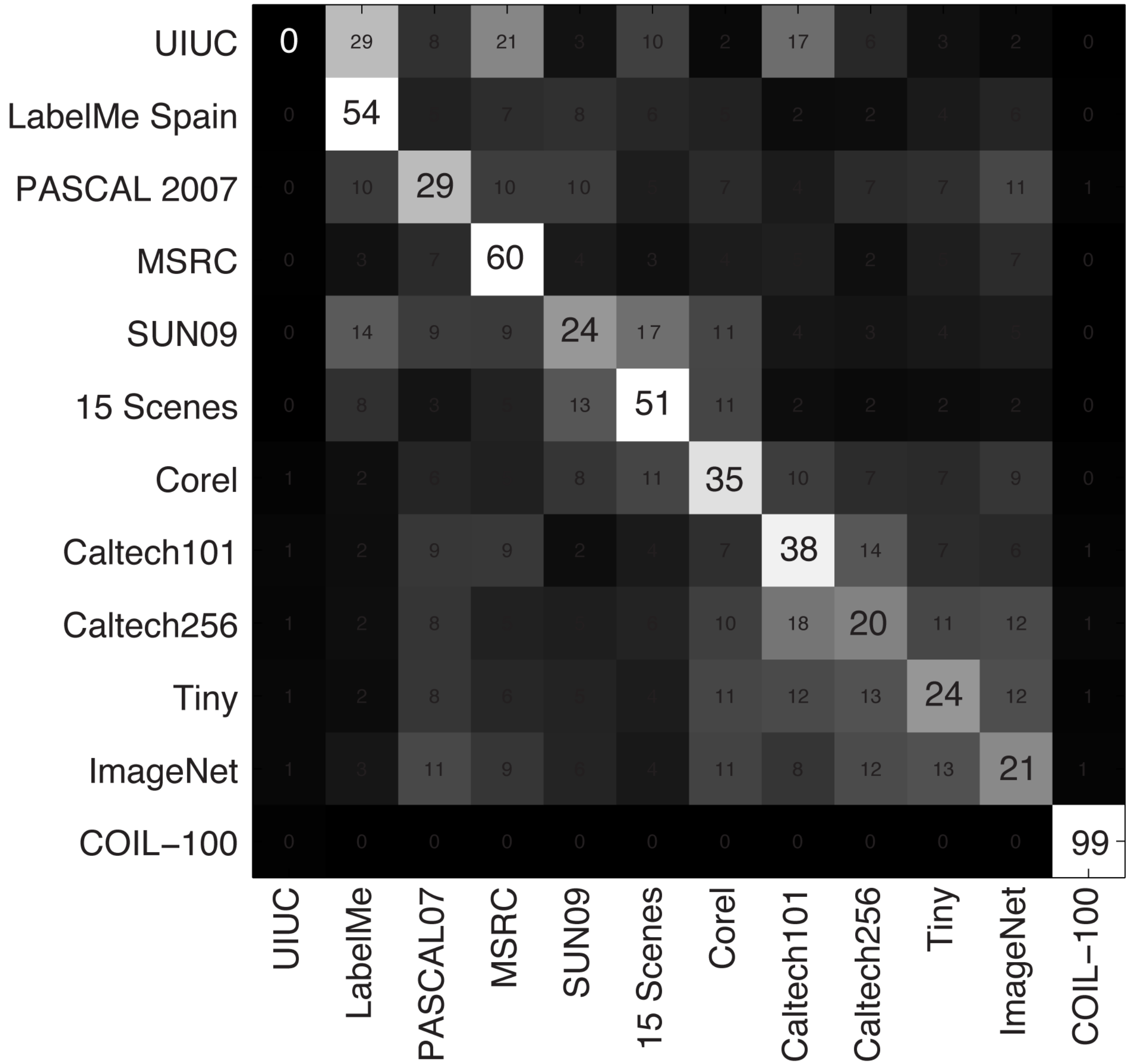
Torralba & Efros, CVPR 2011

- How much does this bias affect standard datasets used for object recognition?

“*Name That Dataset!*” game

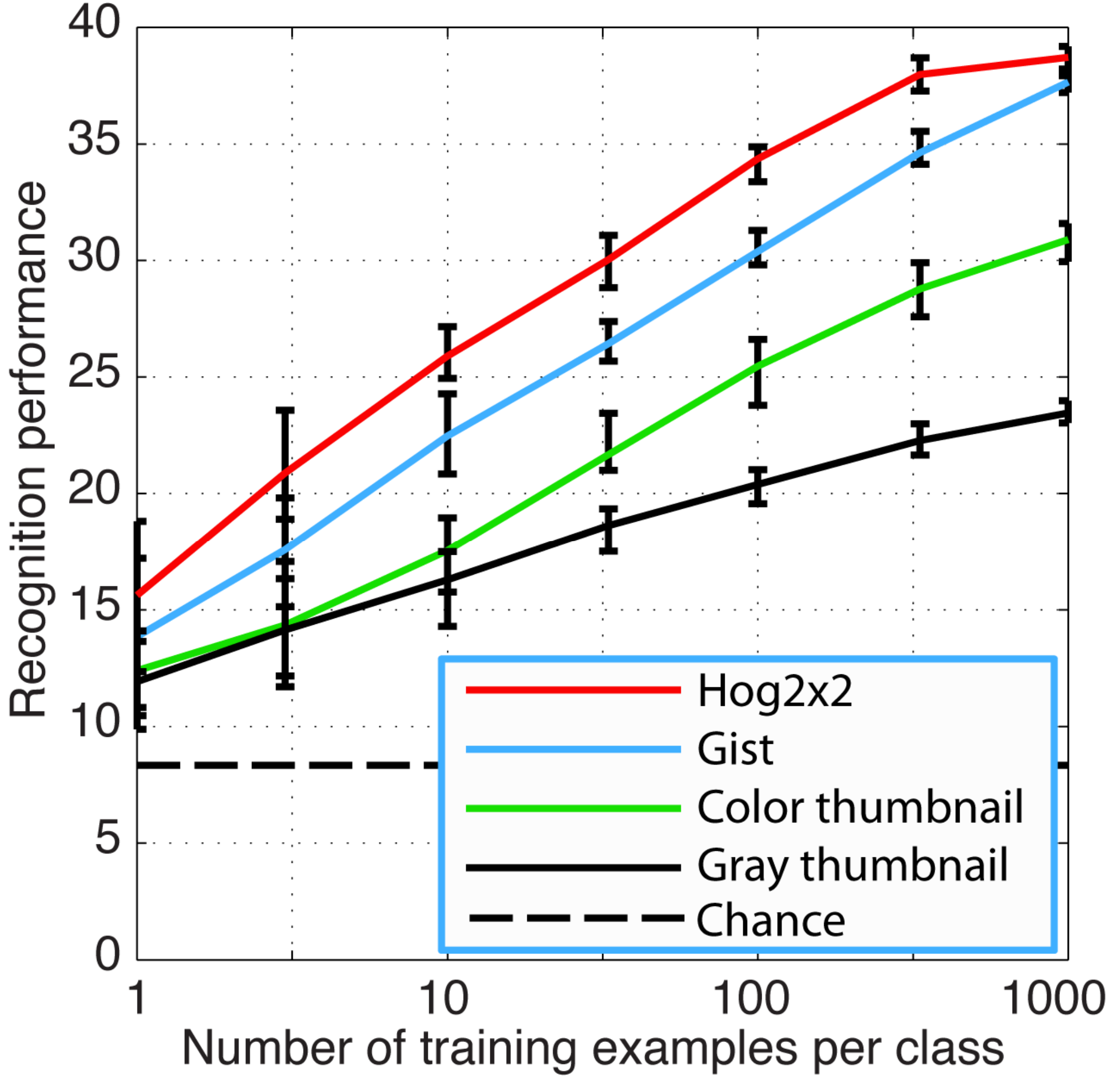
___ Caltech 101
___ Caltech 256
___ MSRC
___ UIUC cars
___ Tiny Images
___ Corel
___ PASCAL 2007
___ LabelMe
___ COIL-100
___ ImageNet
___ 15 Scenes
___ SUN’09

SVM plays *"Name that dataset!"*



- 12 1-vs-all classifiers
- Standard full-image features
- 39% performance (chance is 8%)

SVM plays *"Name that dataset!"*



Dataset look-alikes

ImageNet pretending to be:

Caltech 256 look-alikes from ImageNet

COREL look-alikes from ImageNet

MSRC look-alikes from ImageNet

PASCAL VOC pretending to be:

15 scenes look-a-likes from PASCAL 2007

MSRC look-alikes from PASCAL 2007

Caltech 101 look-alikes from PASCAL 2007

Datasets have different goals…

- Some are object-centric (e.g. Caltech, ImageNet)
- Otherwise are scene-centric (e.g. LabelMe, SUN’09)

- What about playing *“name that dataset”* on bounding boxes?

Similar results

PASCAL cars

SUN cars

Caltech101 cars

Performance: 61%
(chance: 20%)

ImageNet cars

LabelMe cars

Measuring Dataset Bias

Cross-Dataset Generalization

MSRC

Classifier trained on MSRC cars

Cross-dataset Performance

Table 1. Cross-dataset generalization. Object detection and classification performance (AP) for "car" and "person" when training on one dataset (rows) and testing on another (columns), i.e. each row is: training on one dataset and testing on all the others. "Self" refers to training and testing on the same dataset (same as diagonal), and "Mean Others" refers to averaging performance on all except self.

task	Test on: / Train on:	SUN09	LabelMe	PASCAL	ImageNet	Caltech101	MSRC	Self	Mean others	Percent drop
"car" classification	SUN09	**28.2**	29.5	16.3	14.6	16.9	21.9	28.2	19.8	**30%**
	LabelMe	14.7	**34.0**	16.7	22.9	43.6	24.5	34.0	24.5	**28%**
	PASCAL	10.1	25.5	**35.2**	43.9	44.2	39.4	35.2	32.6	**7%**
	ImageNet	11.4	29.6	36.0	**57.4**	52.3	42.7	57.4	34.4	**40%**
	Caltech101	7.5	31.1	19.5	33.1	**96.9**	42.1	96.9	26.7	**73%**
	MSRC	9.3	27.0	24.9	32.6	40.3	**68.4**	68.4	26.8	**61%**
	Mean others	10.6	28.5	22.7	29.4	39.4	34.1	53.4	27.5	48%
"car" detection	SUN09	**69.8**	50.7	42.2	42.6	54.7	69.4	69.8	51.9	**26%**
	LabelMe	61.8	**67.6**	40.8	38.5	53.4	67.0	67.6	52.3	**23%**
	PASCAL	55.8	55.2	**62.1**	56.8	54.2	74.8	62.1	59.4	**4%**
	ImageNet	43.9	31.8	46.9	**60.7**	59.3	67.8	60.7	49.9	**18%**
	Caltech101	20.2	18.8	11.0	31.4	**100**	29.3	100	22.2	**78%**
	MSRC	28.6	17.1	32.3	21.5	67.7	**74.3**	74.3	33.4	**55%**
	Mean others	42.0	34.7	34.6	38.2	57.9	61.7	72.4	44.8	**48%**
"person" classification	SUN09	**16.1**	11.8	14.0	7.9	6.8	23.5	16.1	12.8	**20%**
	LabelMe	11.0	**26.6**	7.5	6.3	8.4	24.3	26.6	11.5	**57%**
	PASCAL	11.9	11.1	**20.7**	13.6	48.3	50.5	20.7	27.1	**-31%**
	ImageNet	8.9	11.1	11.8	**20.7**	76.7	61.0	20.7	33.9	**-63%**
	Caltech101	7.6	11.8	17.3	22.5	**99.6**	65.8	99.6	25.0	**75%**
	MSRC	9.4	15.5	15.3	15.3	93.4	**78.4**	78.4	29.8	**62%**
	Mean others	9.8	12.3	13.2	13.1	46.7	45.0	43.7	23.4	**47%**
"person" detection	SUN09	**69.6**	56.8	37.9	45.7	52.1	72.7	69.6	53.0	**24%**
	LabelMe	58.9	**66.6**	38.4	43.1	57.9	68.9	66.6	53.4	**20%**
	PASCAL	56.0	55.6	**56.3**	55.6	56.8	74.8	56.3	59.8	**-6%**
	ImageNet	48.8	39.0	40.1	**59.6**	53.2	70.7	59.6	50.4	**15%**
	Caltech101	24.6	18.1	12.4	26.6	**100**	31.6	100	22.7	**77%**
	MSRC	33.8	18.2	30.9	20.8	69.5	**74.7**	74.7	34.6	**54%**
	Mean others	44.4	37.5	31.9	38.4	57.9	63.7	71.1	45.6	**36%**

Negative Set Bias

Table 2. Measuring Negative Set Bias.

task	Positive Set: / Negative Set:	SUN09	LabelMe	PASCAL	ImageNet	Caltech101	MSRC	Mean
"car" detection	self	67.6	62.4	56.3	60.5	97.7	74.5	70.0
	all	53.8	51.3	47.1	65.2	97.7	70.0	64.1
	percent drop	20%	18%	16%	-8%	0%	6%	8%
"person" detection	self	67.4	68.6	53.8	60.4	100	76.7	71.1
	all	52.2	58.0	42.6	63.4	100	71.5	64.6
	percent drop	22%	15%	21%	-5%	0%	7%	9%

Dataset Value

Table 3. "Market Value" for a "car" sample across datasets

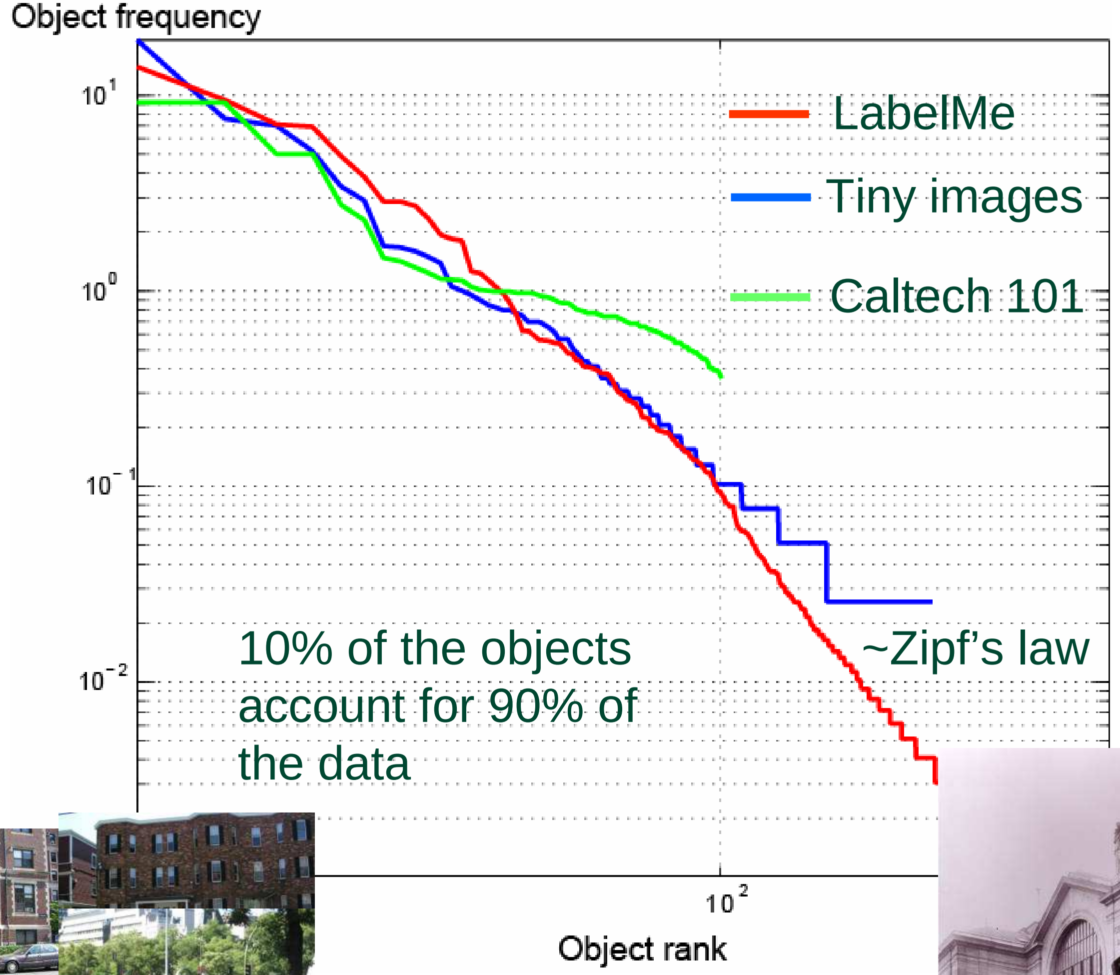
	SUN09 market	LabelMe market	PASCAL market	ImageNet market	Caltech101 market
1 SUN09 is worth	1 SUN09	0.91 LabelMe	0.72 pascal	0.41 ImageNet	0 Caltech
1 LabelMe is worth	0.41 SUN09	1 LabelMe	0.26 pascal	0.31 ImageNet	0 Caltech
1 pascal is worth	0.29 SUN09	0.50 LabelMe	1 pascal	0.88 ImageNet	0 Caltech
1 ImageNet is worth	0.17 SUN09	0.24 LabelMe	0.40 pascal	1 ImageNet	0 Caltech
1 Caltech101 is worth	0.18 SUN09	0.23 LabelMe	0 pascal	0.28 ImageNet	1 Caltech
Basket of Currencies	0.41 SUN09	0.58 LabelMe	0.48 pascal	0.58 ImageNet	0.20 Caltech

Overall…

- Caltech101, MSRC – bad
- PASCAL, ImageNet -- better

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

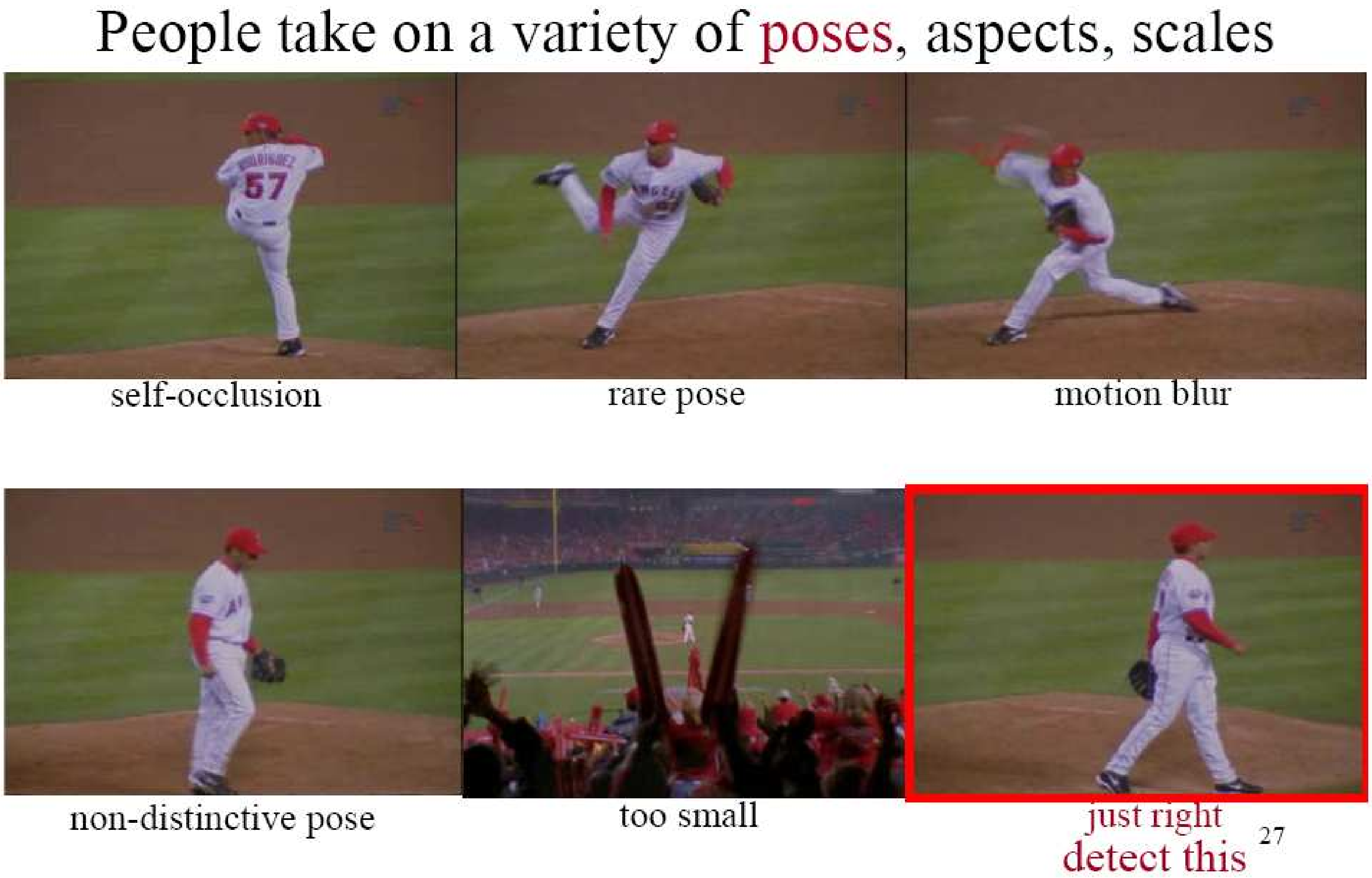
- e.g. reduce resolution, sifting, warping, etc.

segment into chunks

- e.g. segmentation for recognition approaches

Understand the simple stuff first

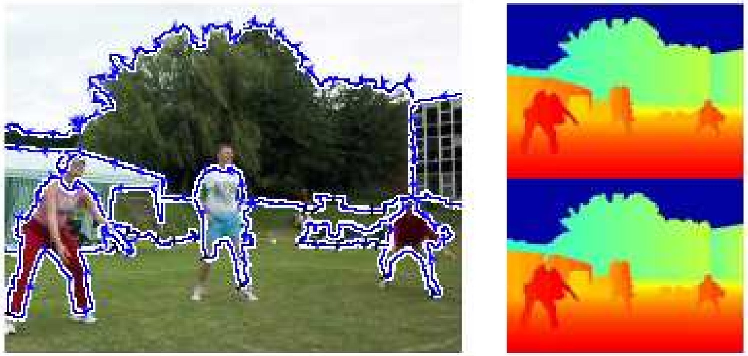
Recognize when it's easy!



Ramanan, Forsyth, Zisserman, 2004

“Poping out” foreground objects

Hoiem et al, ICCV 2007



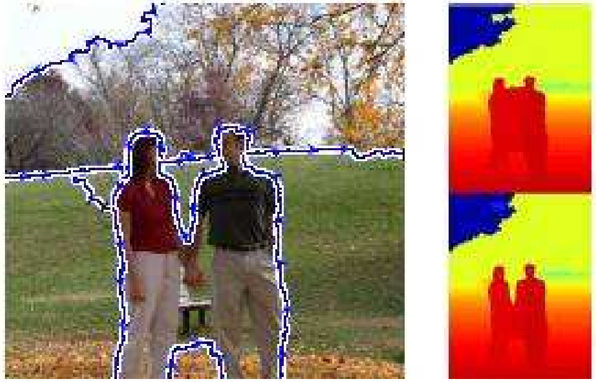
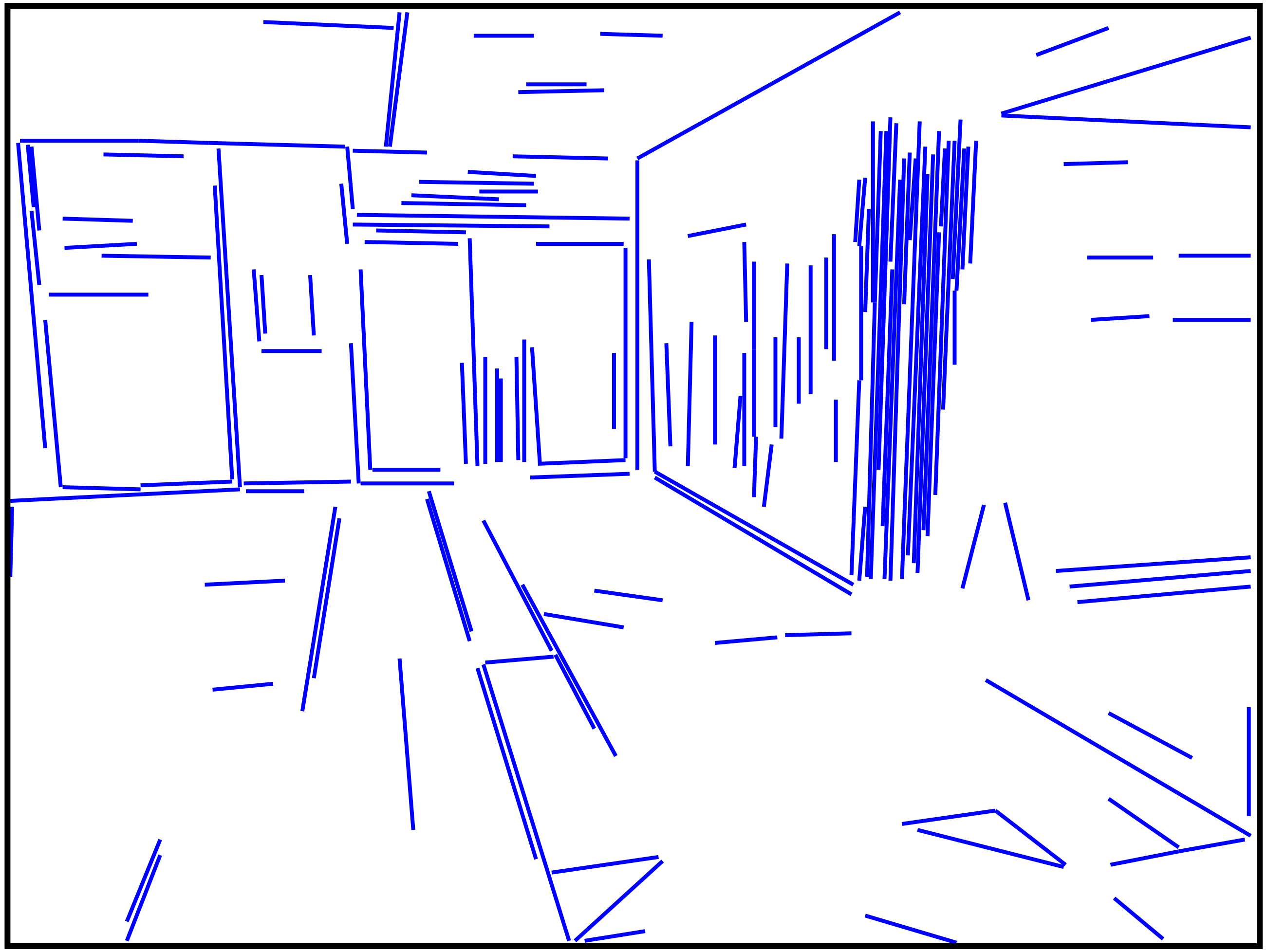


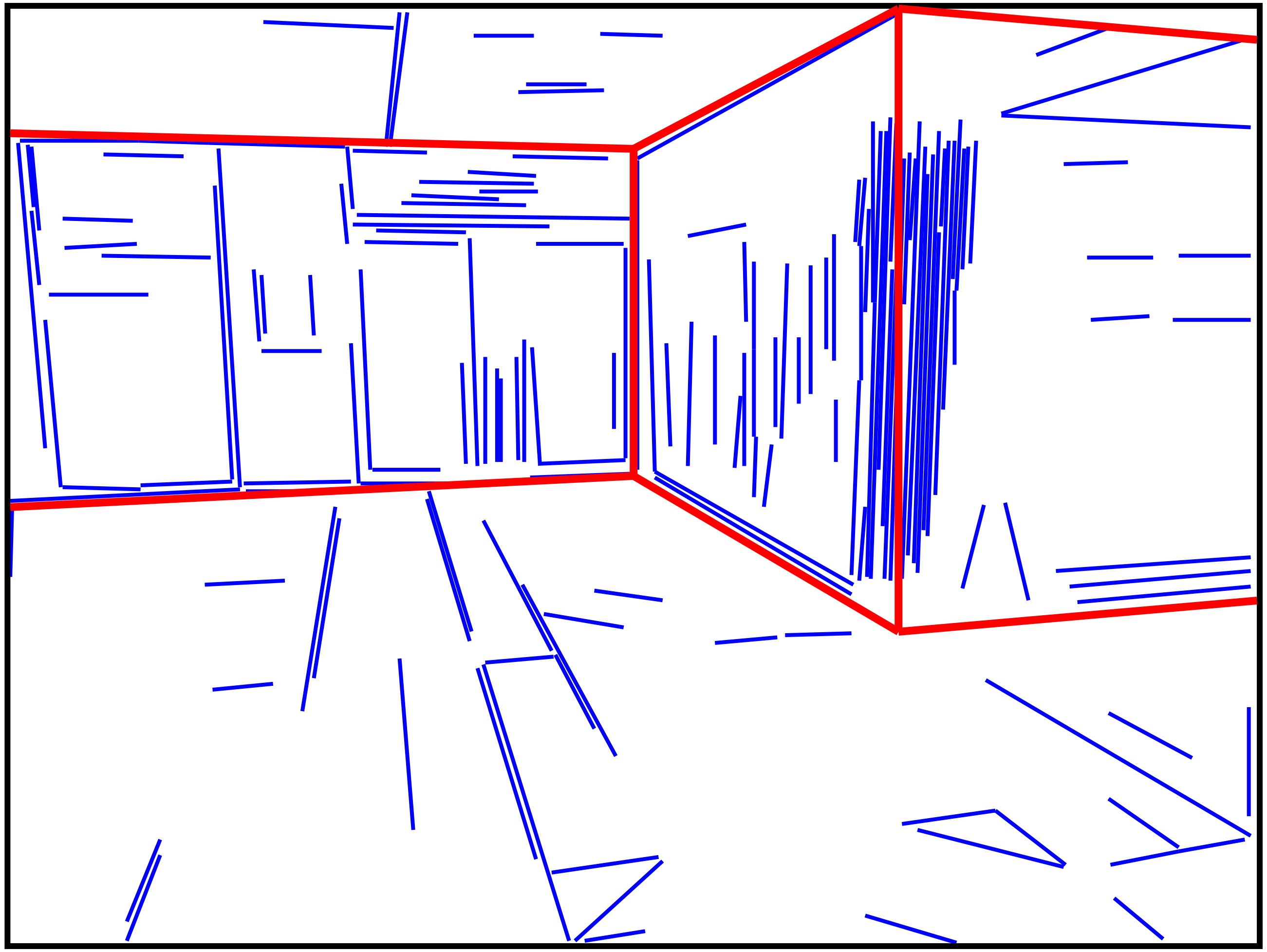
Figure 10. Object popout. We show five out of the fifteen most “solid” regions in the Geometric Context dataset. Our algorithm often finds foreground objects, which would be helpful for unsupervised object discovery [21].

Guess structure



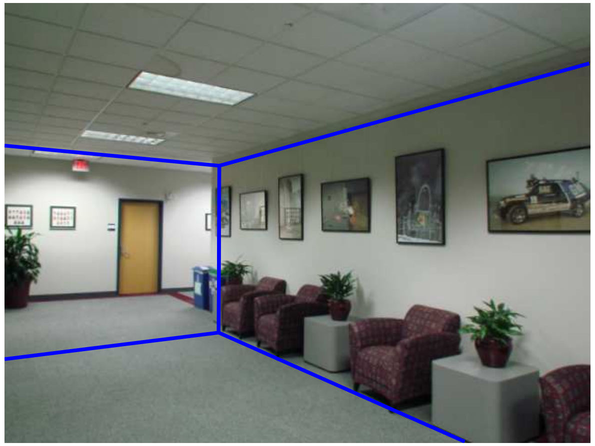
David C. Lee, Martial Hebert, Takeo Kanade, CVPR’09

Guess structure

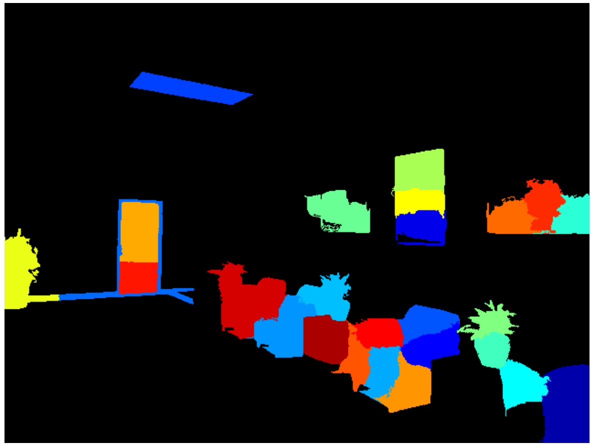


David C. Lee, Martial Hebert, Takeo Kanade, CVPR’09

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:

Attributes – densifying the labels

From categorization to association

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

Categorization vs. The Data

categorization is losing…

YAHOO! vs. Google™

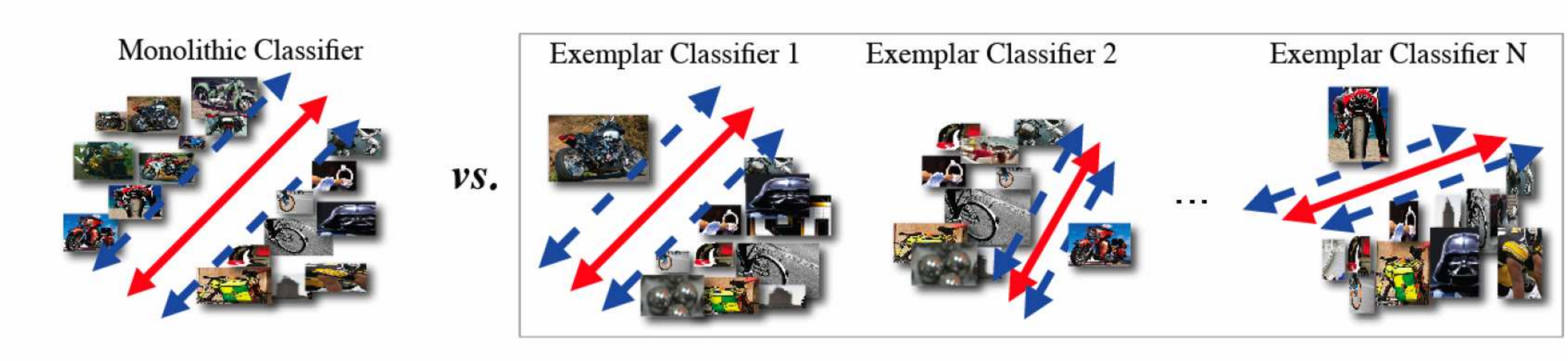
"…That which we call a rose
By any other name would smell as sweet."

"chair" category (PASCAL VOC)

"train" category (PASCAL VOC)

Malisiewicz et al, ICCV'11

Discriminative Exemplar-based Detector

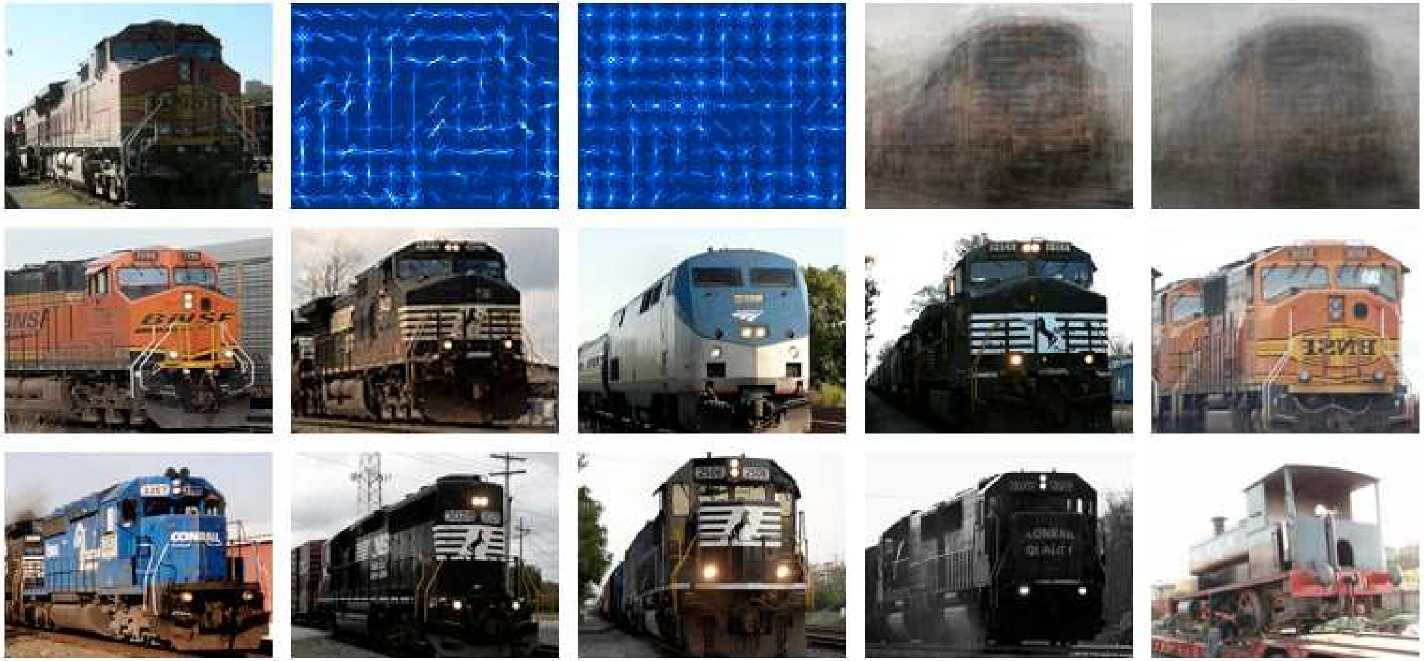


- Train a linear SVM for each positive instance
  - with lots of mined hard-negatives
- Use leave-one-out cross-validation to calibrate detectors
- At test time, run all detectors through non-max suppression to find winner

What’s Going On?

- Instead of one hard problem, many easy problems
- Each detector is an “associator”, an expert in it’s local neighborhood only
- More powerful than local distance learning – exemplar doesn’t have to reside at origin
- The negatives define the boundary
- Related to one-class SVMs, kernel SVMs, kernel learning, KNN-SVM… but no need for common kernel. Also get associations.

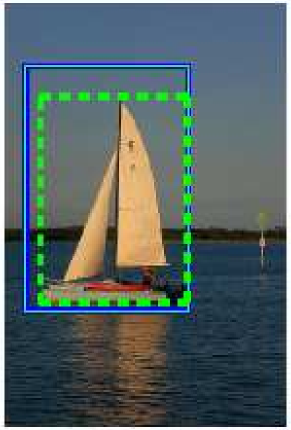
A sample instance detector


BNSF

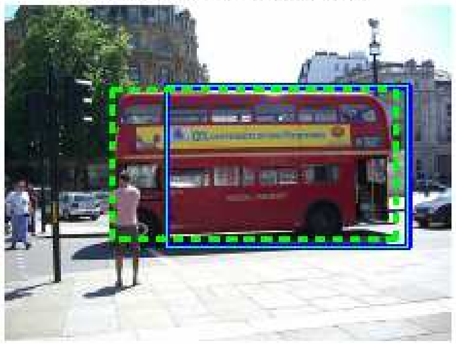
Results

Detection 6: Test Image 001305, OS=0.861

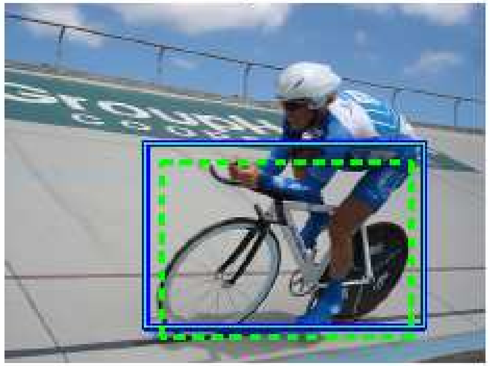
Detection 5: Test Image 002764, OS=0.762



Detection 2: Test Image 005609, OS=0.727



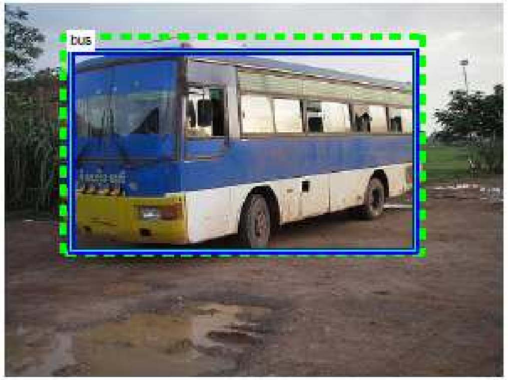
Detection 7: Test Image 006199, OS=0.768



Detection 3: Test Image 000178, OS=0.690

Detection 6: Test Image 001435, OS=0.821

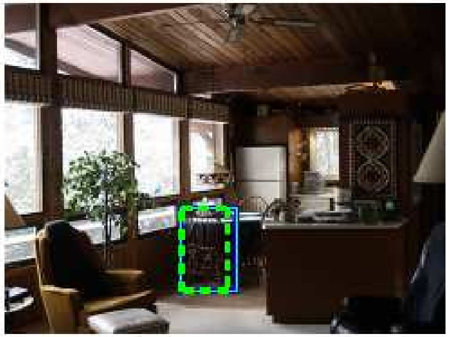
Detection 7: Test Image 005598, OS=0.906



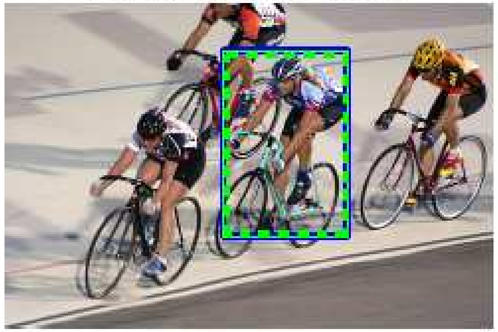
Detection 20: Test Image 002688, OS=0.832

Detection 8: Test Image 008266, OS=0.391

Detection 1: Test image 002353, OS=0.619



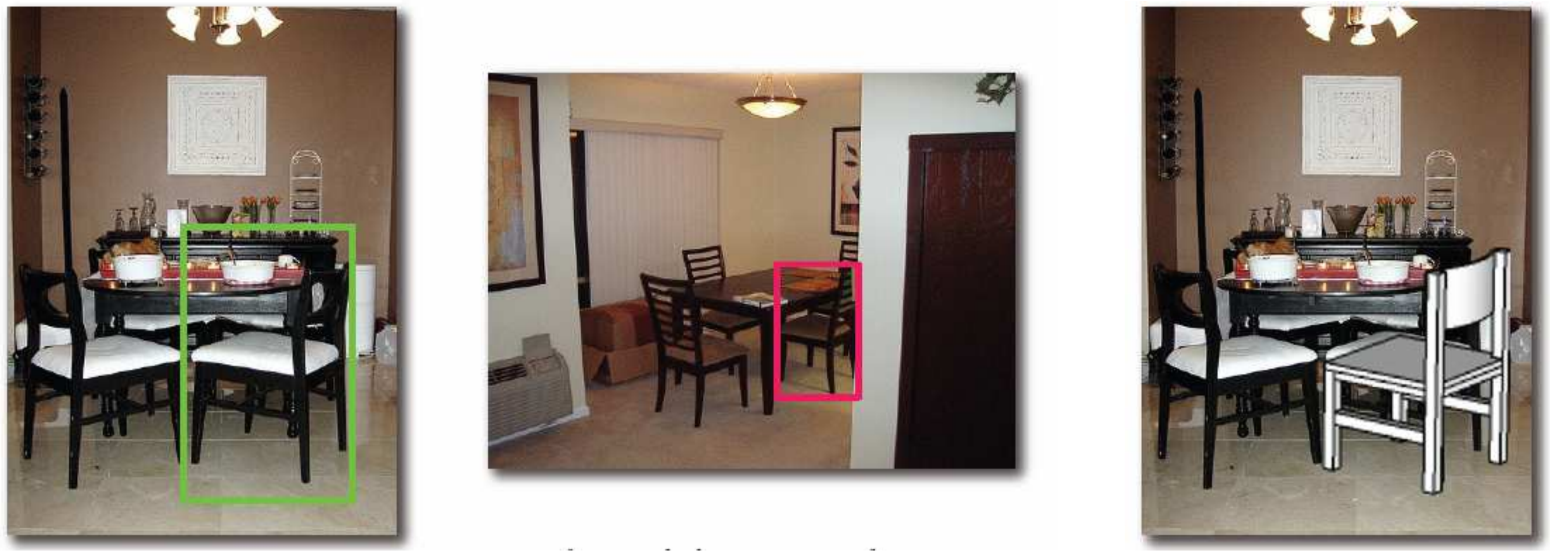
Detection 5: Test image 001496, OS=0.926



Detection 1: Test image 005086, OS=0.601

Label Transfer

- Now can easily transfer labels, segmentations, layouts, even 3D models:



Take-home Message

Large-scale data is necessary, but certainly not sufficient, to solve recognition

Corollary: *all the coolest stuff hasn't been done yet!*