

# Internet vision

Bill Freeman

MIT CSAIL

May 4, 2011

# Schedule the last week

5-minute presentations of your class projects

short homework due (describe the main points of your 5-minute presentation)

May 2011						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4
5	6	7	8	9	10	11

5-8 page papers due.

Time and location of final class presentations:

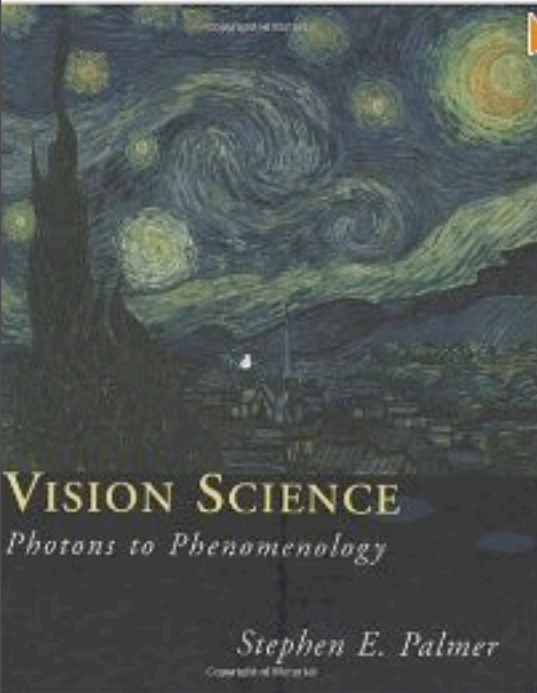
1:00pm - 3:30 or 4:00pm Wednesday

NOTE LOCATION: 3-343,

<http://web.mit.edu/registrar/classrooms/rooms/roompages/Buildings/Building3/3-343.html>

# Prize for best, clearest presentation

Click to **LOOK INSIDE!**



## Vision Science: Photons to Phenomenology [Hardcover] Stephen E. Palmer (Author), Linda A. Palmer (Contributor)

★★★★☆  (6 customer reviews) |  Like (1)

List Price: ~~\$88.00~~

Price: **\$83.81** & this item ships for **FREE with Super Saver Shipping**  
You Save: **\$4.19 (5%)**

**In Stock.**

Ships from and sold by **Amazon.com**. Gift-wrap available.

Only 11 left in stock--order soon (more on the way).

**Want it delivered Thursday, May 5?** Order it in the next 19 hours and 54 minutes.  
**Shipping** at checkout. [Details](#)

**26 new** from \$79.62    **22 used** from \$59.99



FREE Two-Day Shipping for Students. [Learn more](#)

# Outline

- datasets
- applications



# Some internet image datasets

- Flickr
- Facebook



# Dogs Named Banjo

[Group Pool](#) [Discussion](#) [23 Members](#) [Map](#) [Join This Group](#)

[More](#) ▾

**Group Pool** 70 items | Only members can add to the pool. [Join?](#)



by [bastique](#)



by [Jorn Idzerda](#)



by [Jorn Idzerda](#)



by [surfer\\_vero](#)



by [D.Grayson](#)



by [D.Grayson](#)



by [D.Grayson](#)



by [D.Grayson](#)



by [bastique](#)



by [surfer\\_vero](#)



by [Jorn Idzerda](#)



by [Jorn Idzerda](#)

**SEARCH**

[» More photos](#)



**dognamedbanjo (a group admin) says:**

14 Feb 08 - Hi Everyone! This is a group dedicated to all dogs named Banjo. They come in all shapes and sizes and are all lovable in their own way! Do you know a Banjo? Add a picture here and enjoy the infinite doggie cuteness!

**Discussion** 1 post | Only members can post. [Join?](#)

Title

Author

Replies

Latest Post

**NEW** [How did you decide to name your dog Banjo?](#)

[dognamedbanjo](#)

5

28 months ago

facebook

Search

SafeSearch moderate

About 1,520,000,000 results (0.07 seconds)

Advanced search

Related searches: [facebook emoticons](#) [facebook logo](#) [facebook icon](#) [facebook page](#) [facebook funny](#) [facebook png](#)






Photos of Maddie Freeman in  
Wedding Day  
By Daniel Hoadley - 3 of 2,017

In this photo: Taylor Jackman, Maddie Freeman (photos)

Saturday

 Taylor Jackman likes this.

[Tag This Photo](#)

Wednesday, May 4, 2011



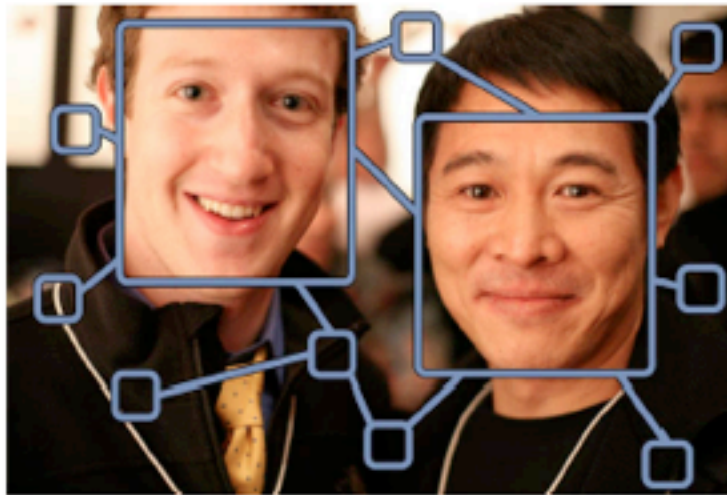
# Toward Large-Scale Face Recognition Using Social Network Context

*The authors of this paper believe that social incentives can be used to obtain numerous facial images of faces and they propose a computational method for using these images.*

By ZAK STONE, Student Member IEEE, TODD ZICKLER, Member IEEE, AND  
TREVOR DARRELL, Member IEEE

[http://www.eecs.harvard.edu/~zickler/papers/SocialContext\\_ProcIEEE2010.pdf](http://www.eecs.harvard.edu/~zickler/papers/SocialContext_ProcIEEE2010.pdf)

**ABSTRACT** | Personal photographs are being captured in digital form at an accelerating rate, and our computational tools for searching, browsing, and sharing these photos are struggling to keep pace. One promising approach is automatic face recognition, which would allow photos to be organized by the identities of the individuals they contain. However, achieving accurate recognition at the scale of the Web requires discriminating among hundreds of millions of individuals and would seem to be a daunting task. This paper argues that social network context may be the key for large-scale face recognition to succeed. Many personal photographs are shared on the Web through online social network sites, and we can leverage the resources and structure of such social networks to improve face recognition rates on the images shared. Drawing upon real photo collections from volunteers who are members of a popular online social network, we assess the availability of resources to improve face recognition and discuss techniques



**Fig. 1.** The billions of personal photographs shared in online social networks present a new opportunity to develop “socially aware” face recognition systems. By leveraging contextual information about the

# Collecting labeled datasets

- ESP game (CMU)

Luis Von Ahn and Laura Dabbish 2004

<http://www.gwap.com/gwap/>

- LabelMe (MIT)

Russell, Torralba, Freeman, 2005

<http://labelme.csail.mit.edu/>

- 80 Million Tiny Images

Torralba, Fergus, Freeman 2008

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

- ImageNet

Li, Fei-Fei, 2009

<http://www.image-net.org/>

- Mechanical Turk

Amazon

<https://www.mturk.com/mturk/welcome>



# Collecting labeled datasets

- ESP game (CMU)

Luis Von Ahn and Laura Dabbish 2004

<http://www.gwap.com/gwap/>

- LabelMe (MIT)

Russell, Torralba, Freeman, 2005

- 80 Million Tiny Images

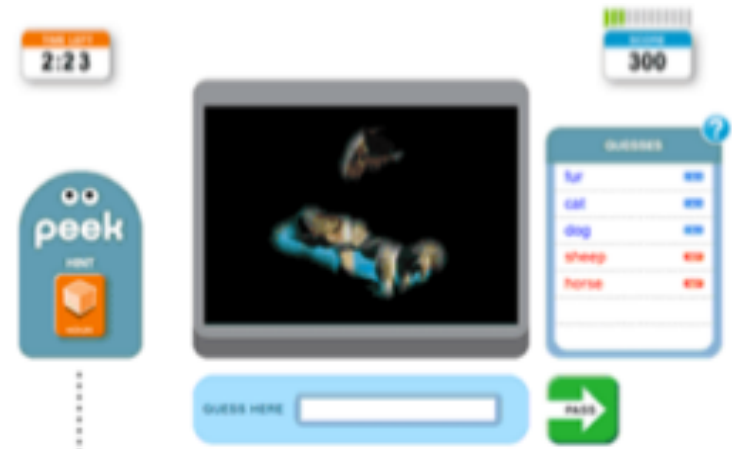
Torralba, Fergus, Freeman 2008

- ImageNet

Li, Fei-Fei, 2009

- Mechanical Turk

Amazon







# Collecting labeled datasets

- ESP game (CMU)

Luis Von Ahn and Laura Dabbish 2004

- LabelMe (MIT)

Russell, Torralba, Freeman, 2005

- 80 Million Tiny Images

Torralba, Fergus, Freeman 2008

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

- ImageNet

Li, Fei-Fei, 2009

- Mechanical Turk

Amazon

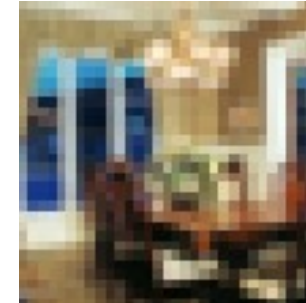


# 32x32 images

256x256



32x32

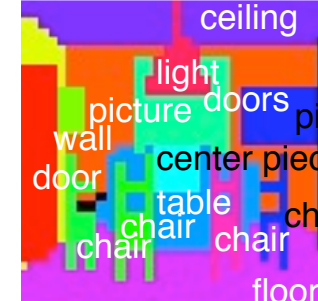
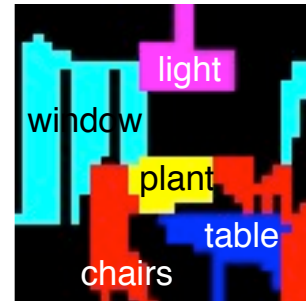
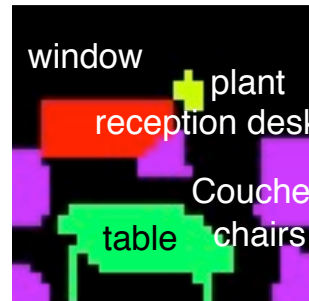
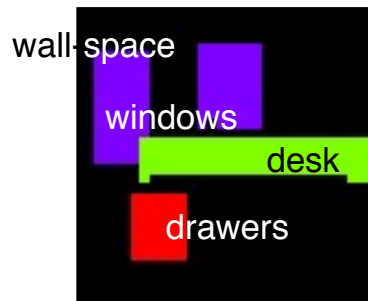


office

waiting area

dining room

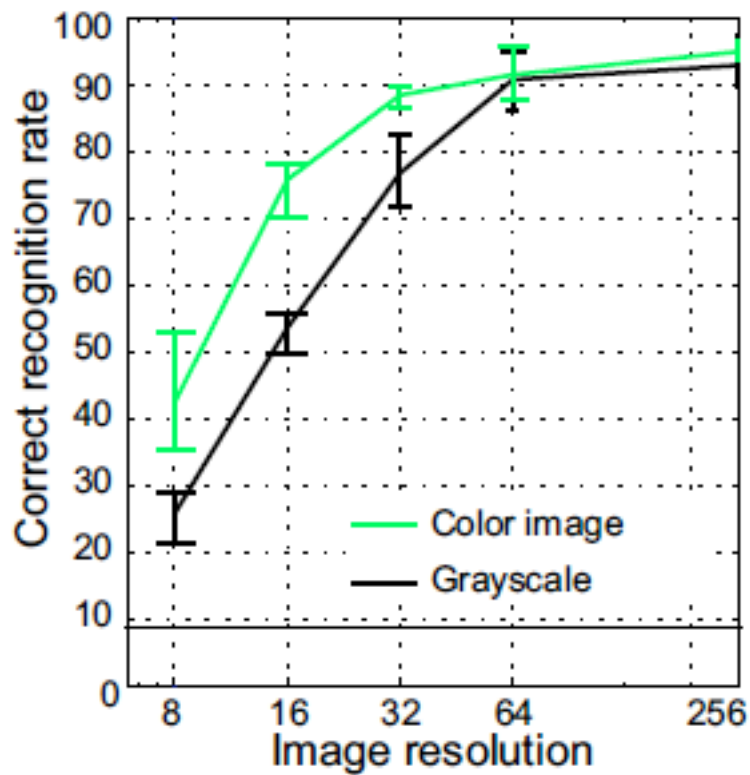
dining room



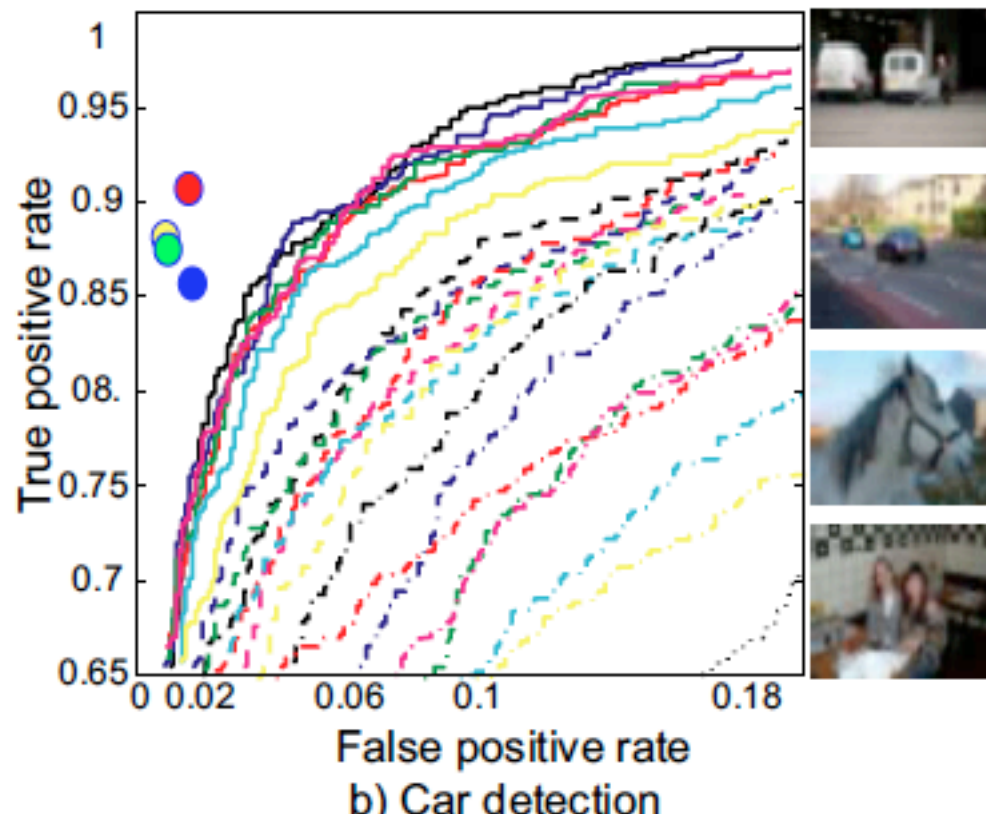
Human's are capable of recognizing and segmenting images with just 32x32 pixels



d) Cropped objects



a) Scene recognition

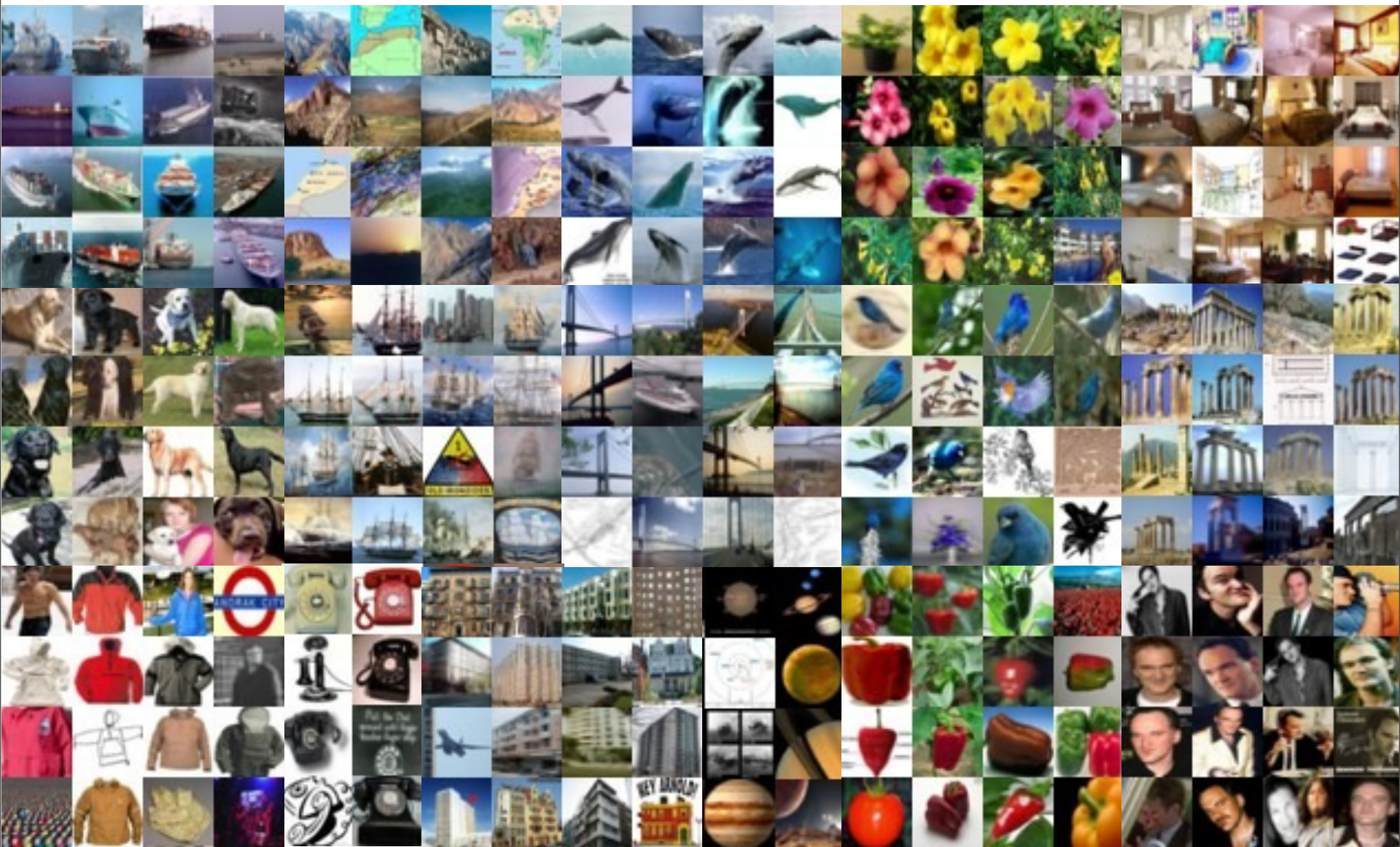


b) Car detection

Fig. 1. a) Human performance on scene recognition as a function of resolution. The green and black curves show the performance on color and gray-scale images respectively. For color  $32 \times 32$  images the performance only drops by 7% relative to full resolution, despite having 1/64th of the pixels. b) Car detection task on the PASCAL 2006 test dataset. The colored dots show the performance of four human subjects classifying tiny versions of the test data. The ROC curves of the best vision algorithms (running on full resolution images) are shown for comparison. All lie below the performance of humans on the tiny images, which rely on none of the high-resolution cues exploited by the computer vision algorithms. c) Humans can correctly recognize and segment objects at very low resolutions, even when the objects in isolation can not be recognized (d).



# Tiny images: 1000 pixels





# WordNet

A lexical database for English



<http://wordnet.princeton.edu/>

## About WordNet

### ■ About WordNet

[Use WordNet online](#)

[Download](#)

[Frequently Asked Questions](#)

[Related projects](#)

[WordNet documentation](#)

[WordNet](#)

WordNet® is a large lexical database of English, developed under the direction of [George A. Miller](#) (Emeritus). Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the [browser](#). WordNet is also freely and publicly available for [download](#). WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

Over the years, many people have contributed to the development of WordNet. Currently, the WordNet team includes the following members, and the WordNet project is housed in the Department of Computer Science:

We appreciate your comments and suggestions, especially when they are constructive and help us improve WordNet. Please contact us at [\[email\]](#).

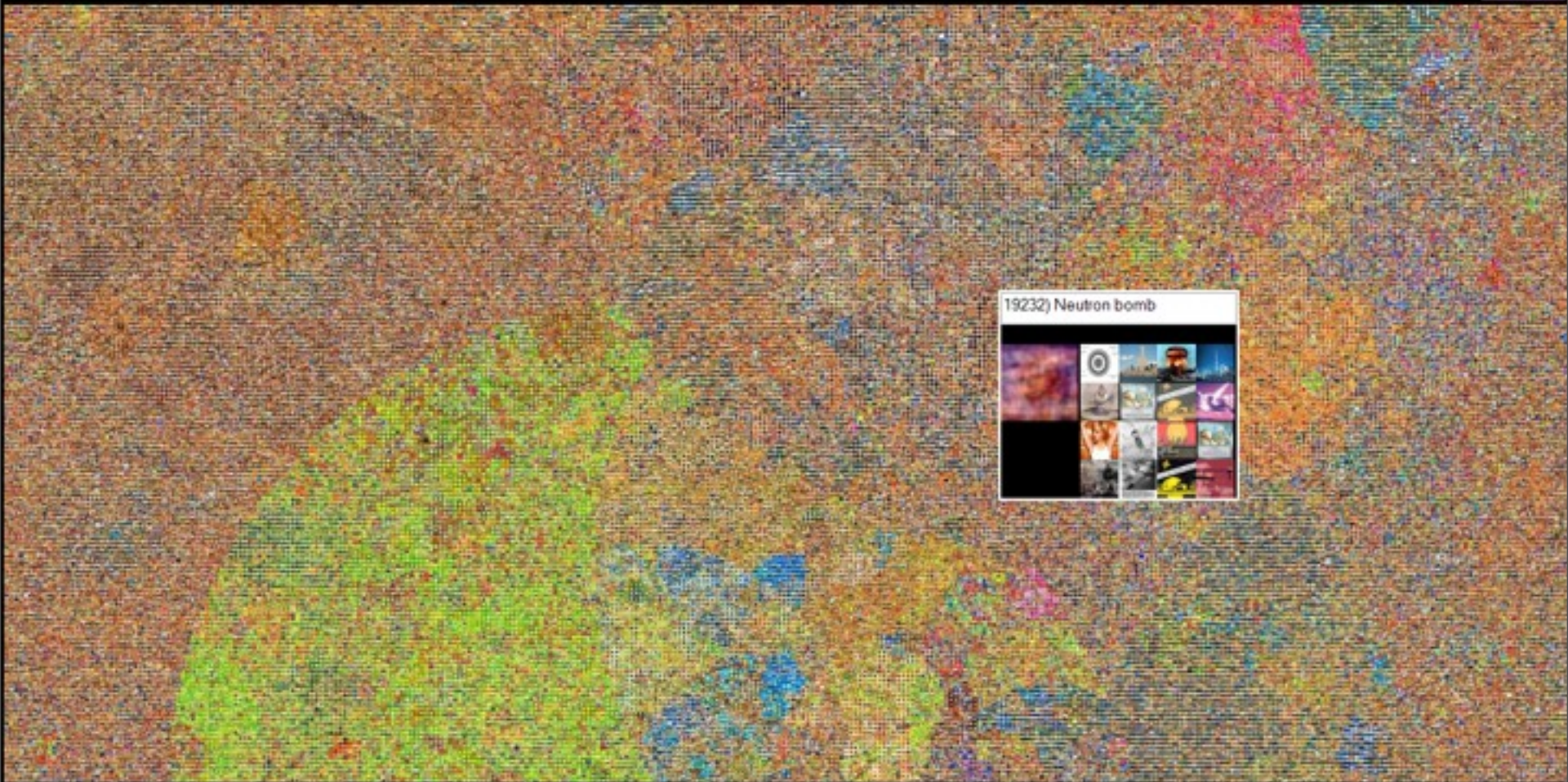
Our staff examines all mail and tries to make appropriate changes, but we hope you understand that due to time constraints we cannot always respond to the sender.

Please note that changes made to the database are not reflected until a new version



# 80 Million Tiny Images

Antonio Torralba, Rob Fergus, William T. Freeman

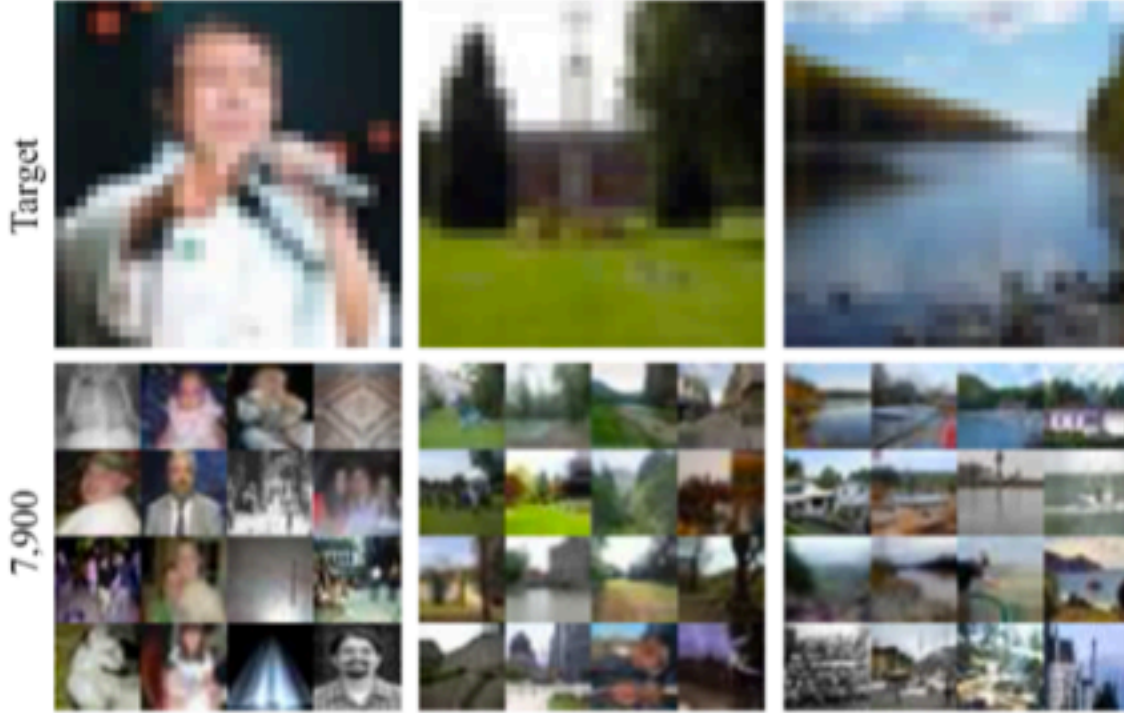


## Visual dictionary

Click on top of the map to visualize the images in that region of the visual dictionary.

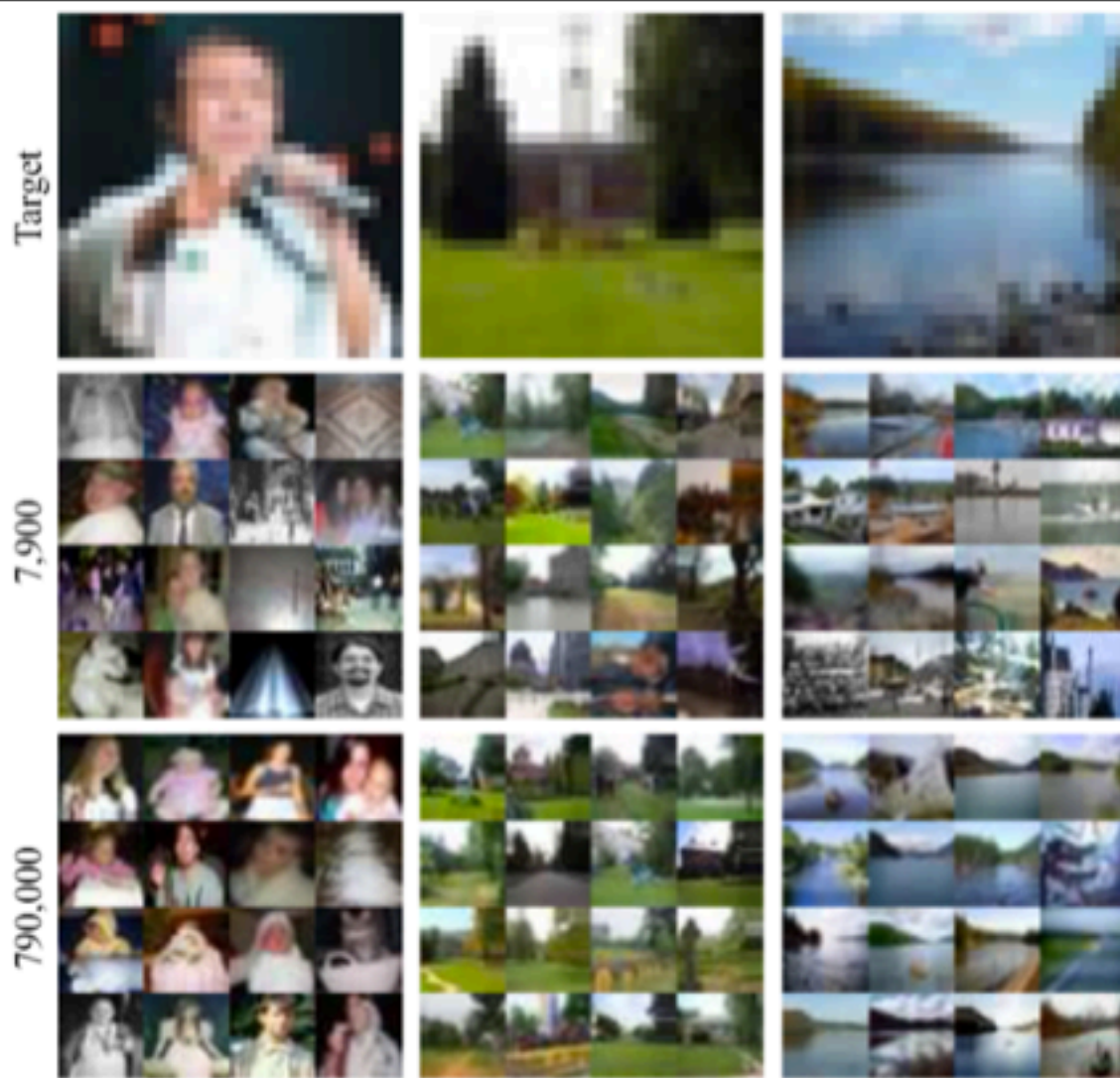
<http://groups.csail.mit.edu/vision/TinyImages/>

# Lots Of Images

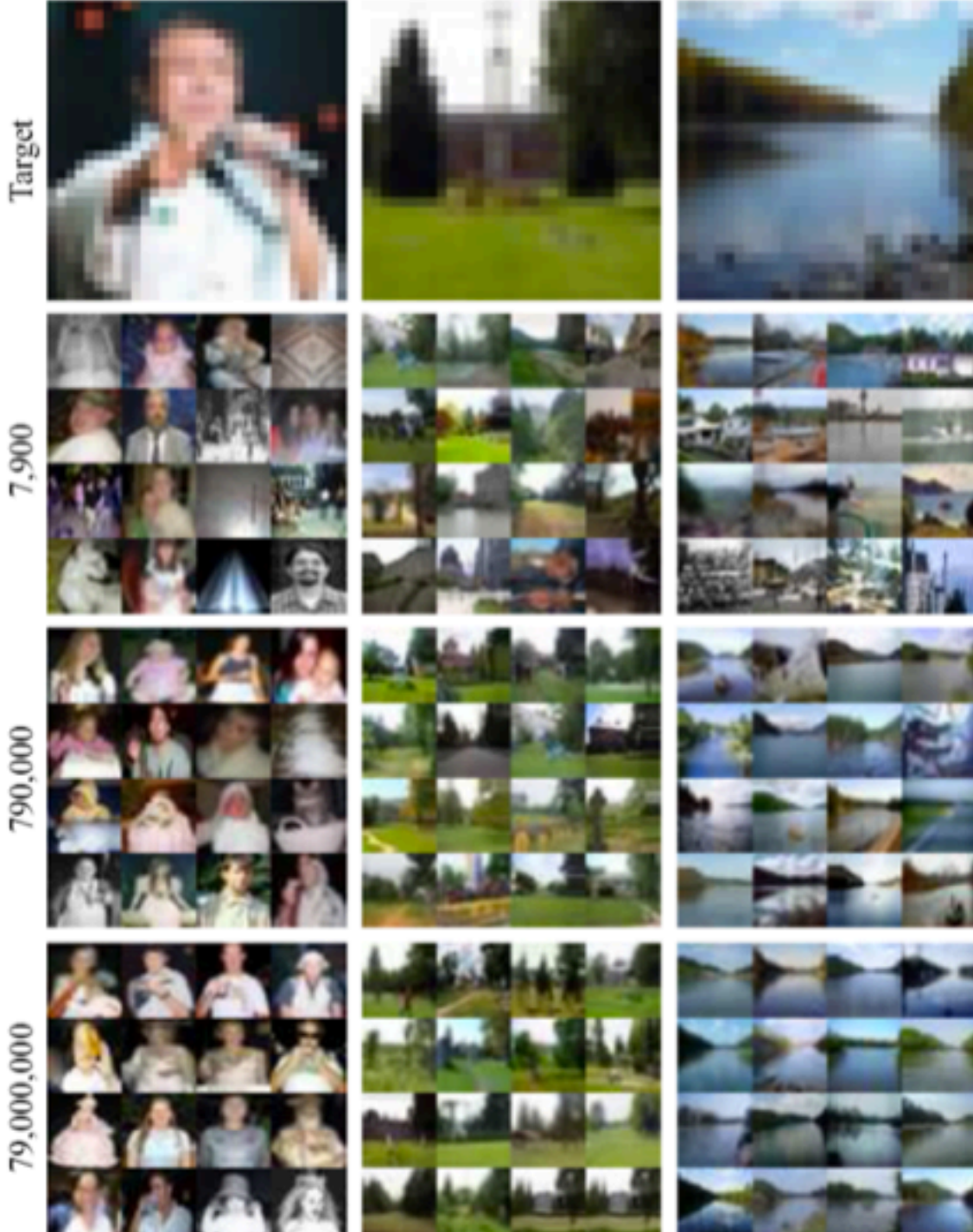




# Lots Of Images



# Lots Of Images



# How hard is it to find a matching image?

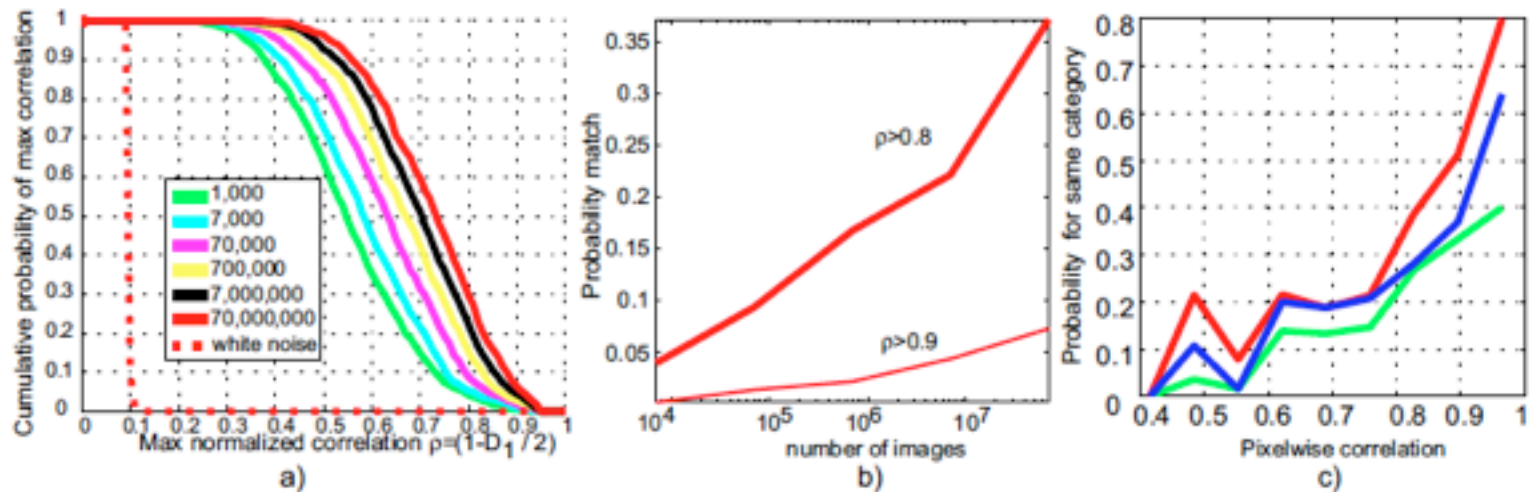


Fig. 4. Exploring the dataset using  $D_{ssd}$ . (a) probability that the nearest neighbor has a correlation greater than  $\rho$ . Each of the colored curves shows the behavior for a different size of dataset. (b) Cross-section of figure (a) plots the probability of finding a neighbor with correlation  $> 0.9$  as a function of dataset size. (c) Probability that two images belong to the same category as a function of pixel-wise correlation (duplicate images are removed). Each curve represents a different human labeler.

(c) . Three human subjects labeled pairs of images as belonging to the same visual class or not (pairs of images that correspond to duplicate images are removed). The plot shows the probability that two images are labeled as belonging to the same class as a function of image similarity. As the normalized correlation exceeds 0.8, the probability of belonging to the same class grows rapidly. Hence a simple K-nearest-neighbor approach might be effective with our size of dataset.



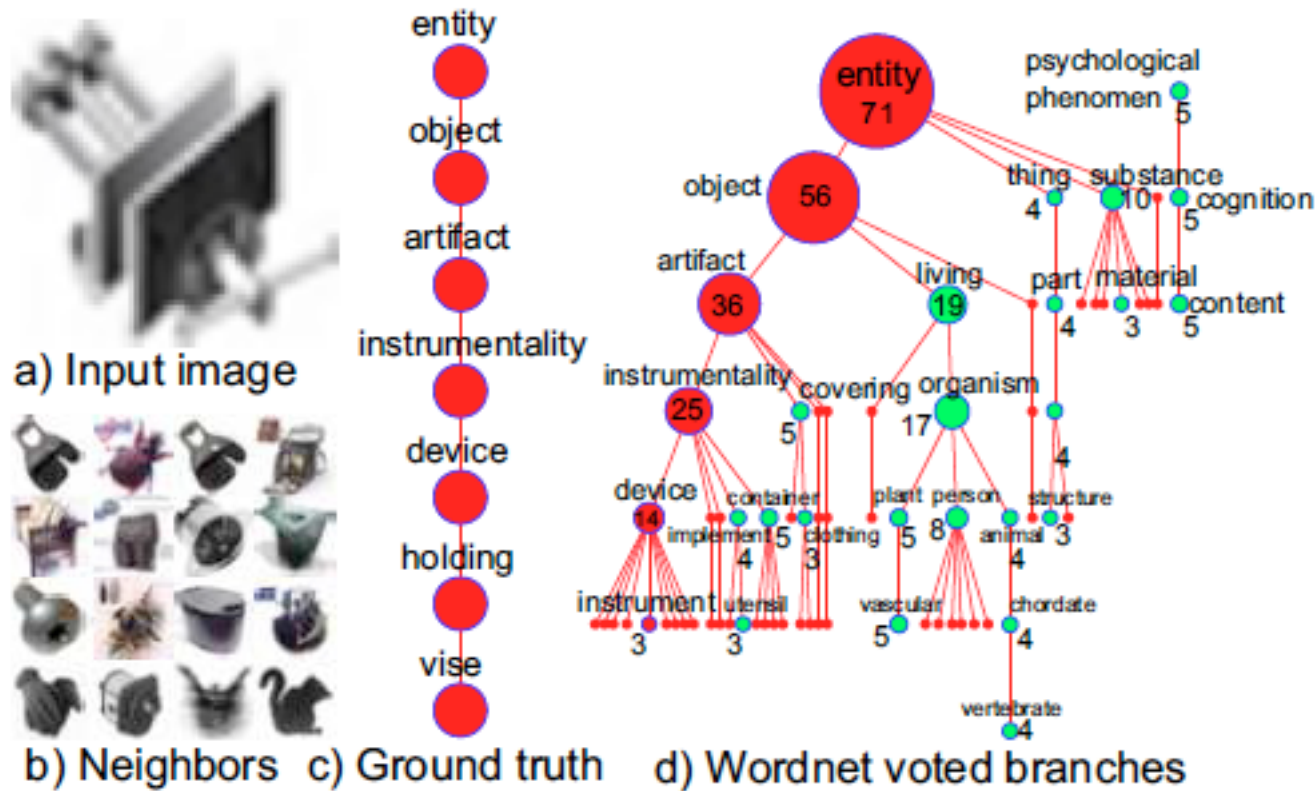


Fig. 7. This figure shows two examples. (a) Query image. (b) First 16 of 80 neighbors found using  $D_{\text{shift}}$ . (c) Ground truth Wordnet branch describing the content of the query image at multiple semantic levels. (d) Sub-tree formed by accumulating branches from all 80 neighbors. The number in each node denotes the accumulated votes. The red branch shows the nodes with the most votes. Note that this branch substantially agrees with the branch for vise and for person in the first and second examples respectively.

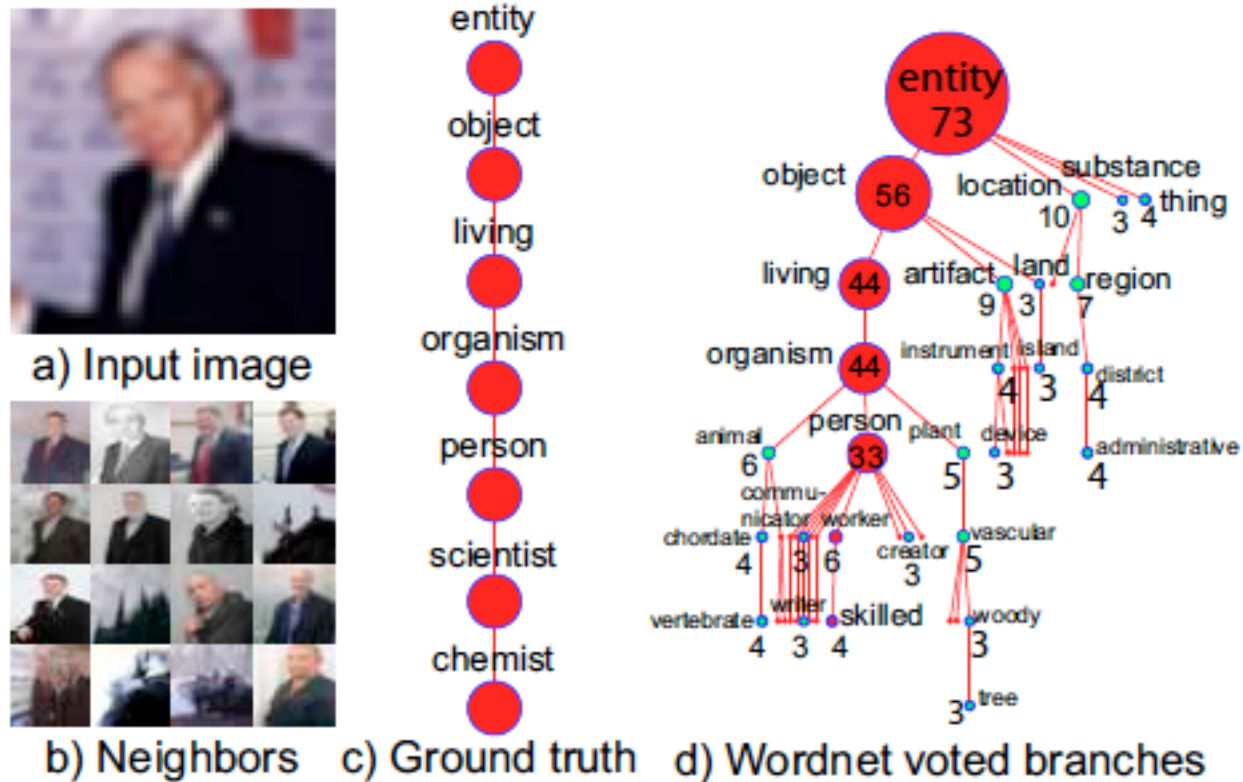


Fig. 7. This figure shows two examples. (a) Query image. (b) First 16 of 80 neighbors found using  $D_{\text{shift}}$ . (c) Ground truth Wordnet branch describing the content of the query image at multiple semantic levels. (d) Sub-tree formed by accumulating branches from all 80 neighbors. The number in each node denotes the accumulated votes. The red branch shows the nodes with the most votes. Note that this branch substantially agrees with the branch for *vise* and for *person* in the first and second examples respectively.

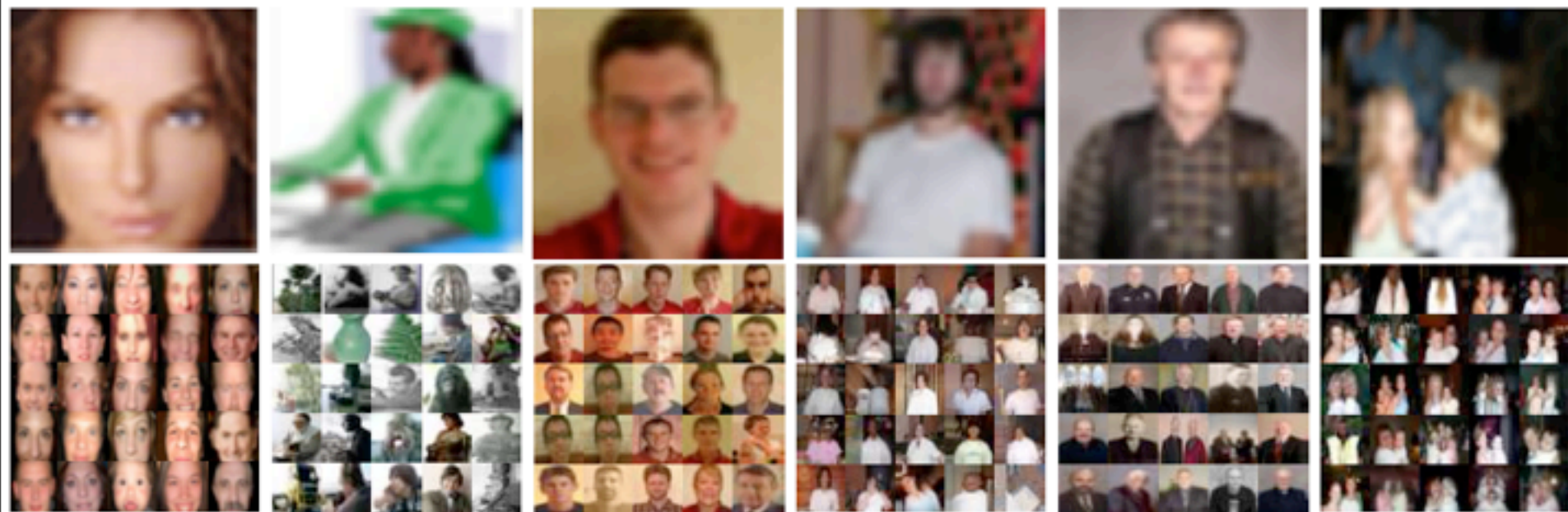


Fig. 8. Some examples of test images belonging to the “person” node of the Wordnet tree, organized according to body size. Each pair shows the query image and the 25 closest neighbors out of 79 million images using  $D_{\text{shift}}$  with  $32 \times 32$  images. Note that the sibling sets contain people in similar poses, with similar clothing to the query images.

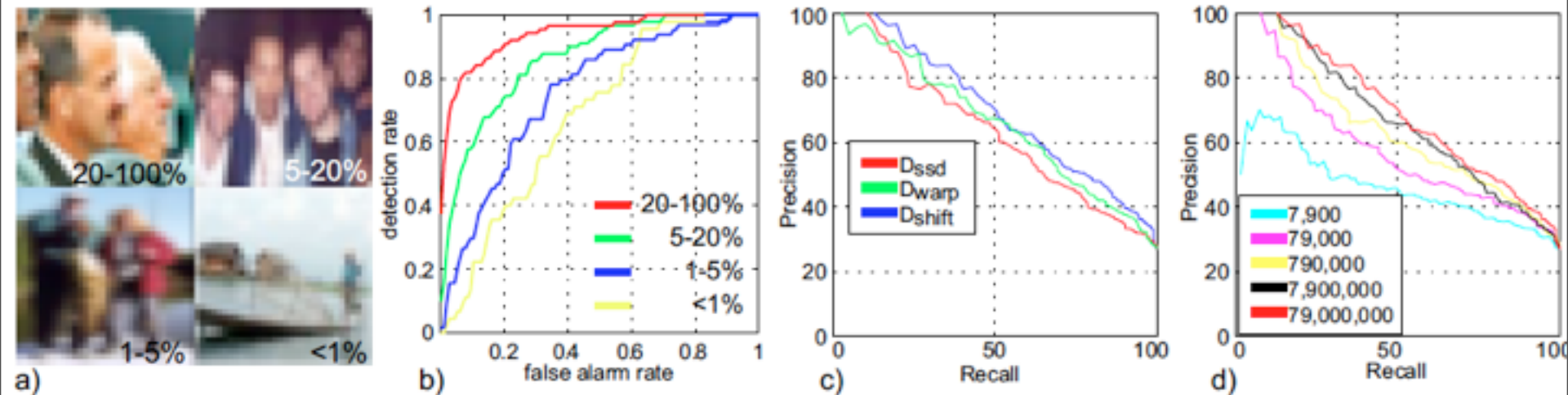


Fig. 9. (a) Examples showing the fraction of the image occupied by the head. (b)–(d): ROC curves for people detection (not localization) in images drawn randomly from the dataset of 79 million as a function of (b) head size; (c) similarity metrics and (d) dataset size using  $D_{shift}$ .



# Automatic Colorization Result

Grayscale input High resolution



Colorization of input using average



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



# Collecting labeled datasets

- ESP game (CMU)  
Luis Von Ahn and Laura Dabbish 2004
- LabelMe (MIT)  
Russell, Torralba, Freeman, 2005
- 80 Million Tiny Images  
Torralba, Fergus, Freeman 2008
- ImageNet  
Li, Fei-Fei, 2009  
<http://www.image-net.org/>
- Mechanical Turk  
Amazon

**Vegetable, veggie, veg**  
Edible seeds or roots or stems or leaves or bulbs or tubers or nonsweet fruits of any of numerous herbaceous plant

1369 pictures 73.58% Popularity Percentile WordNet EN

Numbers in brackets: (the number of synonyms in the subtree.)

ImageNet 2011 Winter Release (17...)

- animal, animate being, beast, br...
- sport, athletics (165)
- fabric, cloth, material, textile (2)
- instrumentality, instrumentation
- appliance (30)
- structure, construction (1238)
- fruit (308)
- flower (461)
- fungus (902)
- tree (992)
- vegetable, veggie, veg (175)**
  - fennel, Florence fennel, finoc...
  - cucumber, cuke (1)
  - squash (16)
  - cruciferous vegetable (18)
  - pieplant, rhubarb (0)
  - root vegetable (25)
  - solanaceous vegetable (25)
  - greens, green, leafy vegetabl...
  - potherb (0)
  - legume (37)
  - raw vegetable, rabbit food (0)
  - artichoke, globe artichoke (0)
  - artichoke heart (0)
  - asparagus (0)
  - plantain (0)
  - truffle, earthnut (0)
  - pumpkin (0)
  - mushroom (0)

Treemap Visualization Images of the Synset Downloads

ImageNet 2011 Winter Release - Vegetable, veggie, veg

# Collecting labeled datasets

- ESP game (CMU)

Luis Von Ahn and Laura Dabbish 2004

- LabelMe (MIT)

Russell, Torralba, Freeman, 2005

- 80 Million Tiny Images

Torralba, Fergus, Freeman 2008

- ImageNet

Li, Fei-Fei, 2009

- Mechanical Turk

Amazon

<https://www.mturk.com/mturk/welcome>

The screenshot shows the Amazon Mechanical Turk website. At the top, there's a navigation bar with 'Your Account', 'HTTs', and 'Qualifications'. Below that, a yellow banner reads 'Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 102,740 HTTs available. View them now.' The main content is split into two columns. The left column, 'Make Money by working on HTTs', lists benefits for workers: 'Can work from home', 'Choose your own work hours', and 'Get paid for doing good work'. It includes a flow diagram: 'Find an interesting task' (gear icon) -> 'Work' (gears icon) -> 'Earn money' (dollar sign icon). The right column, 'Get Results from Mechanical Turk Workers', lists benefits for requesters: 'Have access to a global, on-demand, 24 x 7 workforce', 'Get thousands of HTTs completed in minutes', and 'Pay only when you're satisfied with the results'. It includes a flow diagram: 'Fund your account' (dollar sign icon) -> 'Load your tasks' (gears icon) -> 'Get results' (star icon). The footer contains 'FAQ', 'Contact Us', 'Careers at Amazon', 'Developers', 'Press', 'Privacy', 'Help', and '©2009-2011 Amazon.com, Inc. or its affiliates'.

# Amazon Mechanical Turk

## Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

**102,740 HITs** available. [View them now.](#)

## Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

**As a Mechanical Turk Worker you:**

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

## Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

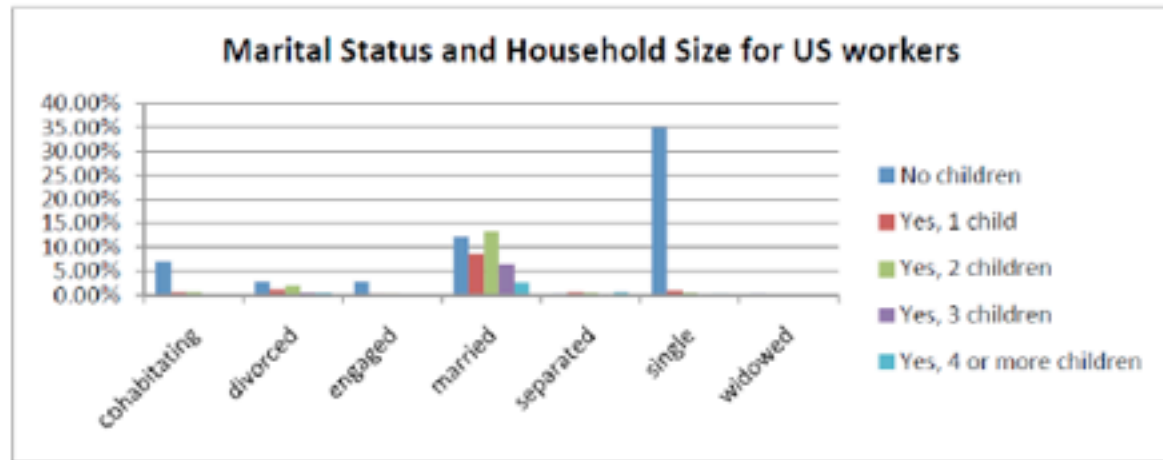
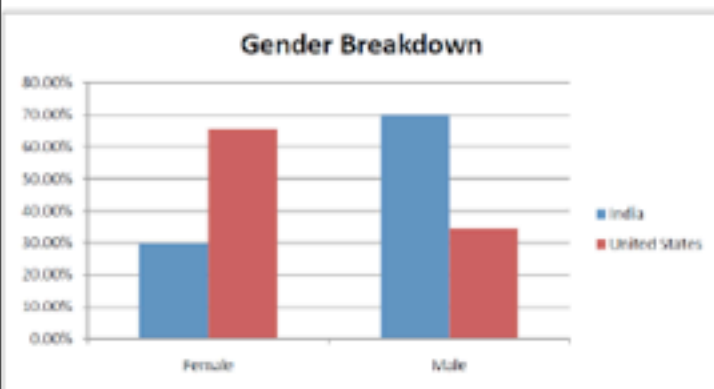
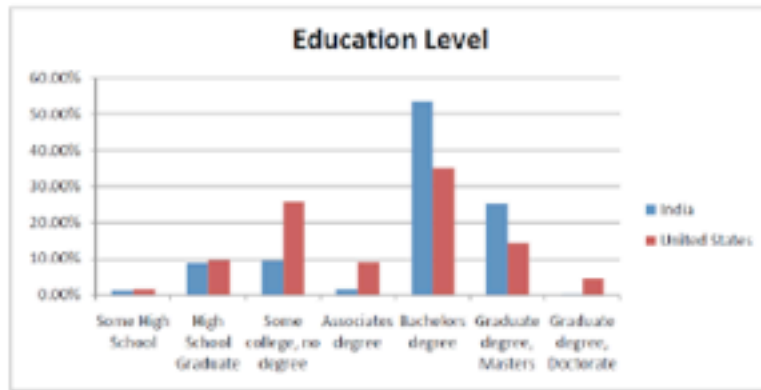
**As a Mechanical Turk Requester you:**

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



# Demography of AMT workers

United States	46.80%
India	34.00%
Miscellaneous	19.20%



Panos Ipeirotis, NYU, Feb, 2010

Slide from Fei-Fei Li, [http://www.image-net.org/papers/ImageNet\\_2010.pdf](http://www.image-net.org/papers/ImageNet_2010.pdf)

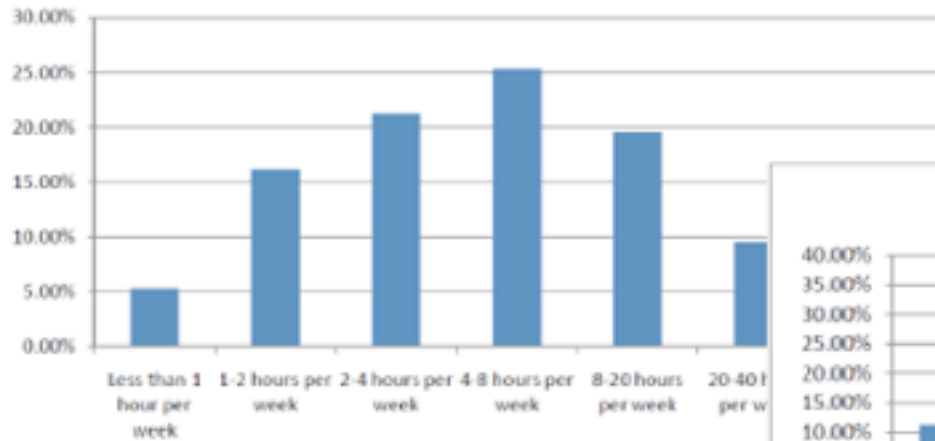
Wednesday, May 4, 2011

# Demography of AMT workers

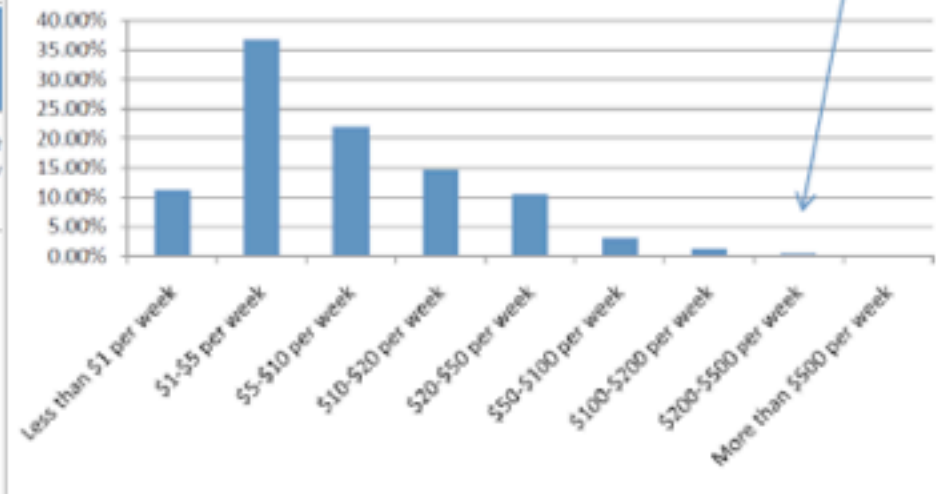


Typical Stanford Graduate student's income

Time spent on Mechanical Turk per week



Weekly Income from Mechanical Turk



Panos Ipeirotis, NYU, Feb, 2010

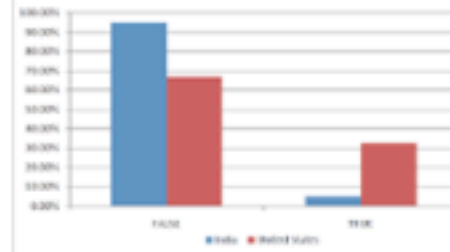
33

# Demography of AMT workers

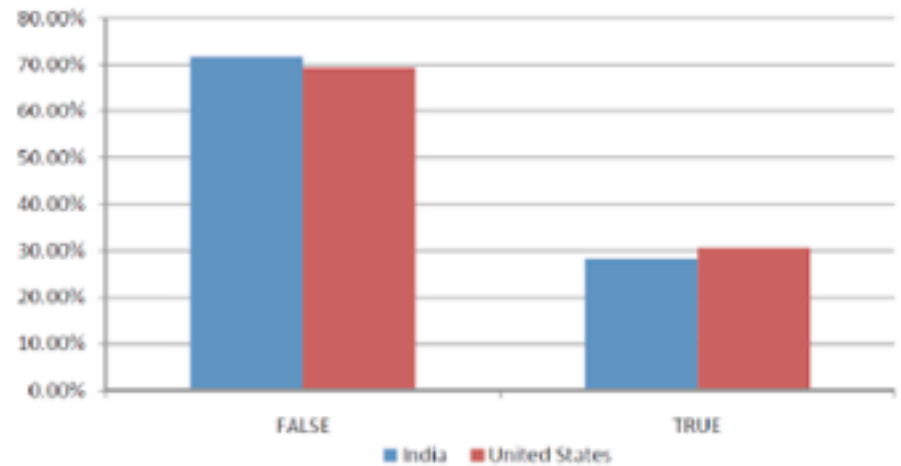
Mechanical Turk is a fruitful way to spend free time and get some cash (e.g., instead of watching TV)



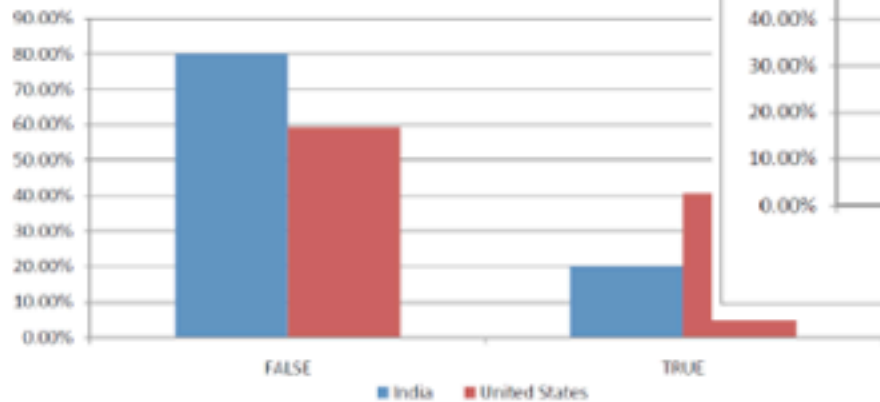
I participate on Mechanical Turk to kill time



I am currently unemployed or only have a part-time job



I participate on Mechanical Turk because the tasks are



Panos Ipeirotis, NYU, Feb, 2010

# things to do with images on an internet scale

- Object recognition
  - 80 million tiny images
- Image editing/completion
  - Hayes and Efros
  - Infinite images



# Scene Completion Using Millions of Photographs



SIGGRAPH2007



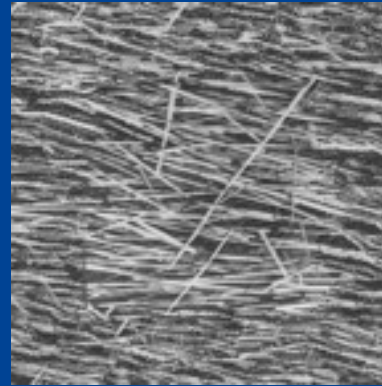
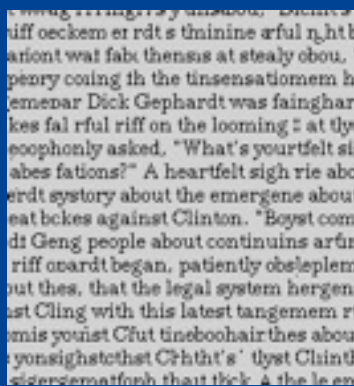
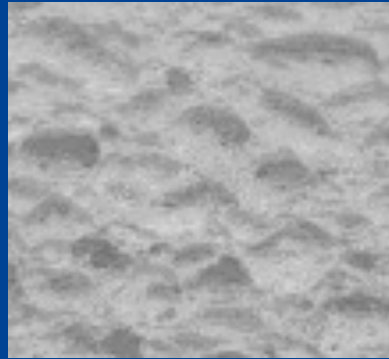
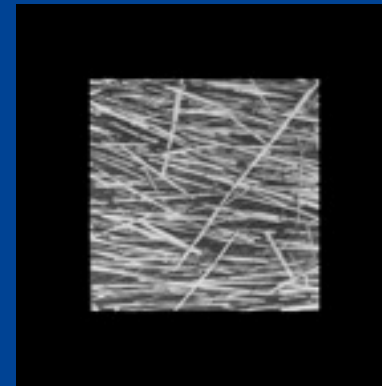
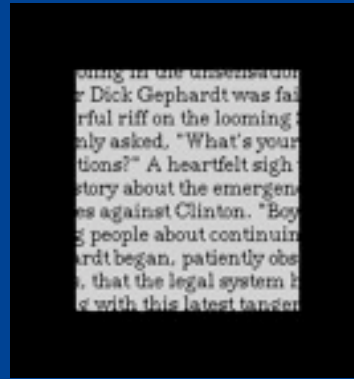
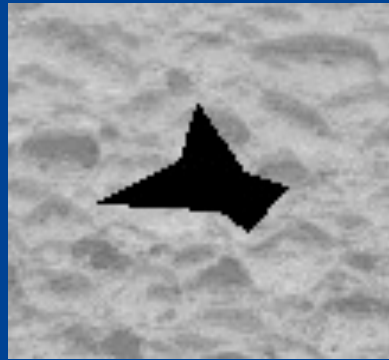
James Hays and Alexei A. Efros  
Carnegie Mellon University











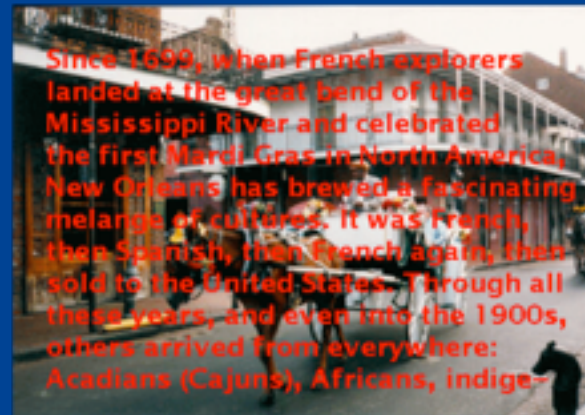
Efros and Leung. Texture synthesis by non-parametric sampling. ICCV 1999.



Efros and Leung result

Hays and Efros, SIGGRAPH 2007



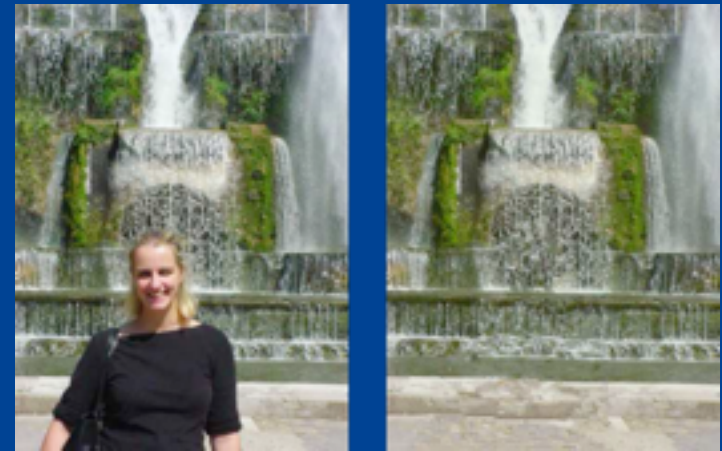
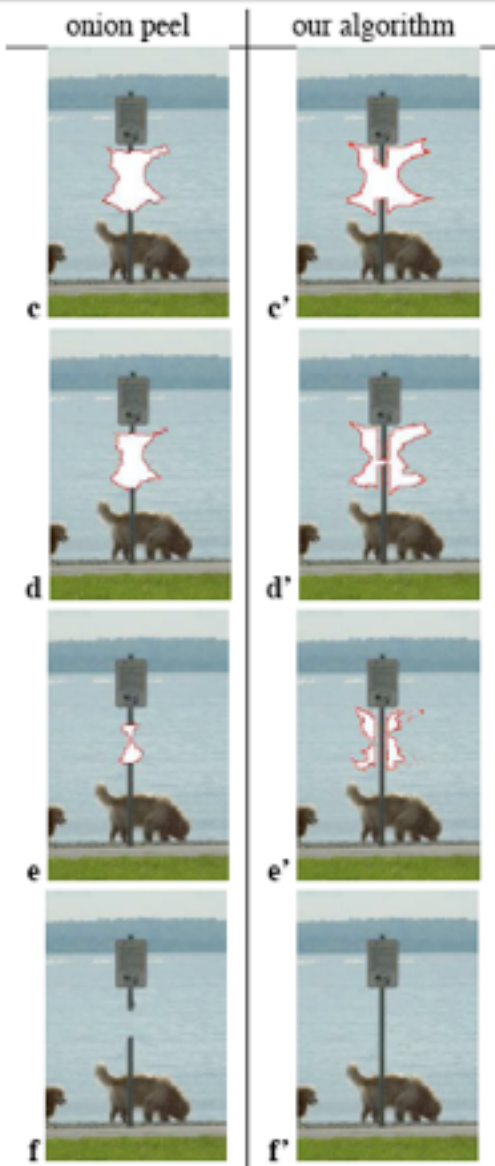


Bertalmio, Sapiro, Caselles, and Ballester.  
Image Inpainting. SIGGRAPH 2000.



## Diffusion Result

Hays and Efros, SIGGRAPH 2007



Criminisi, Perez, and Toyama.  
 Region filling and object  
 removal by exemplar-based  
 inpainting. IEEE Transactions  
 on Image Processing. 2004.

Fig. 11. Onion peel vs. structure-guided filling. (a) Original image. (b) The target region has been selected and marked with a red boundary. (c,d,e,f) Results of filling by concentric layers. (c',d',e',f') Results of filling with our algorithm. Thanks to the *data term* in (1) the sign pole is reconstructed correctly by our algorithm.

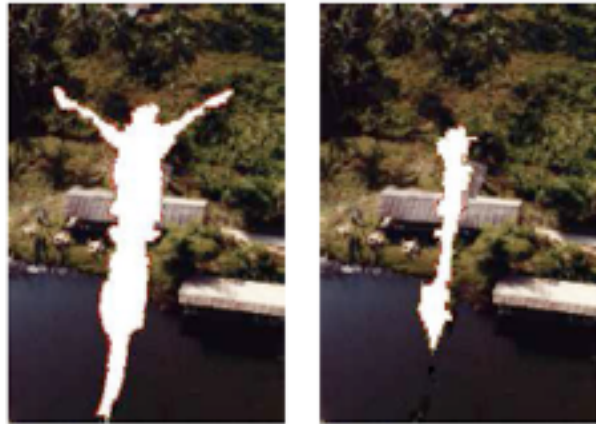


Criminisi, Perez, and Toyama. Region filling and object removal by exemplar-based inpainting. IEEE Transactions on Image Processing. 2004.



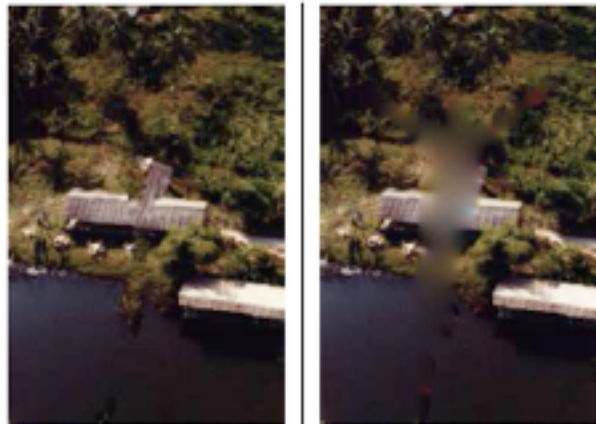
a

b



c

d



e

f

a: original image

b: edited region

c, d: different stages of the filling process.

e: Criminisi et al result

f: diffusion-based in-filling result.





## Criminisi et al. result

Hays and Efros, SIGGRAPH 2007



Criminisi et al. result



## Microsoft Digital Image Pro Smart erase result

Hays and Efos, SIGGRAPH 2007



Jian Sun, Lu Yuan, Jiaya Jia and Heung-Yeung Shum.  
Image Completion with Structure Propagation. SIGGRAPH 2005

Hays and Efros, SIGGRAPH 2007





Jian Sun, Lu Yuan, Jiaya Jia and Heung-Yeung Shum.  
Image Completion with Structure Propagation. SIGGRAPH 2005

Hays and Efros, SIGGRAPH 2007



# Scene Matching for Image Completion



Hays and Efros, SIGGRAPH 2007

# Scene Matching for Image Completion



Hays and Efros, SIGGRAPH 2007





Change **Alley** Aerial Plaza with its The Printer's **Alley** sign looking ...  
...  
300 x 400 - 21k  
[en.wikipedia.org](http://en.wikipedia.org)



Looking west past Printers **Alley**.  
679 x 450 - 469k - jpg  
[franklin.thefuntimesguide.com](http://franklin.thefuntimesguide.com)



More Bubble Gum **Alley** photos  
can be ...  
679 x 450 - 464k - jpg  
[franklin.thefuntimesguide.com](http://franklin.thefuntimesguide.com)



Gasoline **Alley** gang  
692 x 430 - 177k - jpg  
[newcritics.com](http://newcritics.com)



Change **Alley** : interior  
550 x 413 - 98k  
[infopedia.nlb.gov.sg](http://infopedia.nlb.gov.sg)



Earl G. **Alley** ...  
321 x 383 - 19k - jpg  
[www.msstate.edu](http://www.msstate.edu)



Gun **Alley** 8.5x11 Full Color Ink  
Wash ...  
390 x 301 - 14k - jpg  
[www.rorschachentertainment.com](http://www.rorschachentertainment.com)



Grace Court **Alley**  
732 x 549 - 98k - jpg  
[www.bridgeandtunnelclub.com](http://www.bridgeandtunnelclub.com)



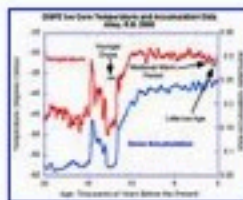
2007 **Alley** Loop Sponsors  
300 x 453 - 51k - jpg  
[www.cbnordic.org](http://www.cbnordic.org)



panoramic photo of Alligator **Alley**  
4902 x 460 - 1048k - jpg  
[sflwww.er.usgs.gov](http://sflwww.er.usgs.gov)



Richard B. **Alley**  
450 x 361 - 29k - gif  
[www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)



Also, Chicken **Alley** is reported to  
...  
450 x 337 - 82k  
[phidoux.typepad.com](http://phidoux.typepad.com)



Ego **Alley**  
500 x 375 - 48k - jpg  
[dc.about.com](http://dc.about.com)



Wednesday, May 4, 2011







Hays and Efos, SIGGRAPH 2007

Wednesday, May 4, 2011



## Scene Completion Result

Hays and Efros, SIGGRAPH 2007



# The Algorithm

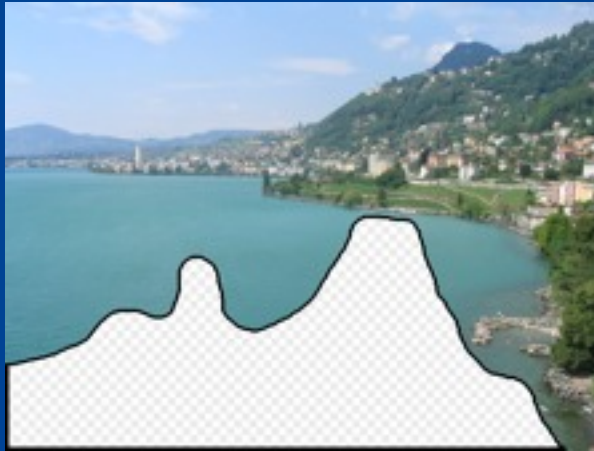


# The Algorithm

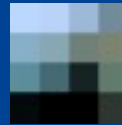
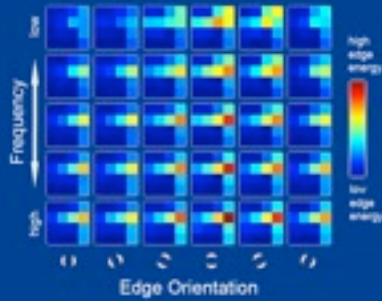


**Input image**

# The Algorithm

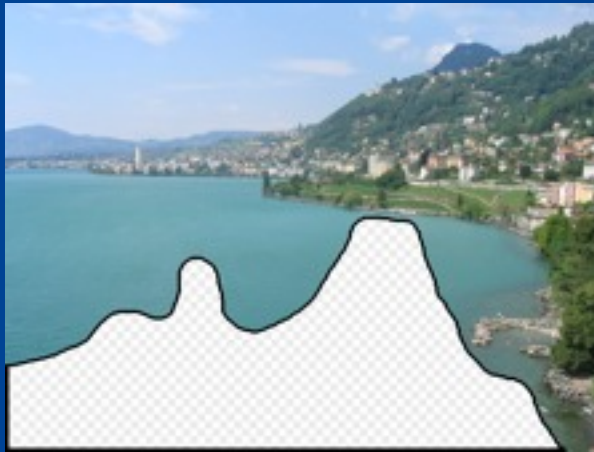


Input image

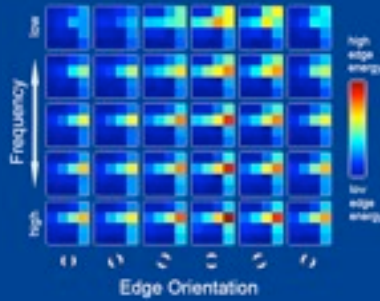


Scene Descriptor

# The Algorithm



Input image



Scene Descriptor

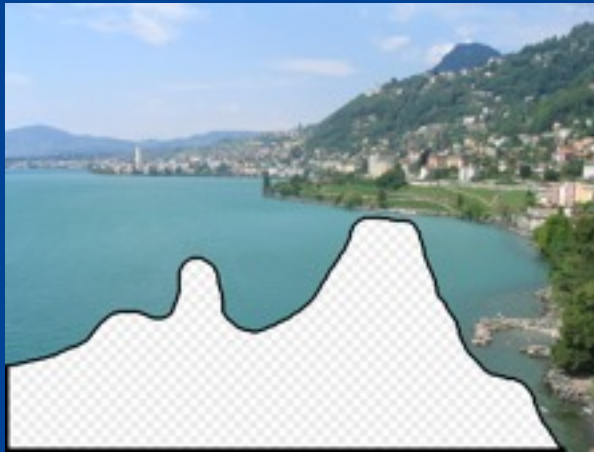


Image Collection

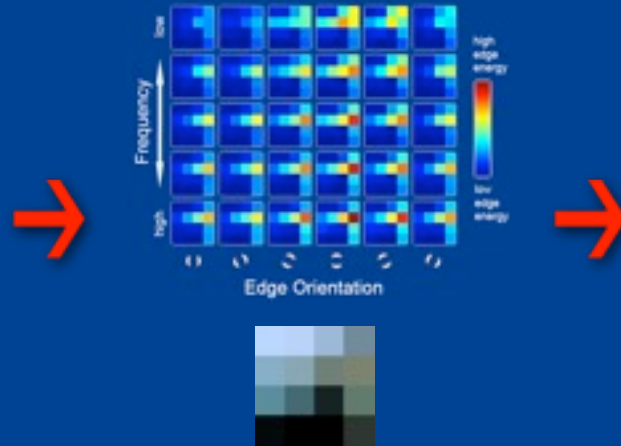
...



# The Algorithm



Input image



Scene Descriptor



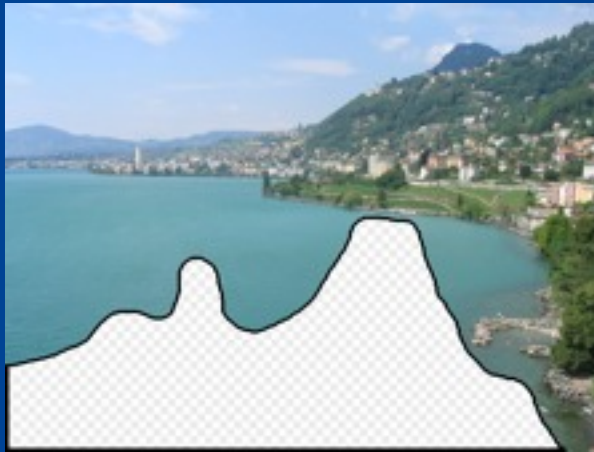
Image Collection



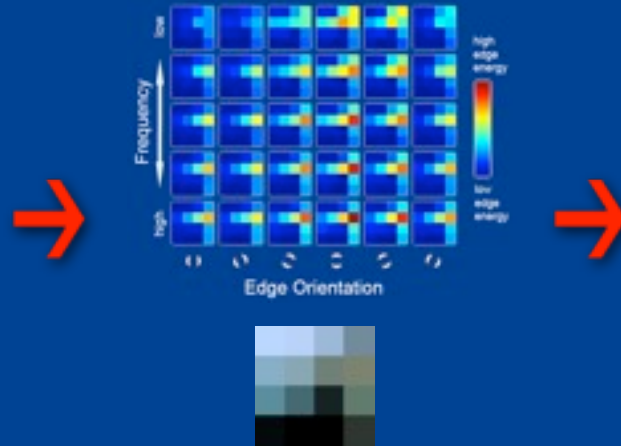
200 matches

Hays and Efros, SIGGRAPH 2007

# The Algorithm



Input image



Scene Descriptor



Image Collection



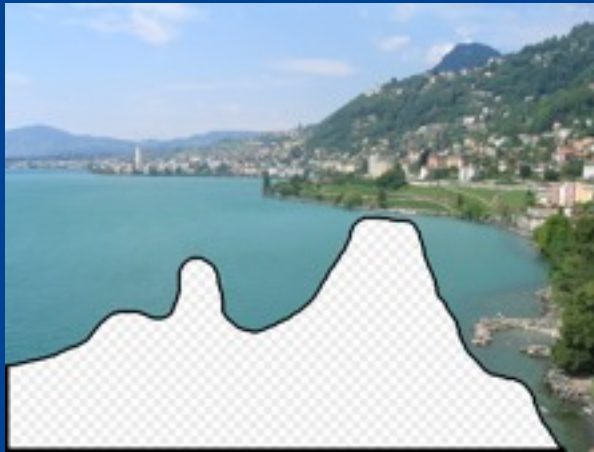
Context matching  
+ blending



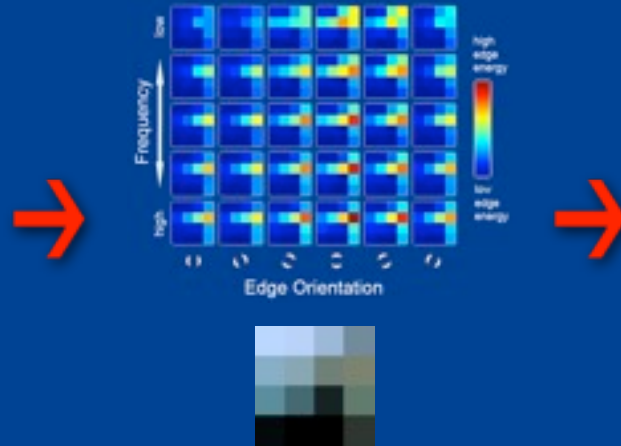
200 matches

Hays and Efros, SIGGRAPH 2007

# The Algorithm



Input image



Scene Descriptor



Image Collection



Context matching  
+ blending

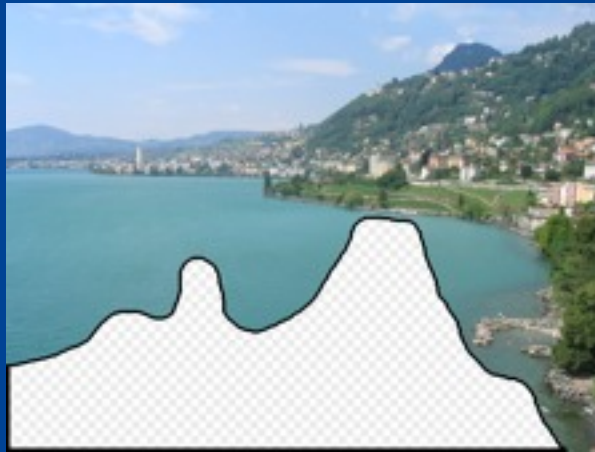


200 matches

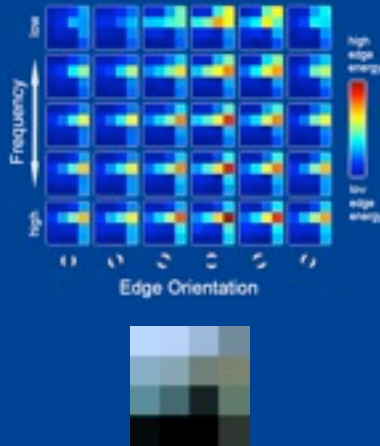
Hays and Efros, SIGGRAPH 2007



# The Algorithm



Input image



Scene Descriptor



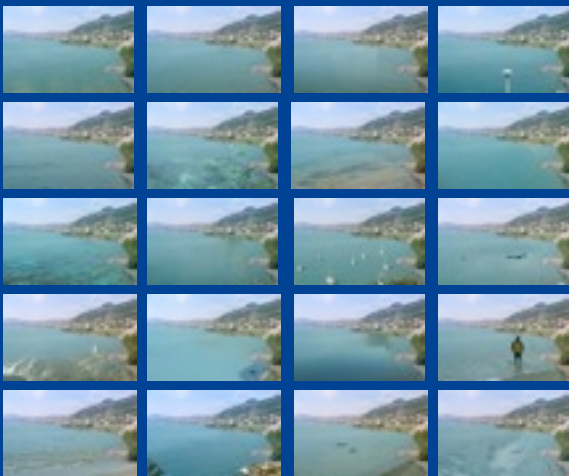
Image Collection



200 matches



Context matching  
+ blending



20 completions



# Data

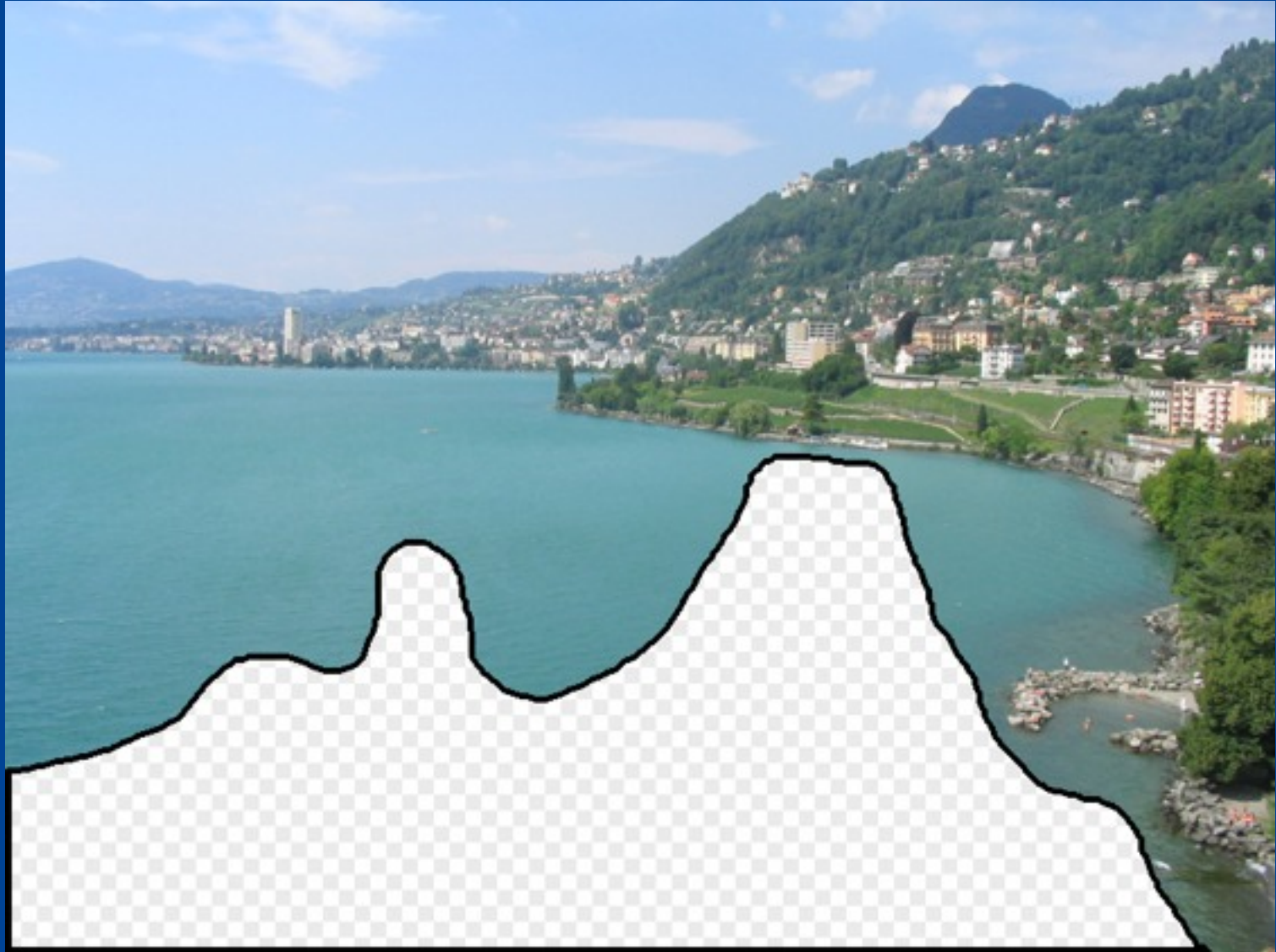
We downloaded 2.3 Million unique images from Flickr groups and keyword searches.

# Data

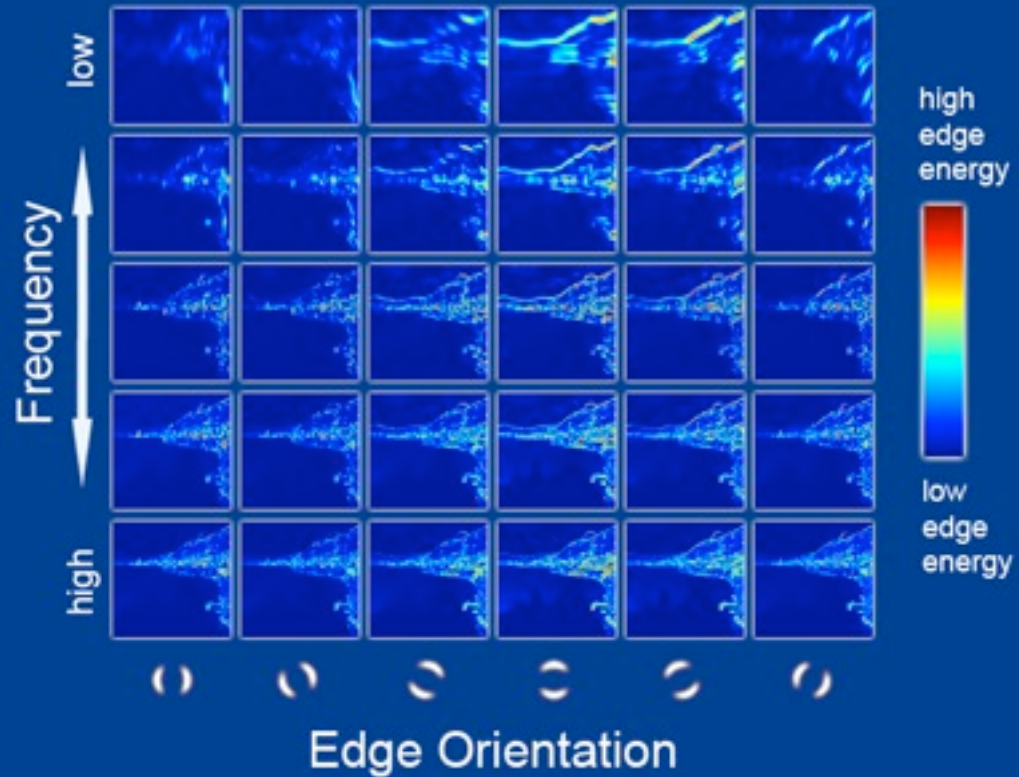
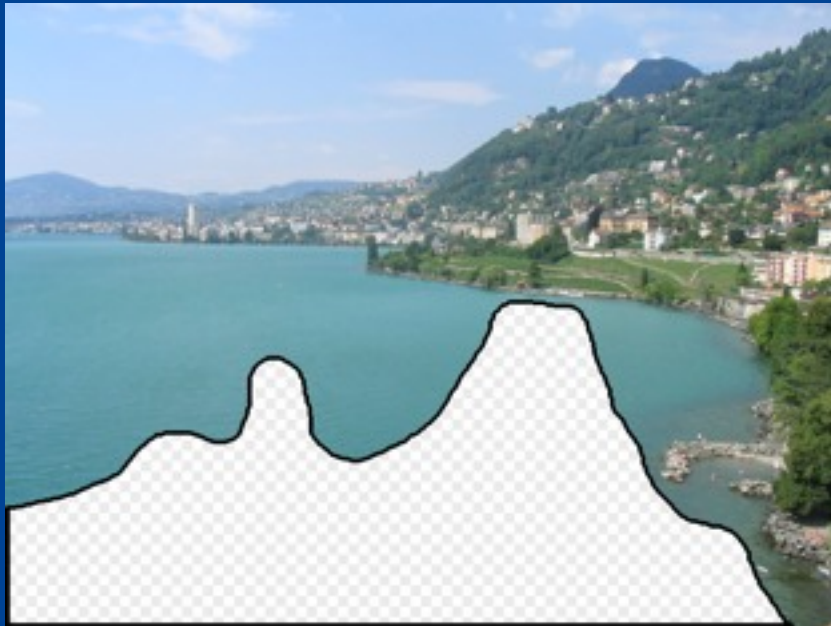
We downloaded **2.3 Million** unique images from Flickr groups and keyword searches.



# Scene Matching

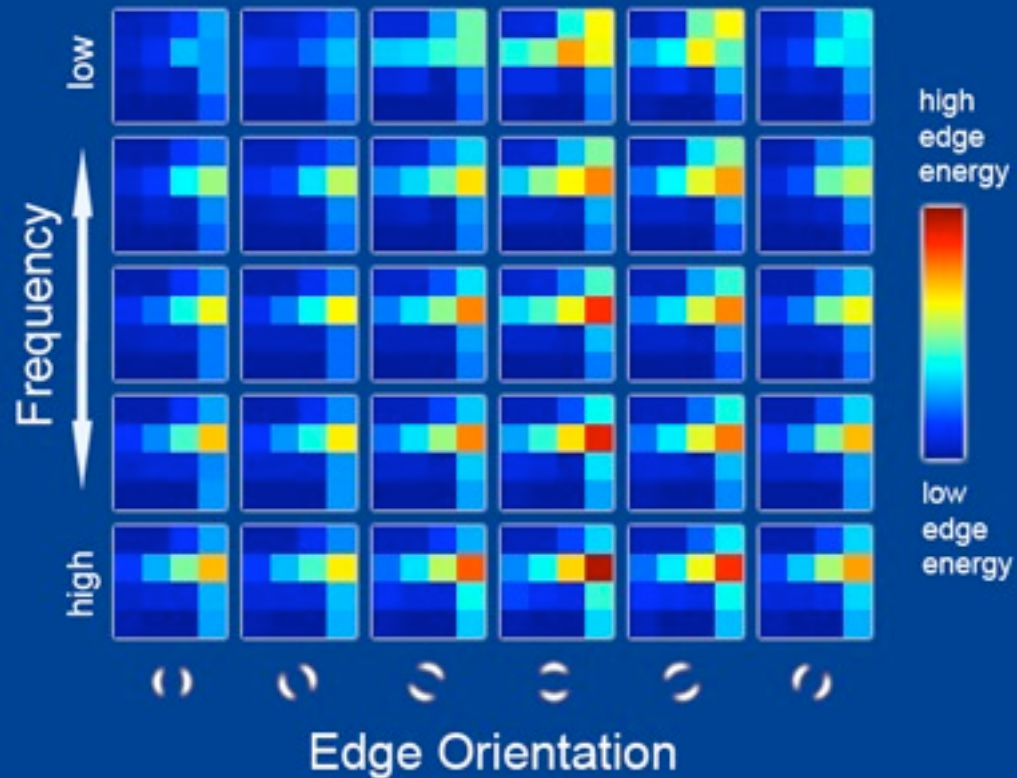
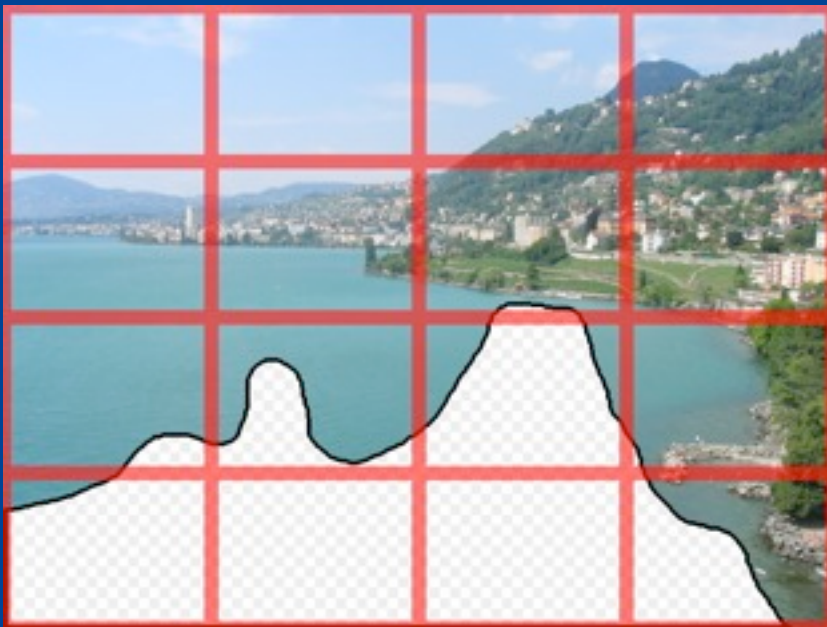


# Scene Descriptor





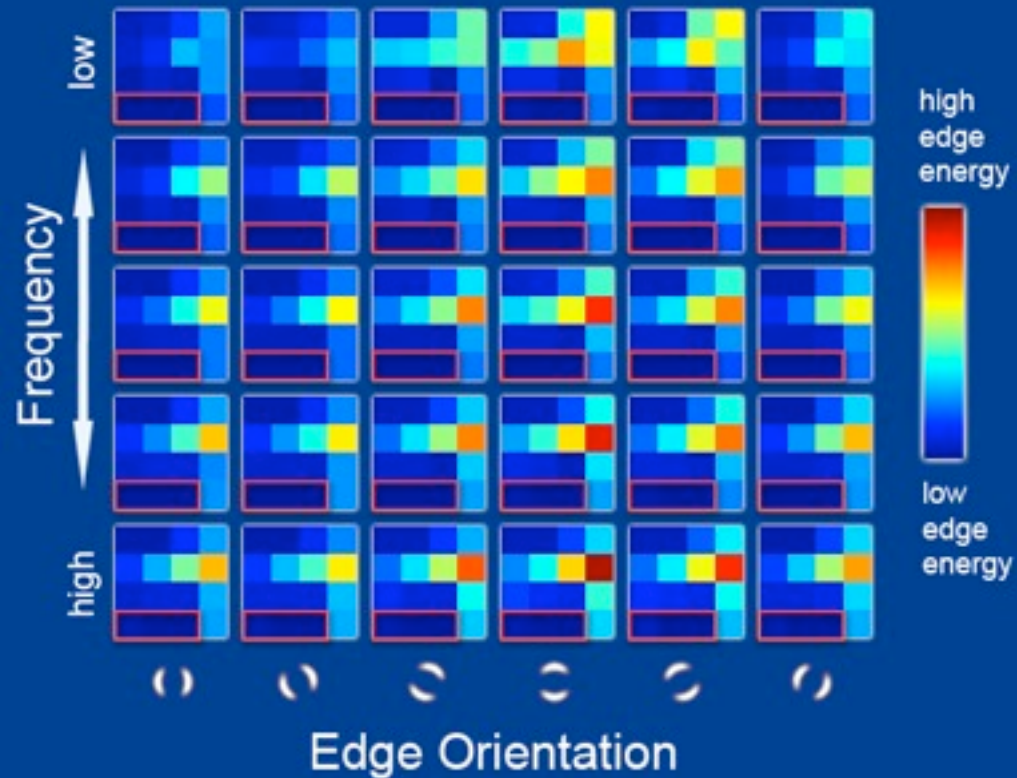
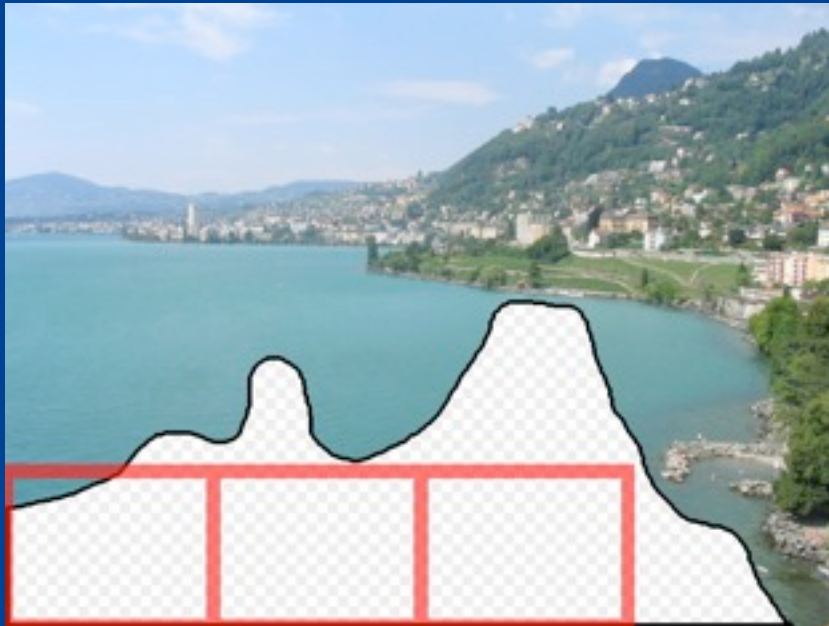
# Scene Descriptor



Gist scene descriptor  
(Oliva and Torralba 2001)

Hays and Efros, SIGGRAPH 2007

# Scene Descriptor



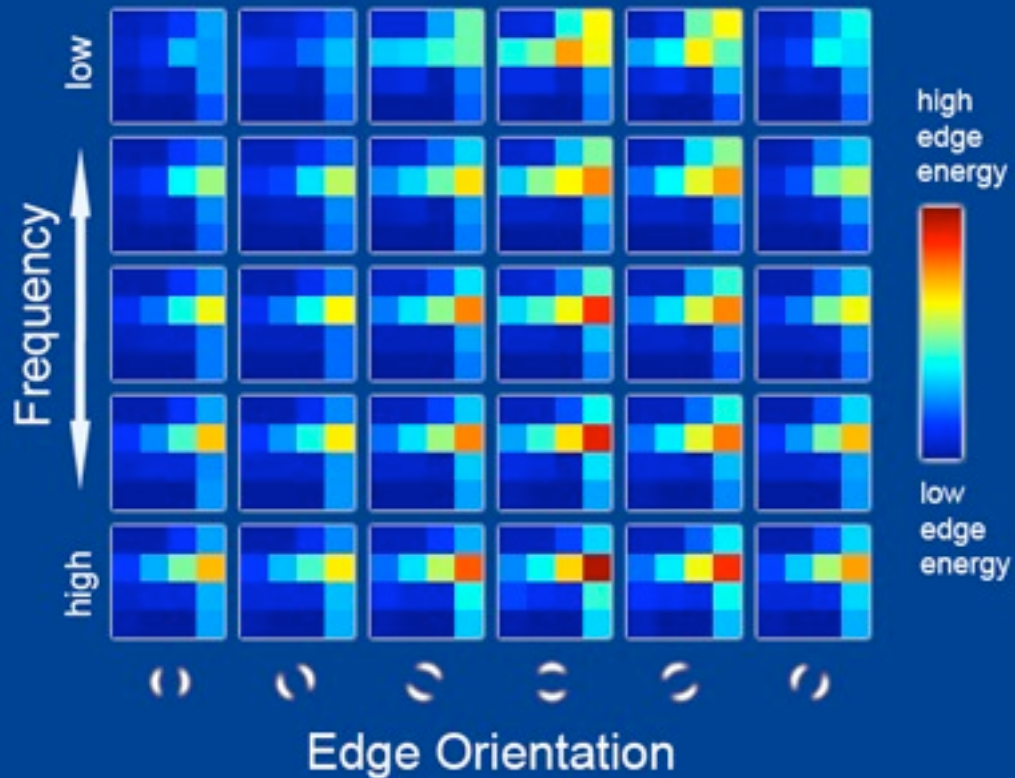
Gist scene descriptor  
(Oliva and Torralba 2001)

Hays and Efros, SIGGRAPH 2007

# Scene Descriptor



+



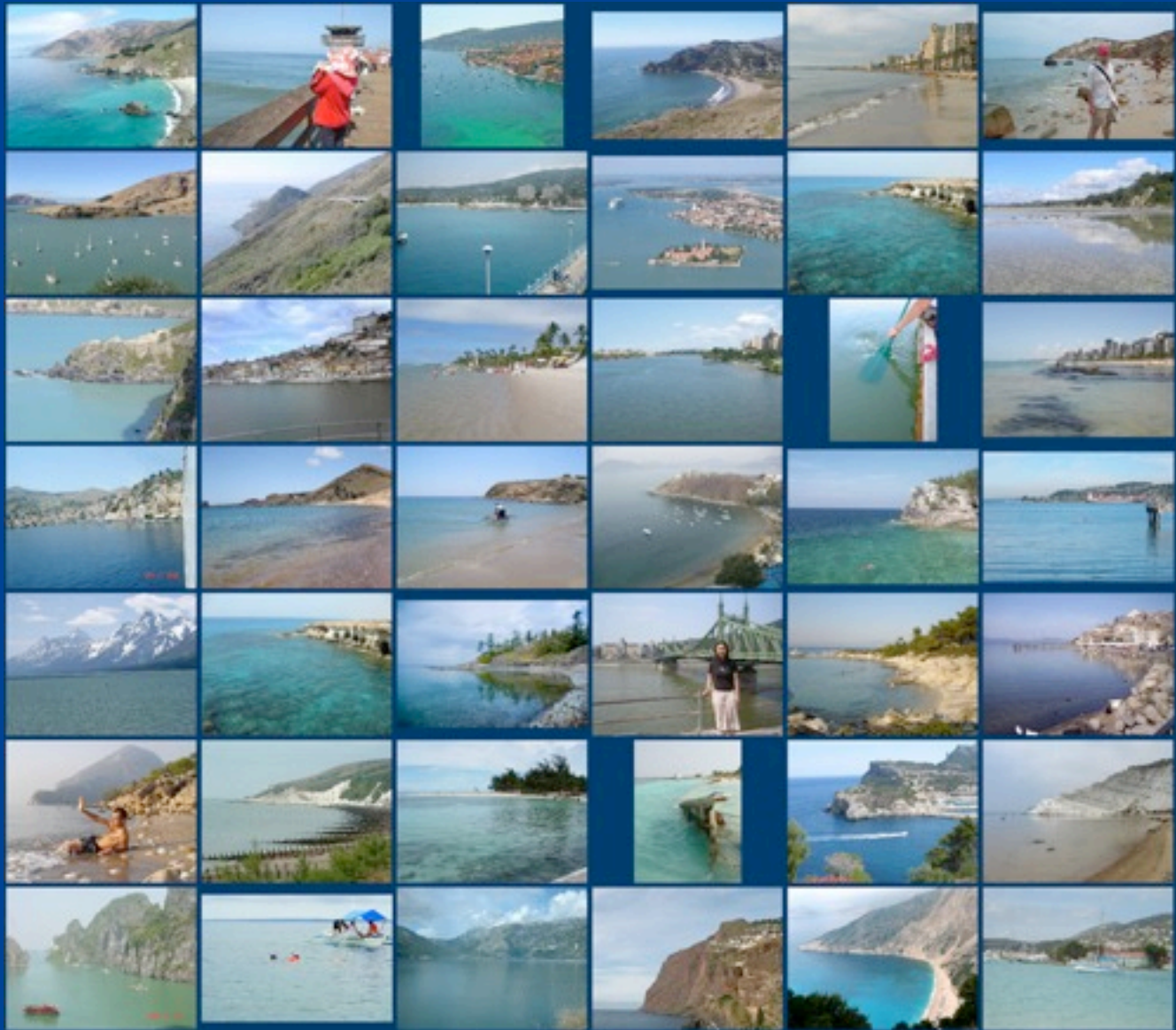
Gist scene descriptor  
(Oliva and Torralba 2001)

Hays and Efros, SIGGRAPH 2007



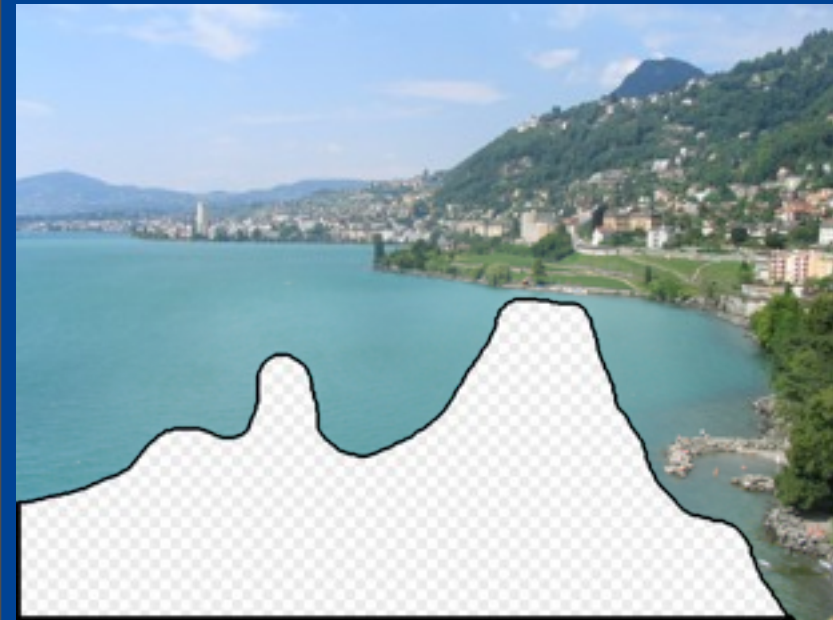




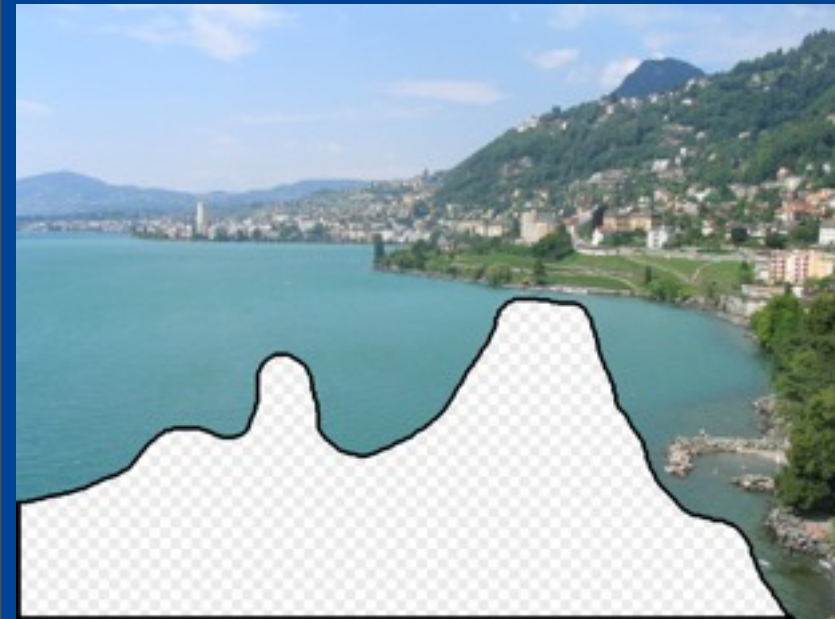


... 200 total

# Context Matching



# Context Matching











# Graph cut + Poisson blending

Hays and Efros, SIGGRAPH 2007



# Graph cut + Poisson blending

Hays and Efros, SIGGRAPH 2007

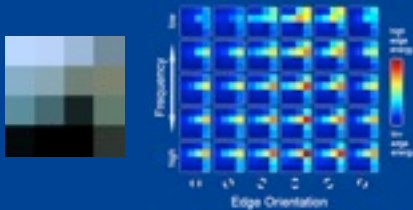
# Result Ranking

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



# Result Ranking

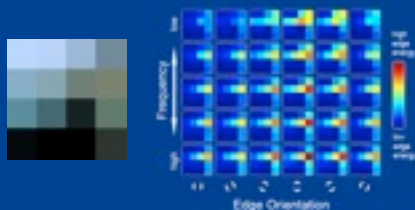
We assign each of the 200 results a score which is the sum of:



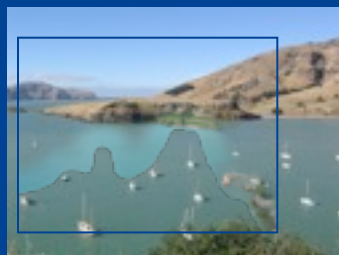
The scene matching distance

# Result Ranking

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



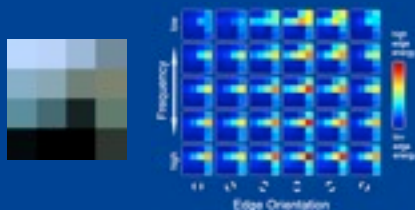
The scene matching distance



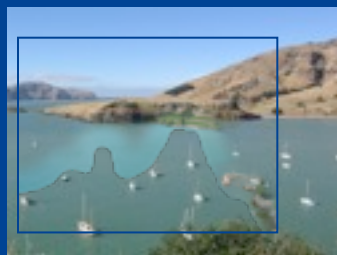
The context matching distance  
(color + texture)

# Result Ranking

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



The scene matching distance



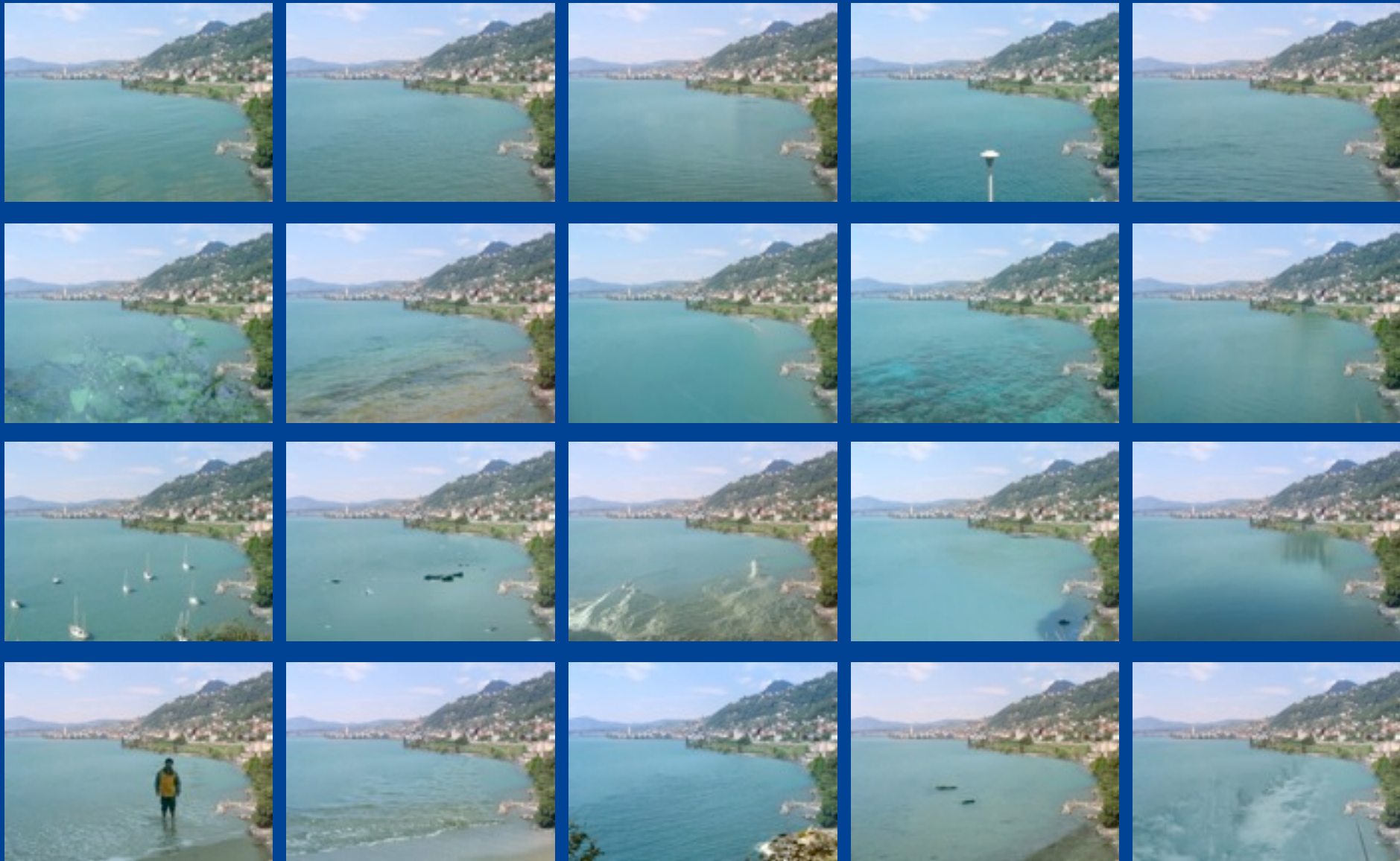
The context matching distance  
(color + texture)



The graph cut cost



# Top 20 Results















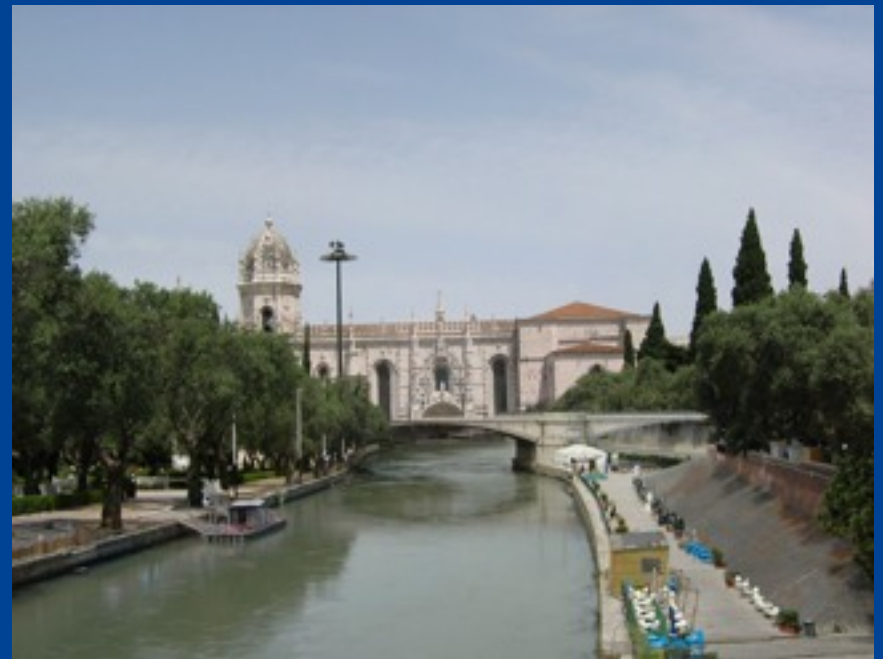
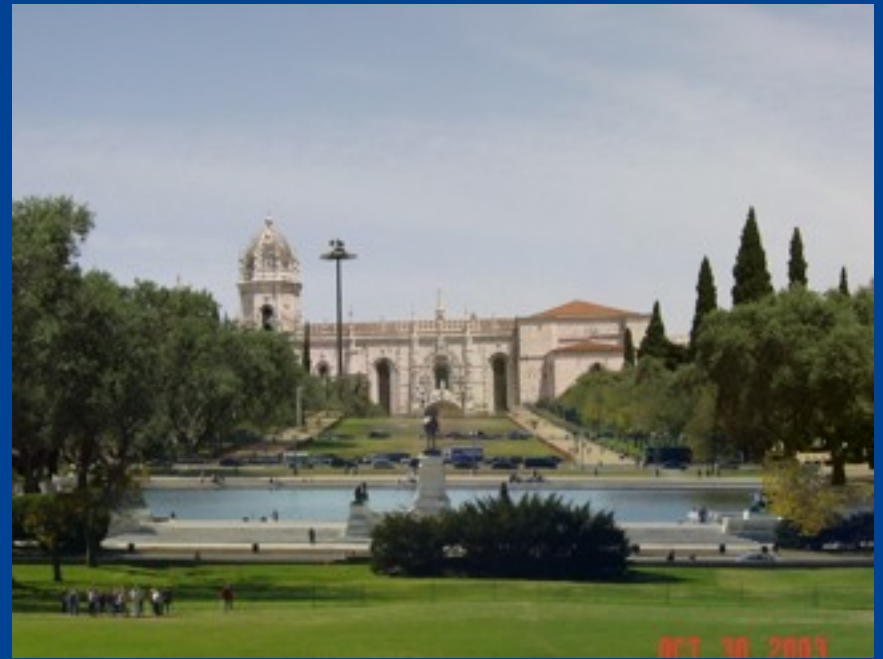












Hays and Efron, SIGGRAPH 2007

Wednesday, May 4, 2011



Hays and Efron, SIGGRAPH 2007

Wednesday, May 4, 2011





Hays and Efros, SIGGRAPH 2007

Wednesday, May 4, 2011





Hays and Efron, SIGGRAPH 2007

Wednesday, May 4, 2011



... 200 scene matches

Hays and Efros, SIGGRAPH 2007





... 200 scene matches

Hays and Efros, SIGGRAPH 2007







Hays and Efros, SIGGRAPH 2007

Wednesday, May 4, 2011



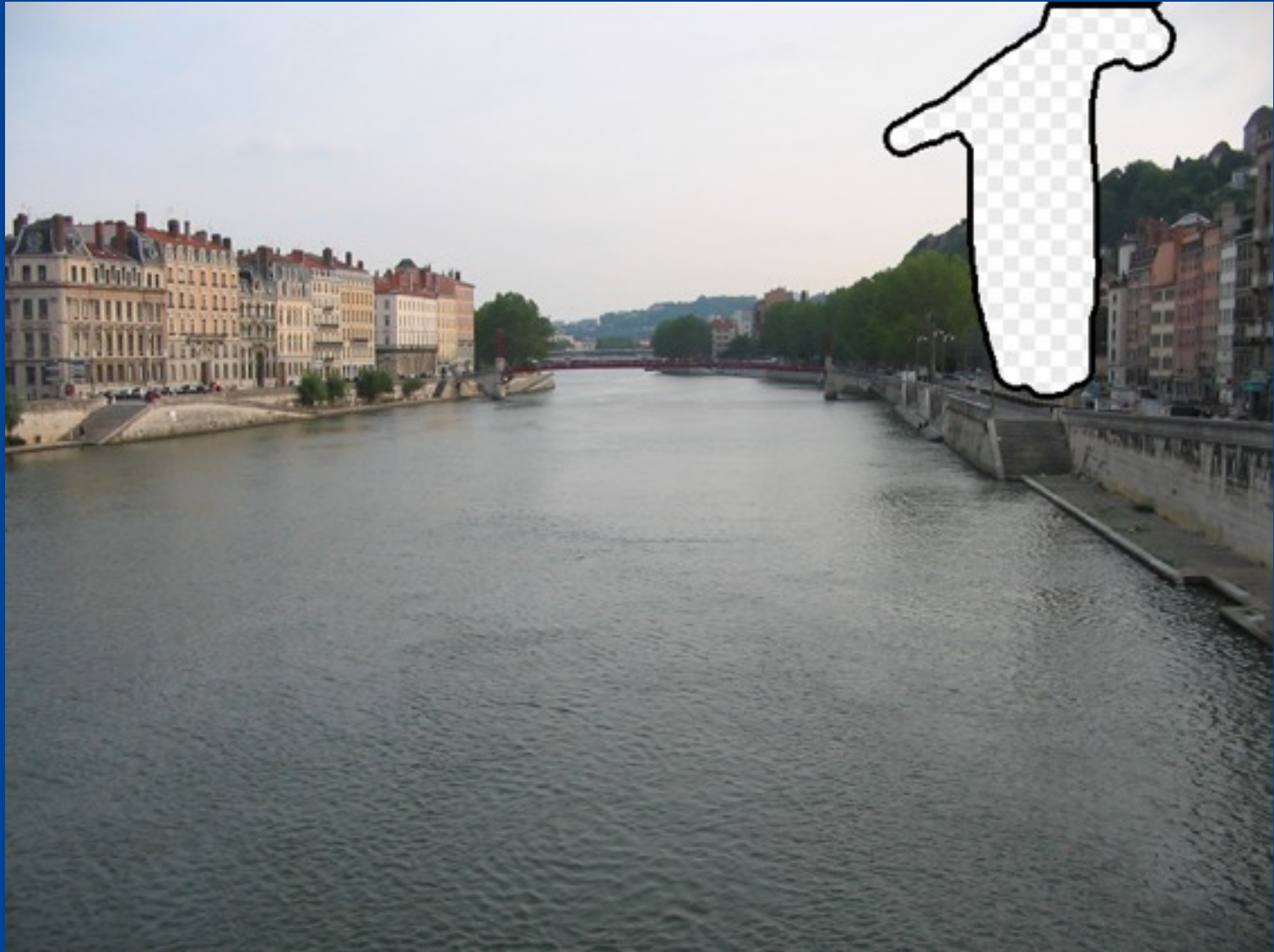


Hays and Efros, SIGGRAPH 2007

Wednesday, May 4, 2011

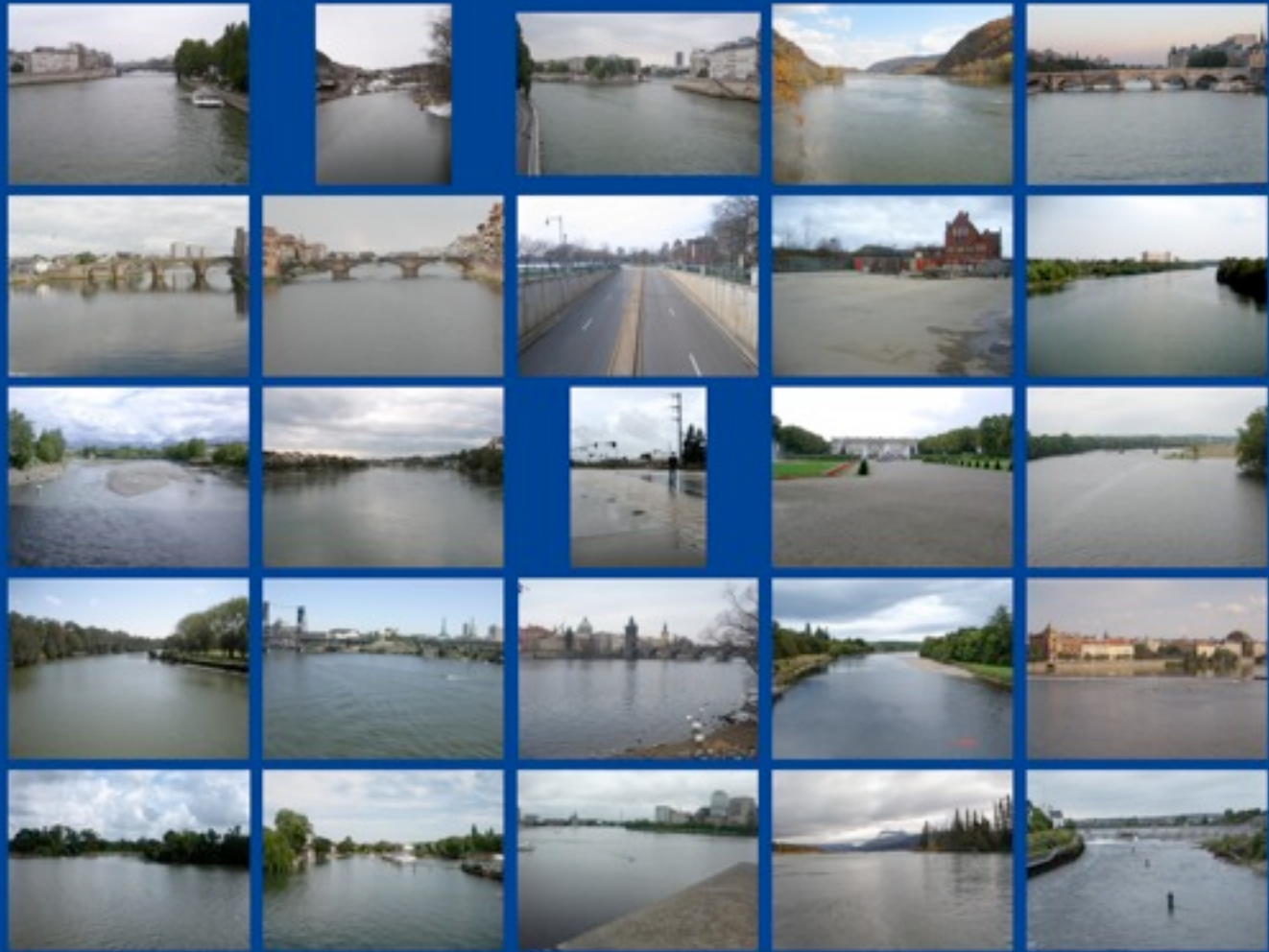
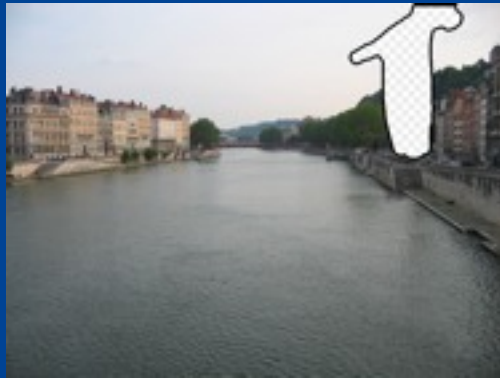




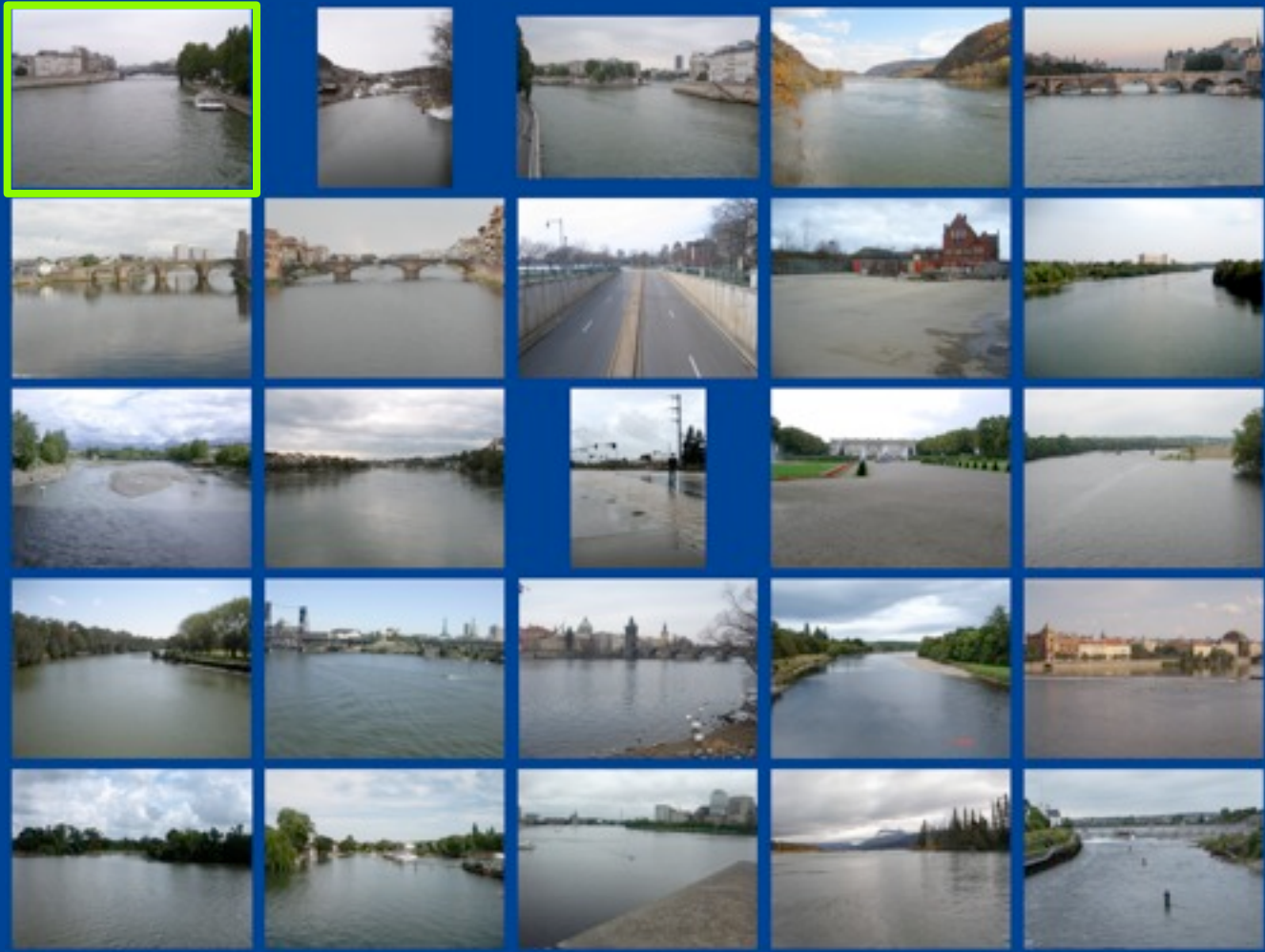




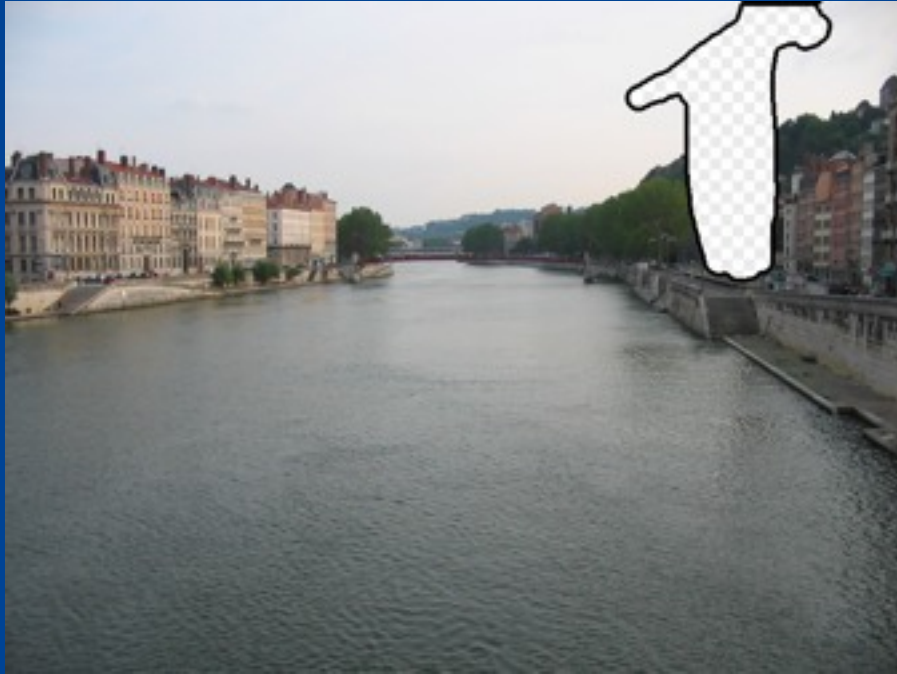




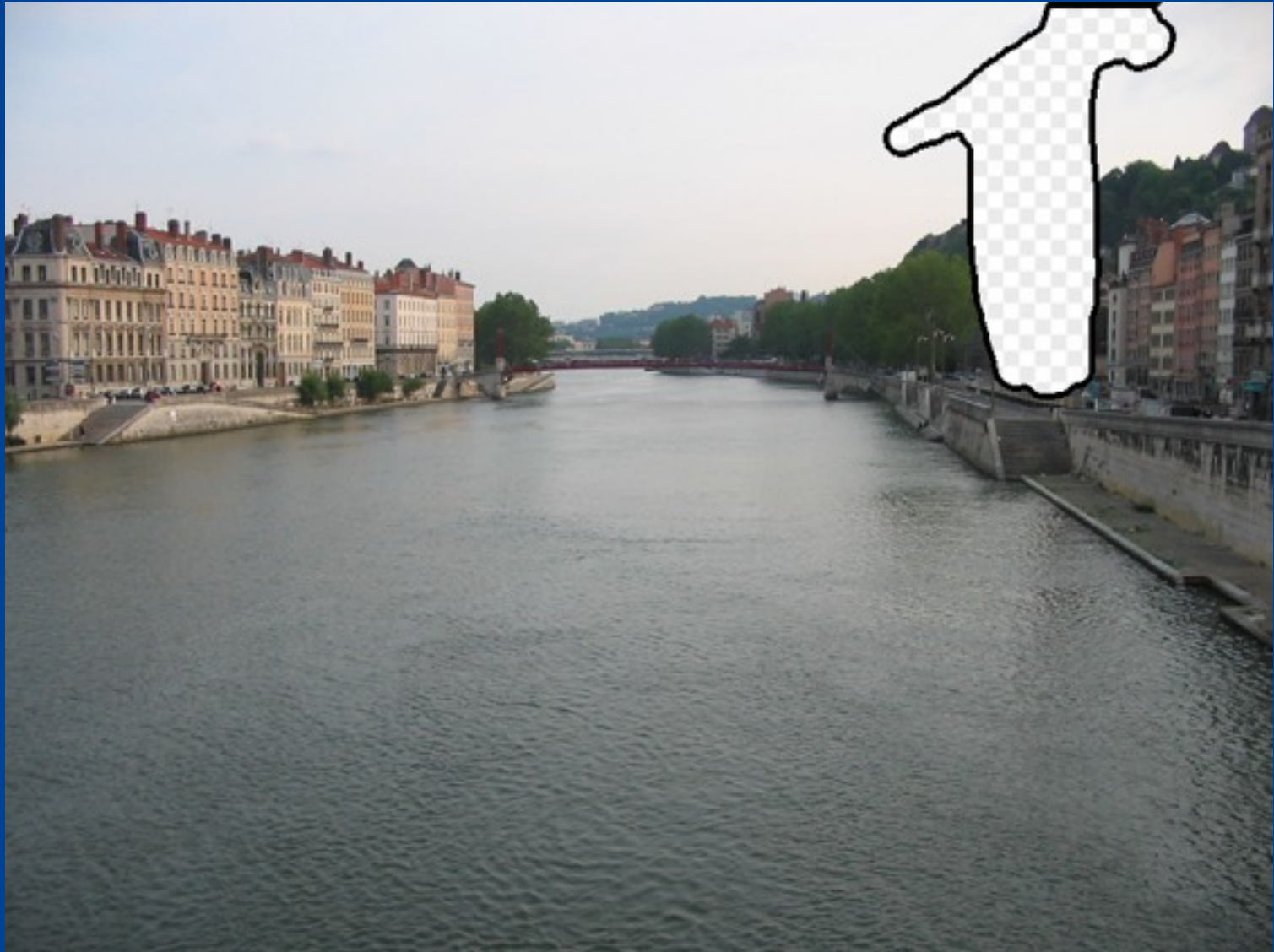
... 200 scene matches



... 200 scene matches







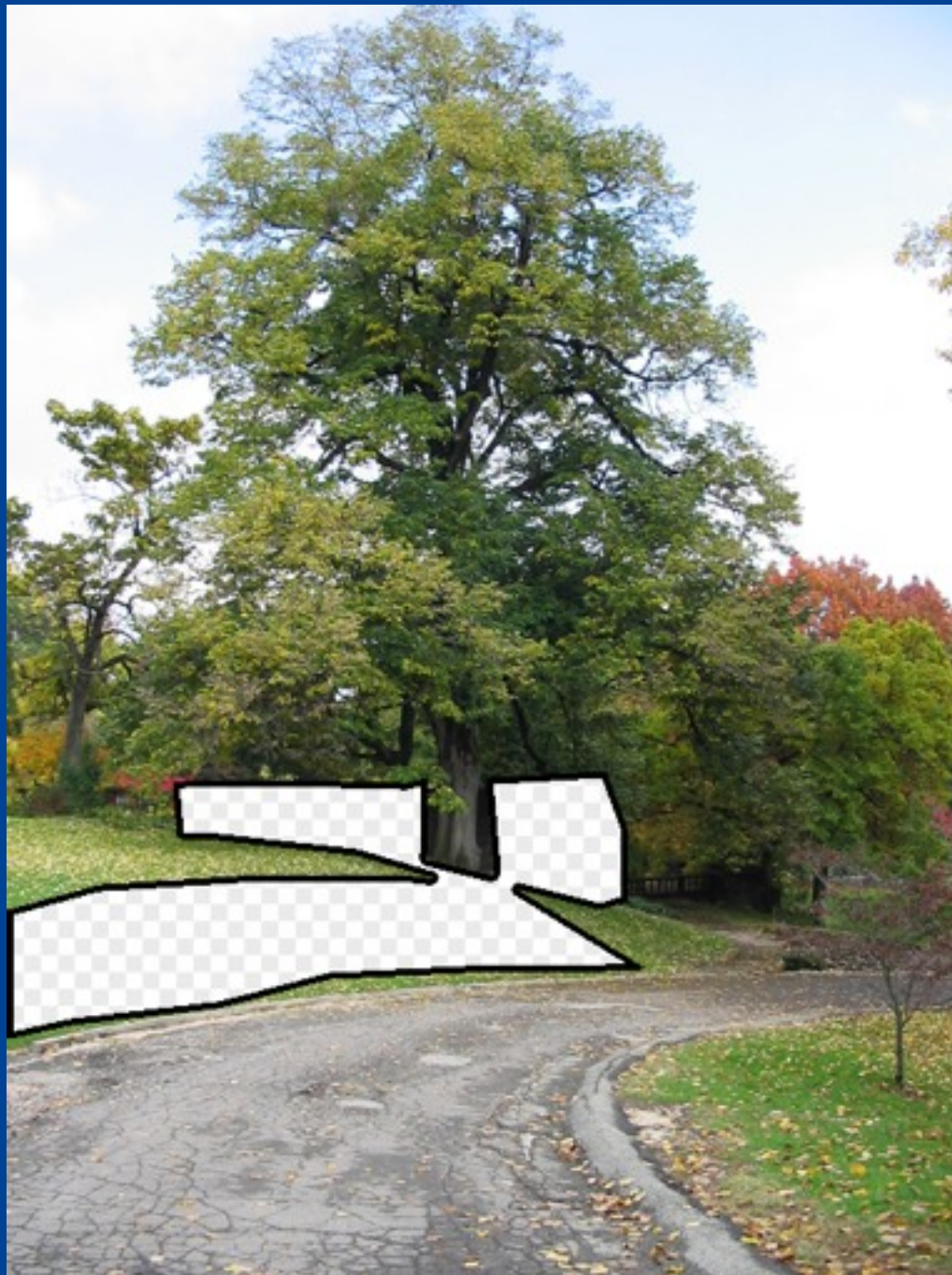




Hays and Efros, SIGGRAPH 2007

Wednesday, May 4, 2011





Hays and Efros, SIGGRAPH 2007

Wednesday, May 4, 2011



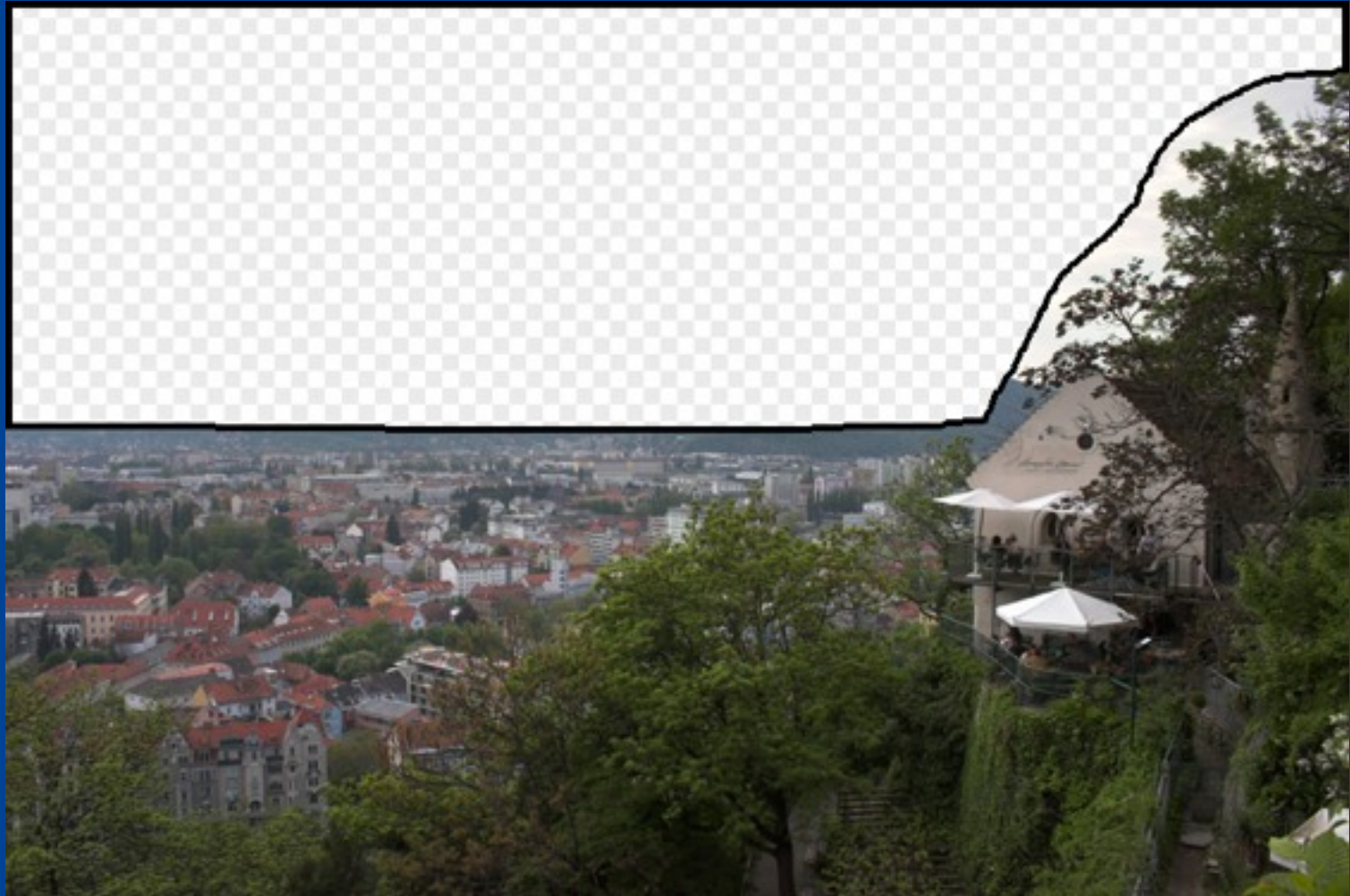
Hays and Efros, SIGGRAPH 2007

Wednesday, May 4, 2011















Hays and Efos, SIGGRAPH 2007

Wednesday, May 4, 2011





Hays and Efos, SIGGRAPH 2007



Hays and Efos, SIGGRAPH 2007

Wednesday, May 4, 2011







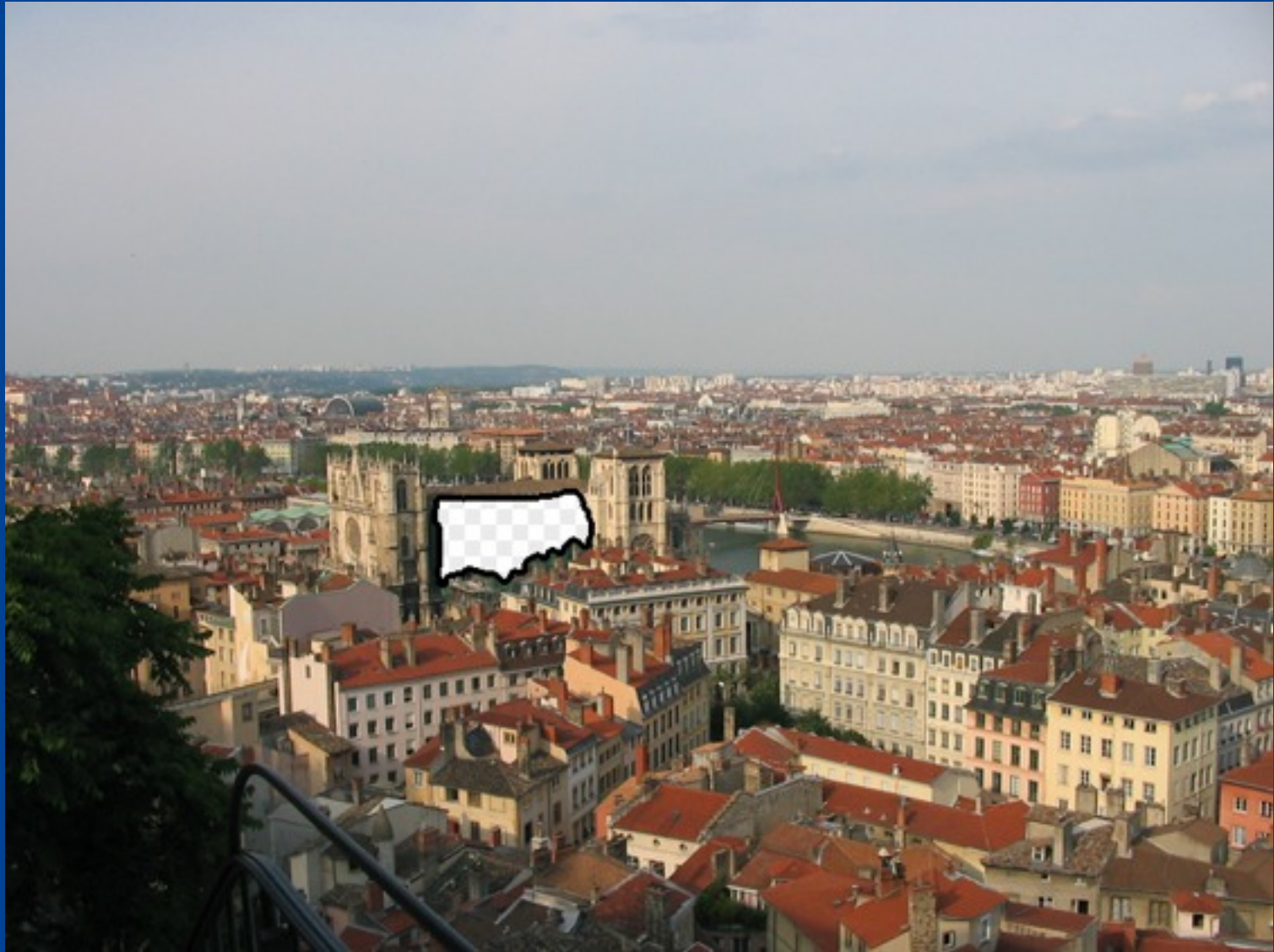






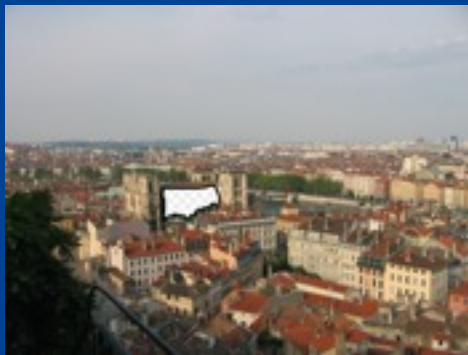






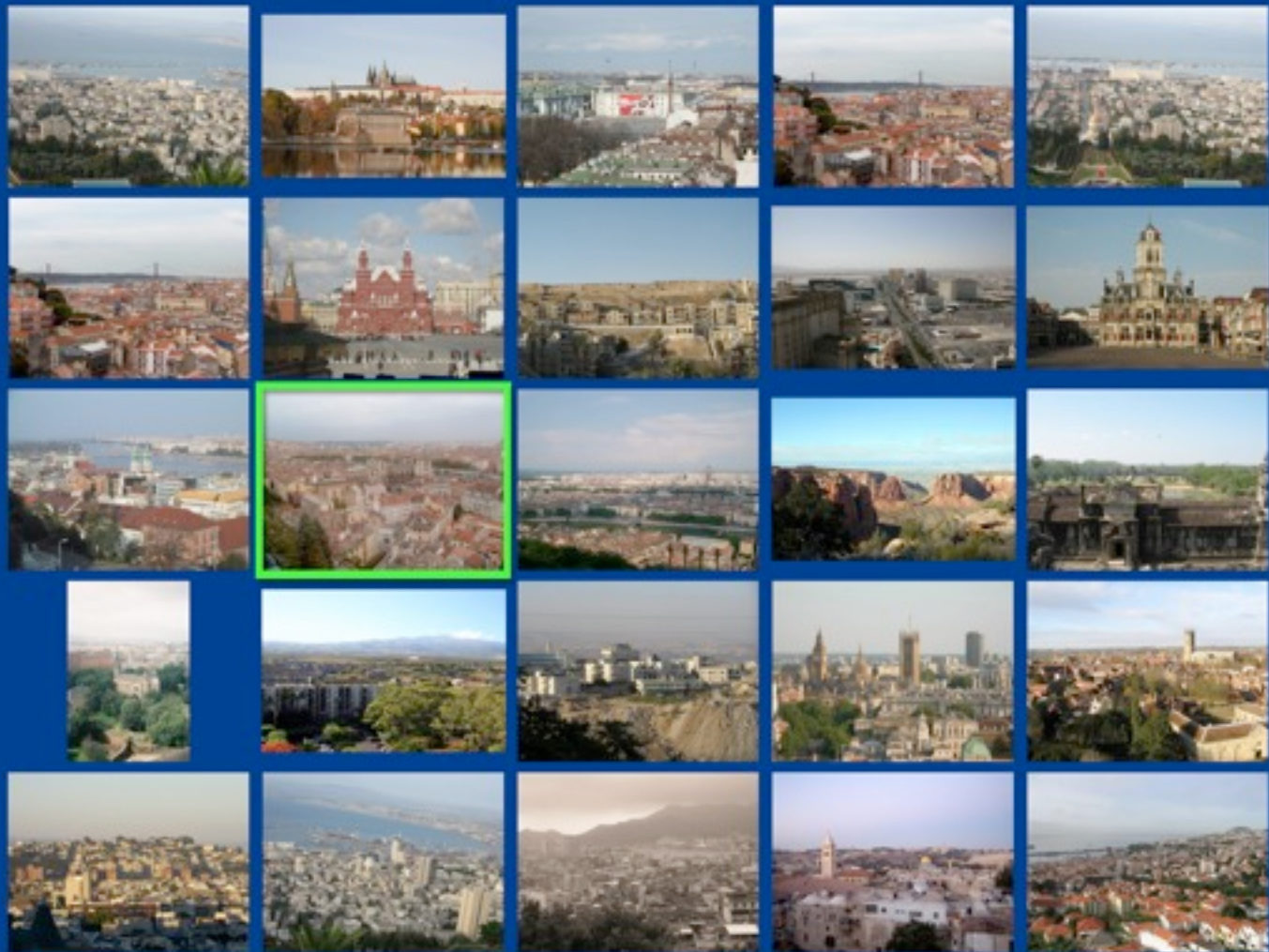
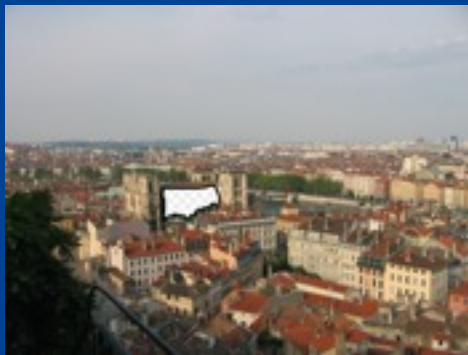




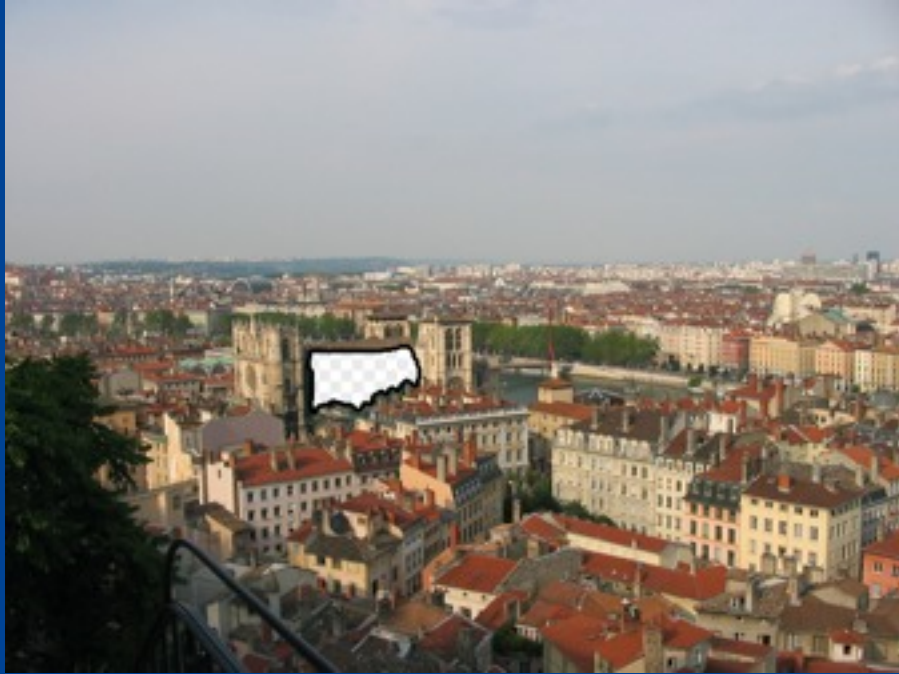


... 200 scene matches

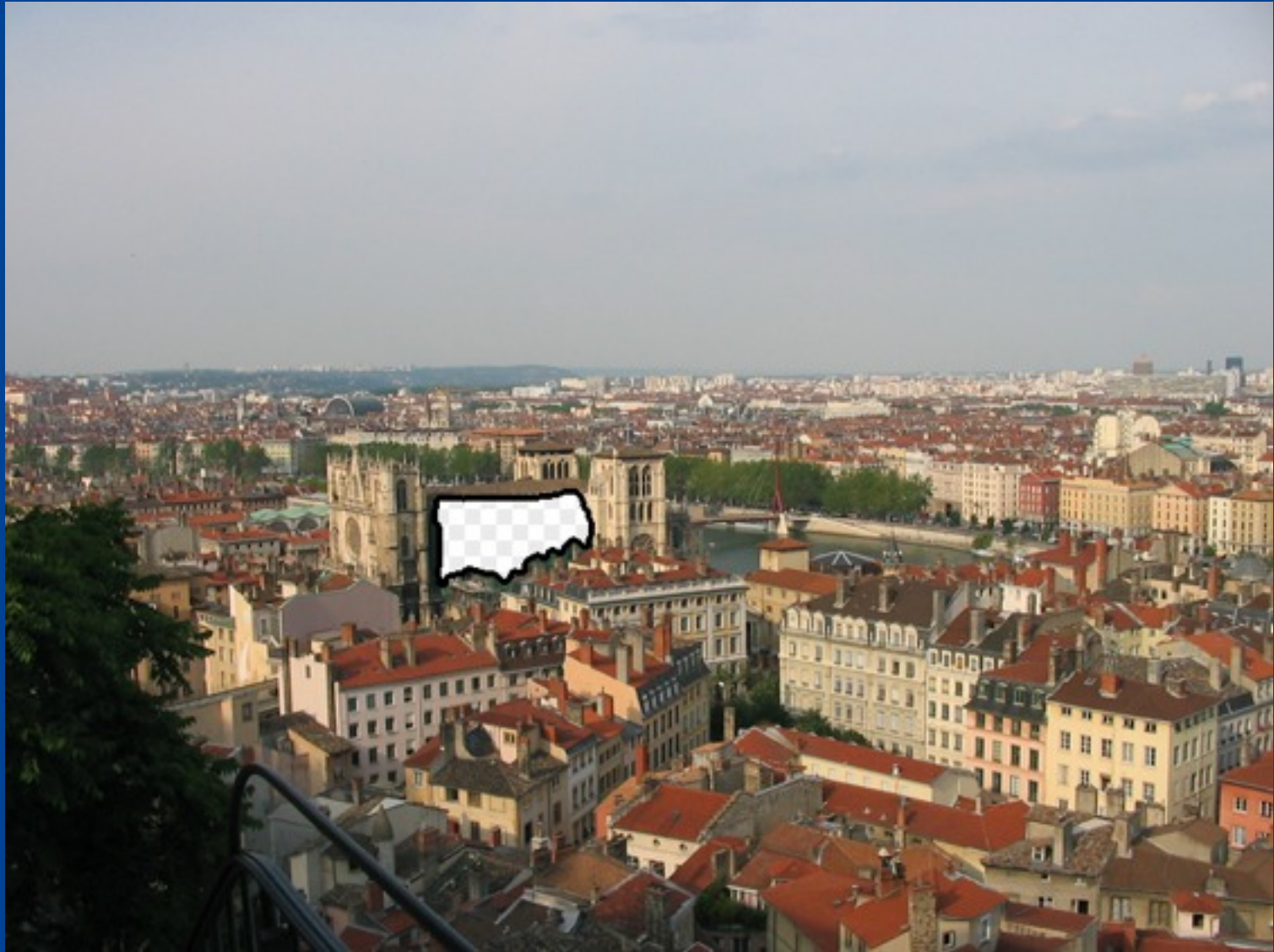




... 200 scene matches











# Failures

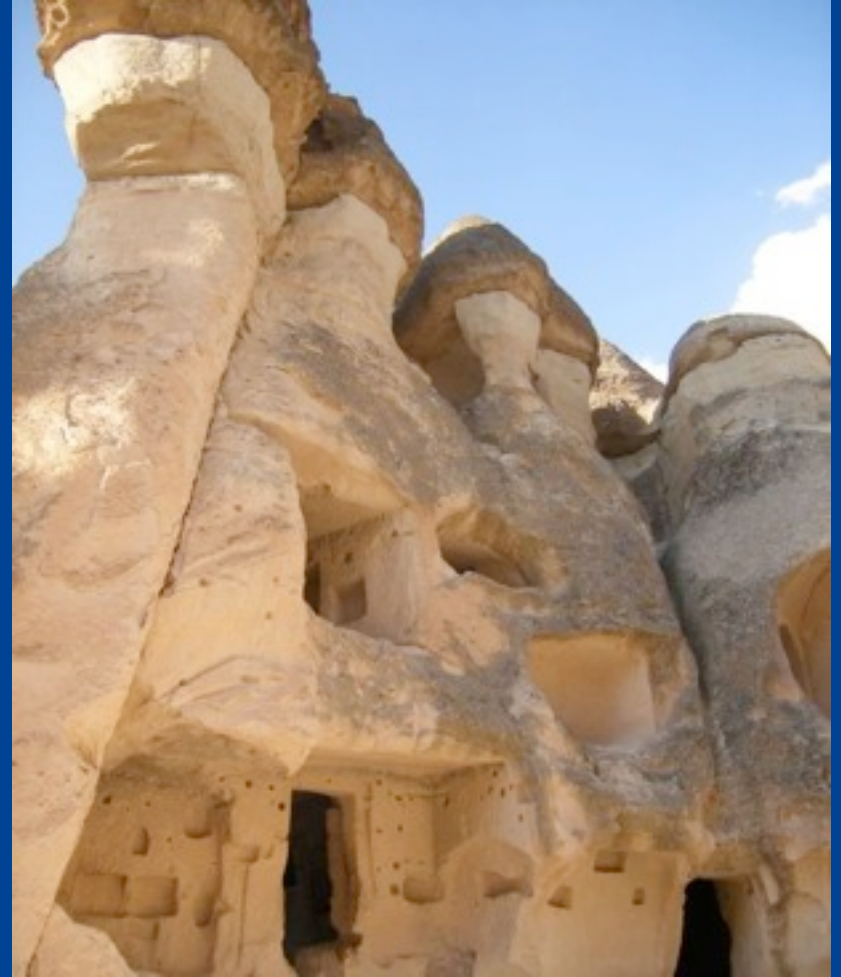


# Failures





# Failures



# Failures

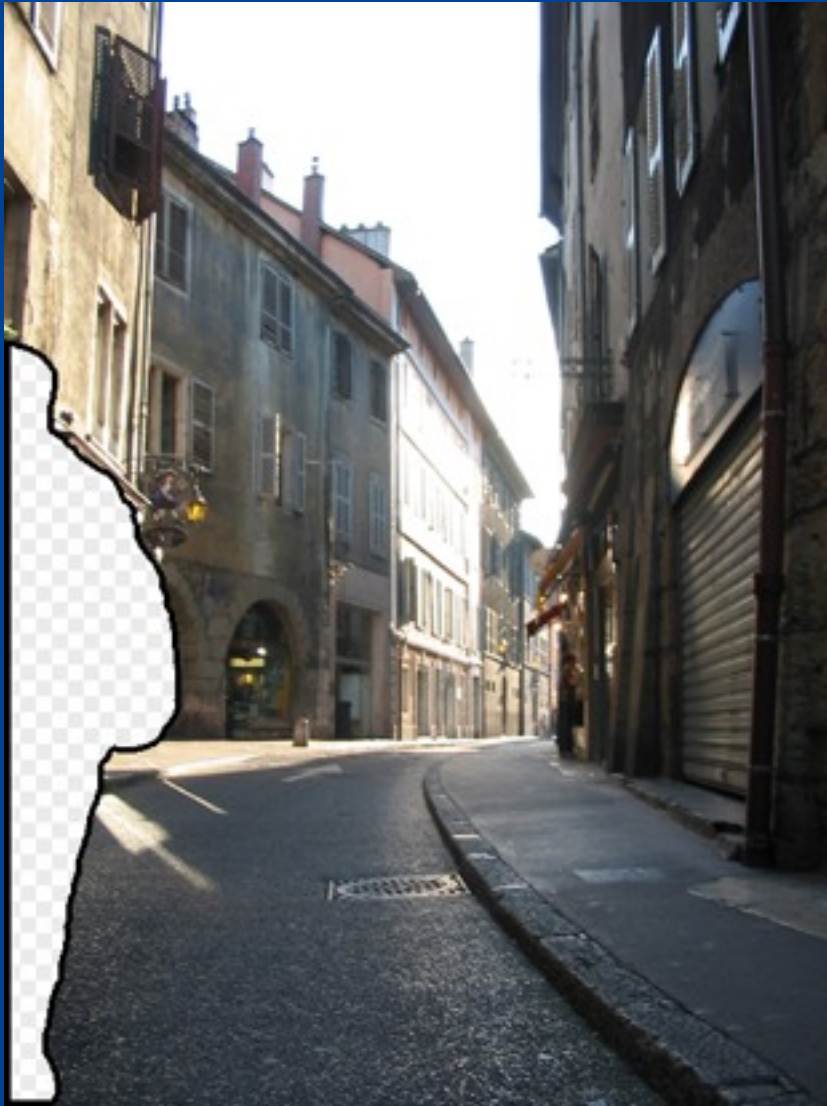


# Failures

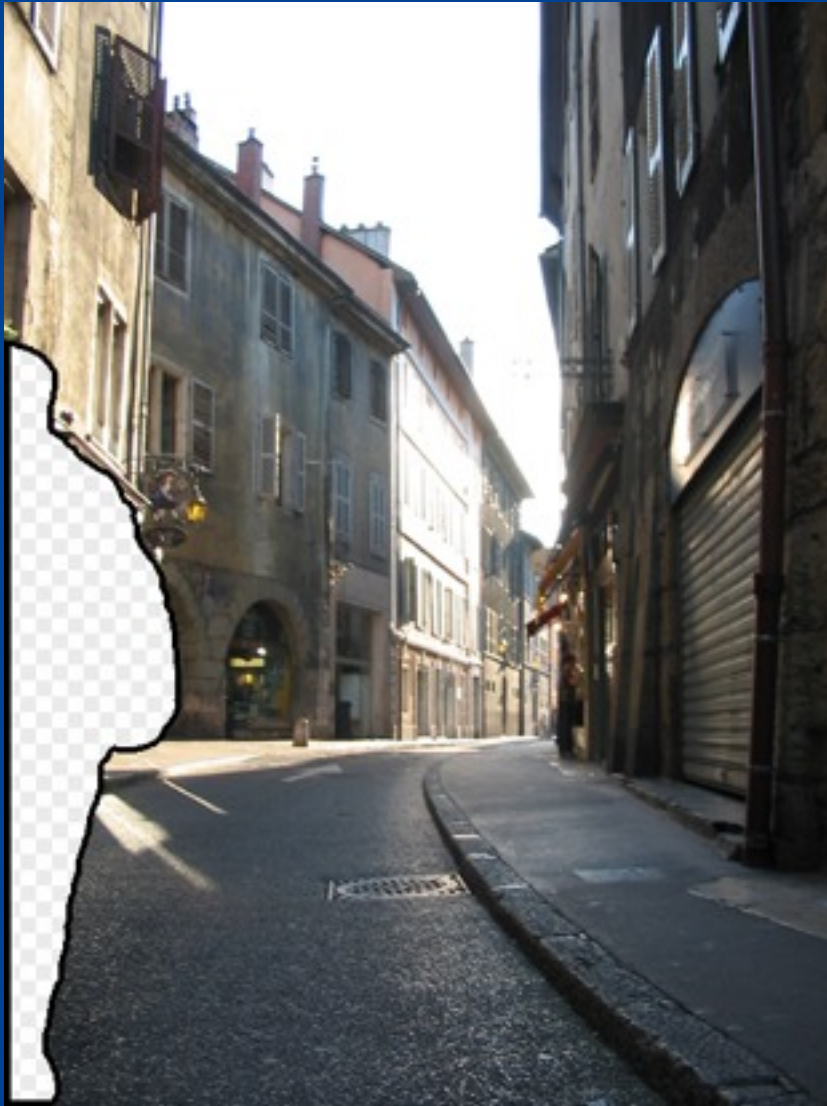




# Failures

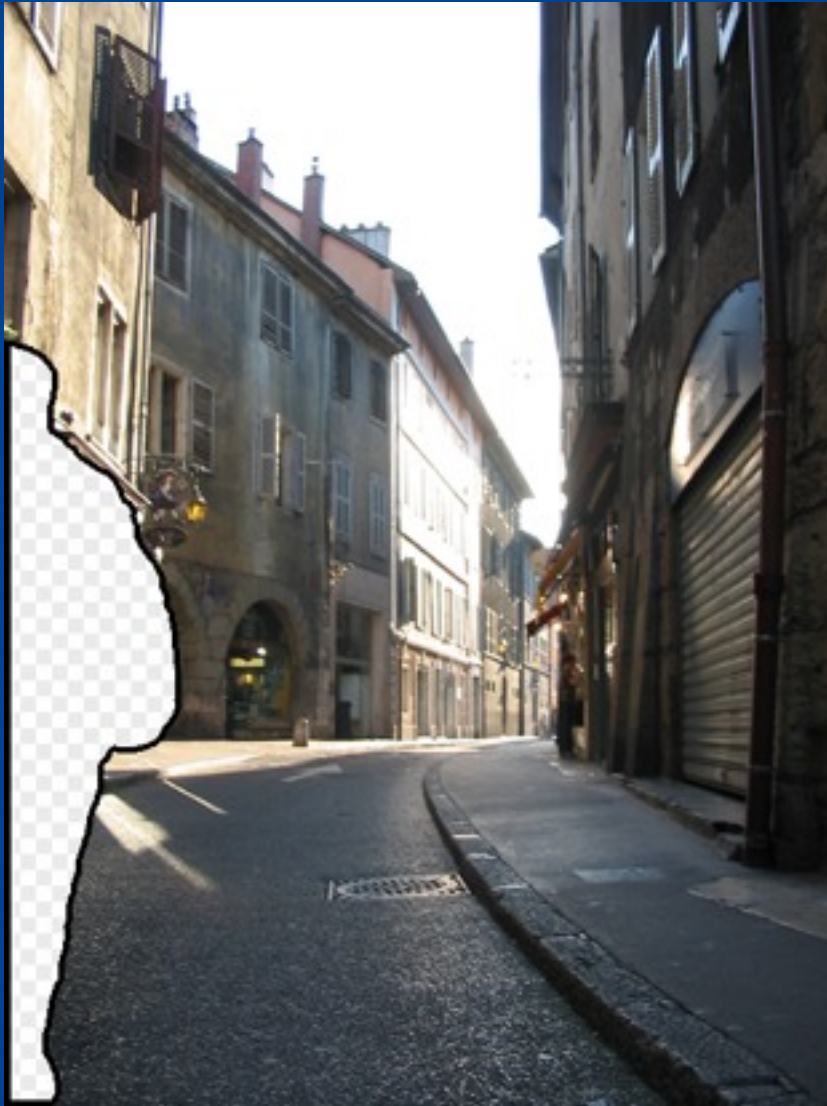


# Failures



Hays and Efros, SIGGRAPH 2007

# Failures



Hays and Efros, SIGGRAPH 2007



# Failures



# Failures



# Failures

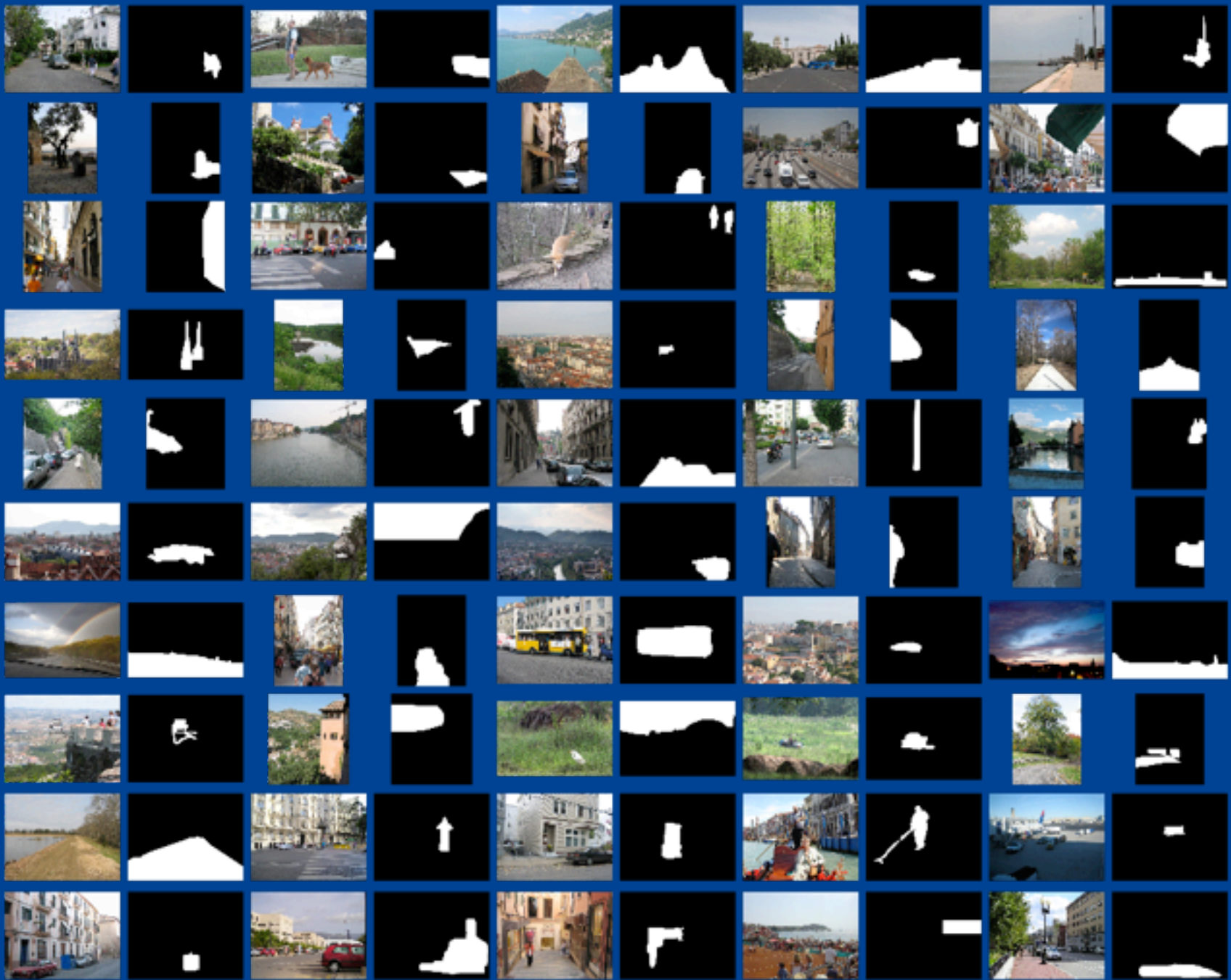




# Failures



# Evaluation







Criminisi et al.

Single result



Scene Completion

Each result  
selected from 20



Original Images



Criminisi et al.

Single result

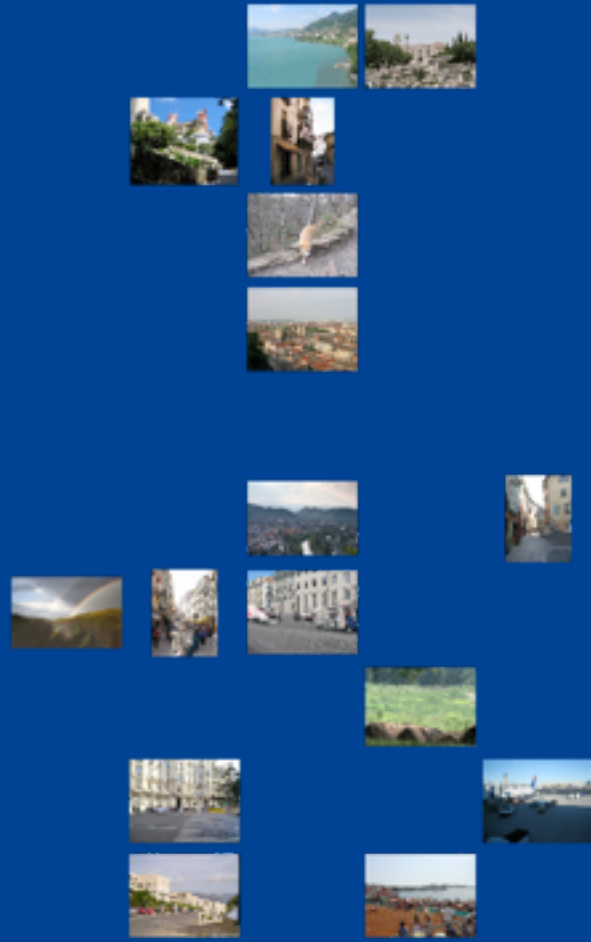


Scene Completion

Each result  
selected from 20



Original Images



Criminisi et al.

Single result



Scene Completion

Each result  
selected from 20







Real Image. This image  
has not been manipulated

or

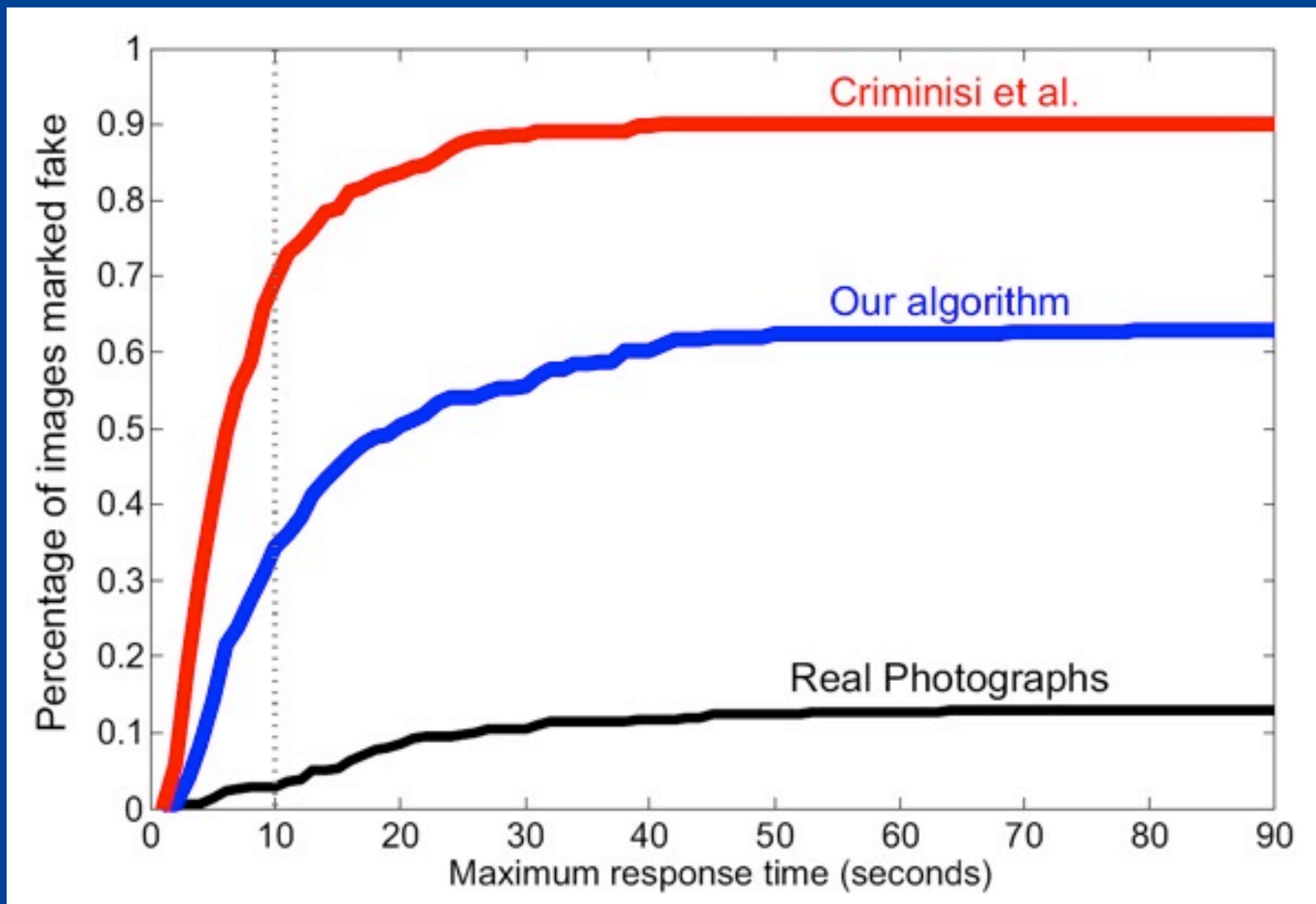
Fake Image. This image  
has been manipulated







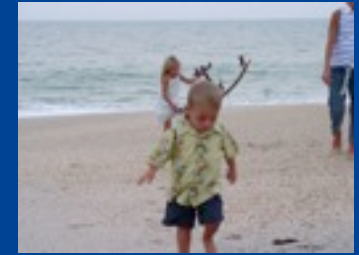
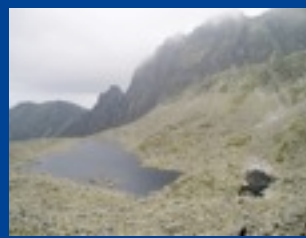
# User Study Results - 20 Participants



Why does it work?





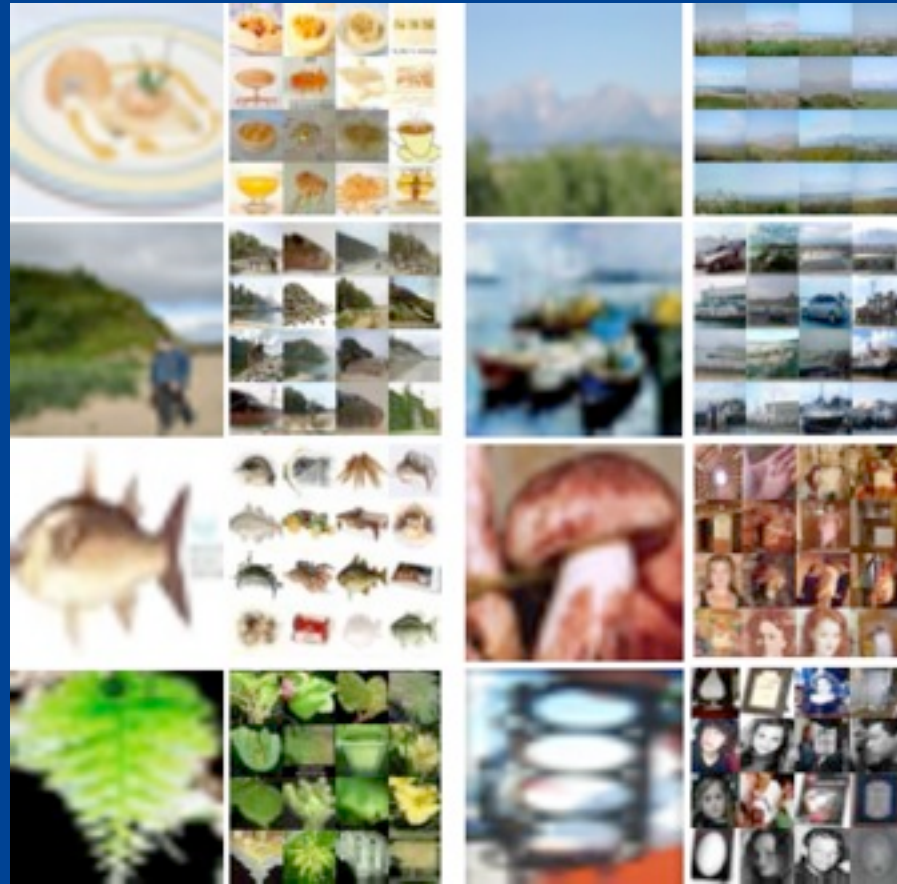


10 nearest neighbors from a collection of 20,000 images



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.

Hays and Efros, SIGGRAPH 2007

# The Big Picture



# The Big Picture





# The Big Picture



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



# The Big Picture



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

Brute-force Image Understanding







# Infinite images

by:

Biliana Kaneva

Josef Sivic

Shai Avidan

Antonio Torralba

Bill Freeman

# Infinite images



Wednesday, May 4, 2011

# Infinite images





# Image representation

Original image



GIST  
[Oliva and Torralba'01]



Color layout



# Obtaining semantically coherent themes

We further break-up the collection into **themes** of semantically coherent scenes:



Train SVM-based classifiers from 1-2k training images  
[Oliva and Torralba, 2001]



# Basic camera motions

Starting from a **single image**,  
images to simulate a camera motion:

find a sequence of

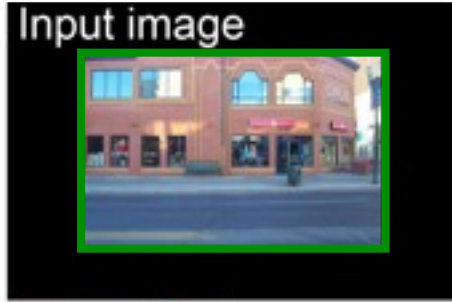
Forward motion

Camera rotation

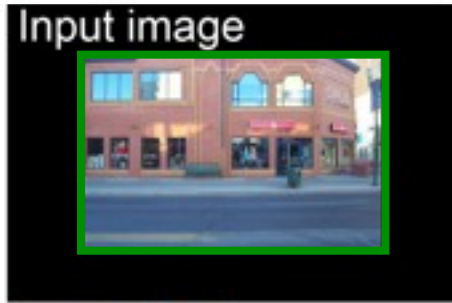
Camera pan



# Scene matching with camera view transformations: Translation



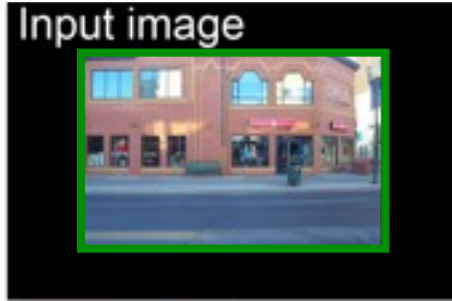
# Scene matching with camera view transformations: Translation



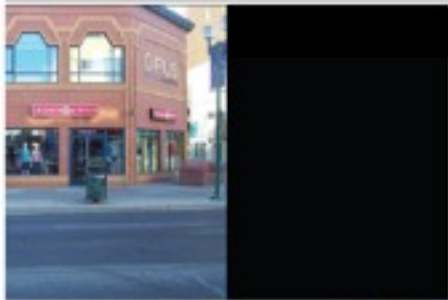
1. Move camera



# Scene matching with camera view transformations: Translation

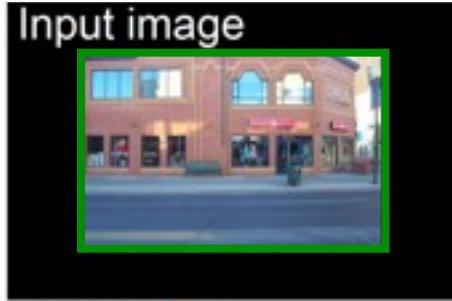


1. Move camera

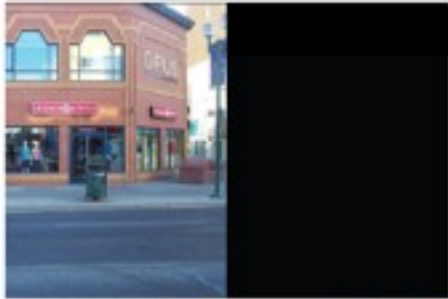


2. View from the virtual camera

# Scene matching with camera view transformations: Translation



1. Move camera

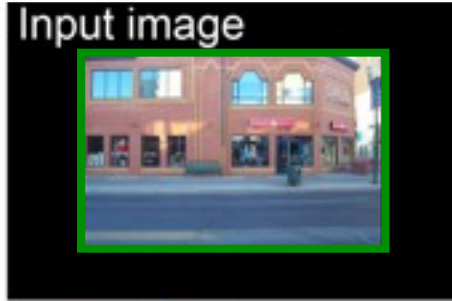


2. View from the virtual camera

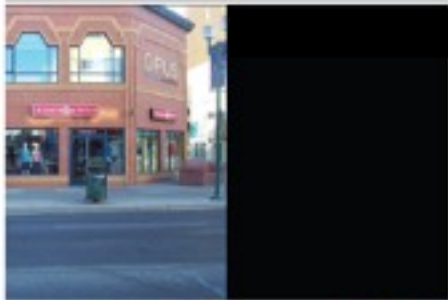


3. Find a match to fill the missing pixels

# Scene matching with camera view transformations: Translation



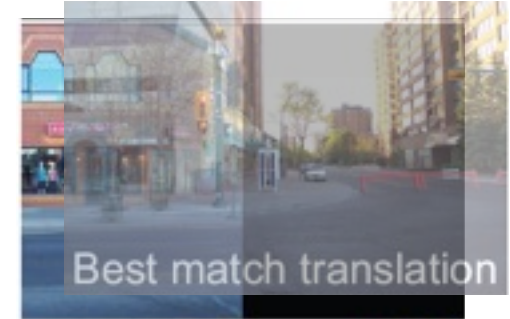
1. Move camera



2. View from the virtual camera



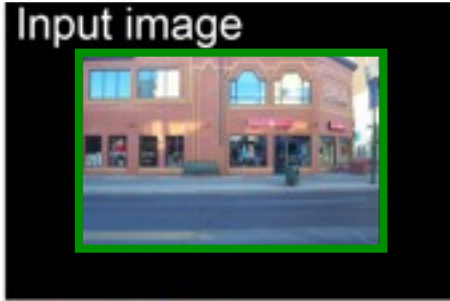
3. Find a match to fill the missing pixels



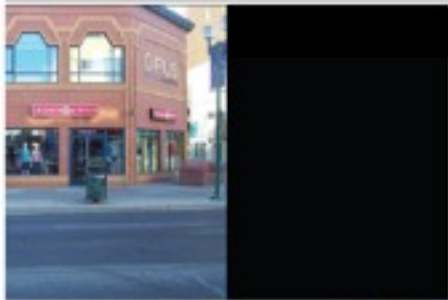
4. Locally align images



# Scene matching with camera view transformations: Translation



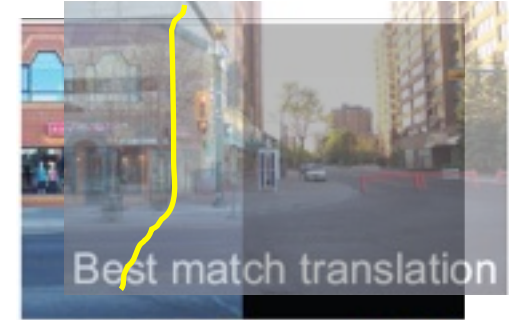
1. Move camera



2. View from the virtual camera



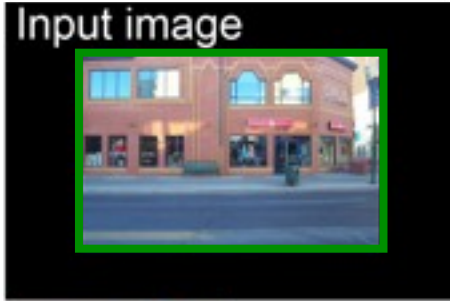
3. Find a match to fill the missing pixels



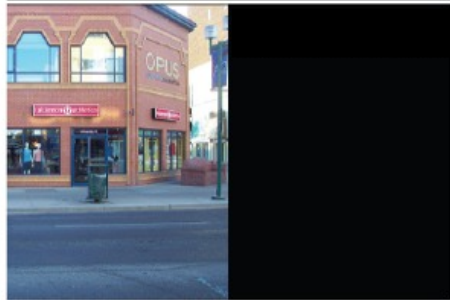
4. Locally align images

5. Find a seam

# Scene matching with camera view transformations: Translation



1. Move camera



2. View from the virtual camera



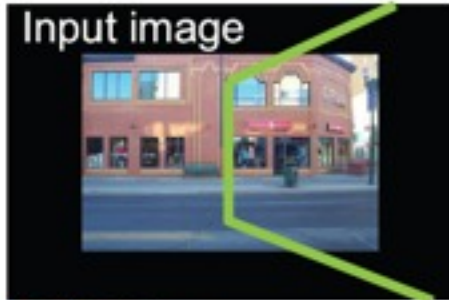
3. Find a match to fill the missing pixels

4. Locally align images

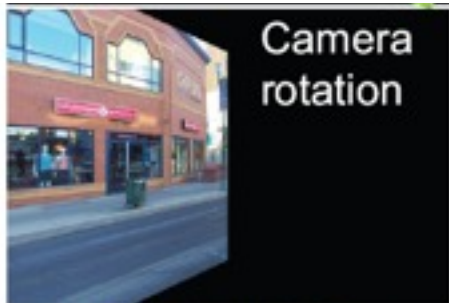
5. Find a seam

6. Blend in the gradient domain

# Scene matching with camera view transformations: Camera rotation



1. Rotate camera



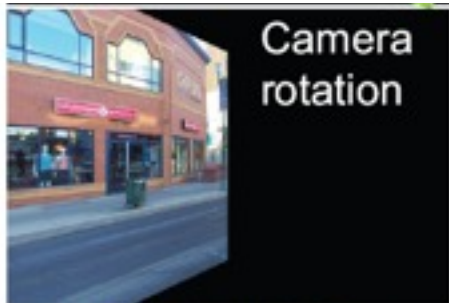
2. View from the virtual camera



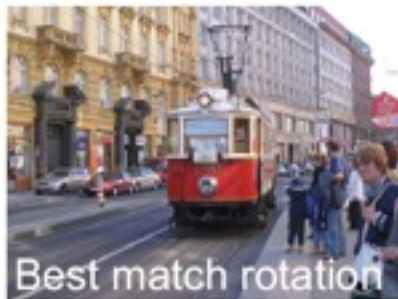
# Scene matching with camera view transformations: Camera rotation



1. Rotate camera

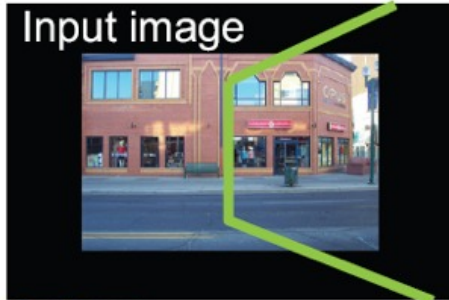


2. View from the virtual camera

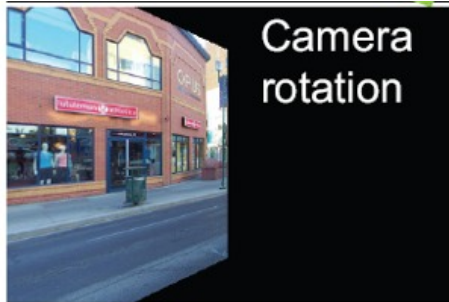


3. Find a match to fill-in the missing pixels

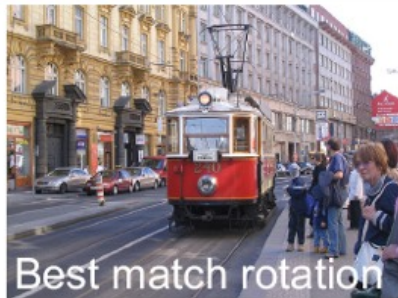
# Scene matching with camera view transformations: Camera rotation



1. Rotate camera



2. View from the virtual camera

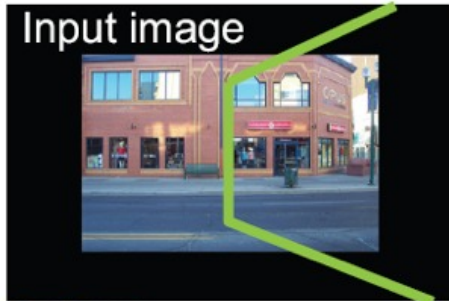


3. Find a match to fill-in the missing pixels

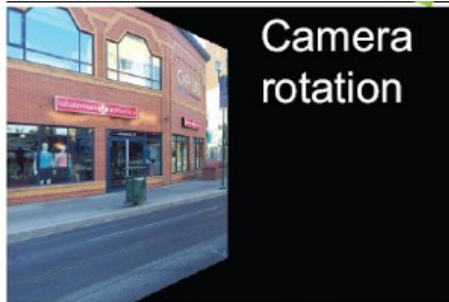


4. Stitched rotation

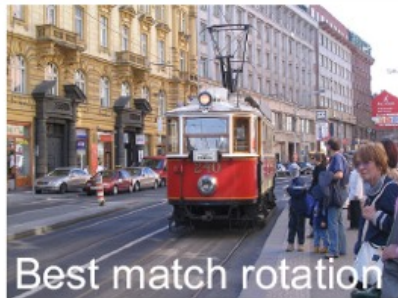
# Scene matching with camera view transformations: Camera rotation



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



# More “infinite” images – camera translation





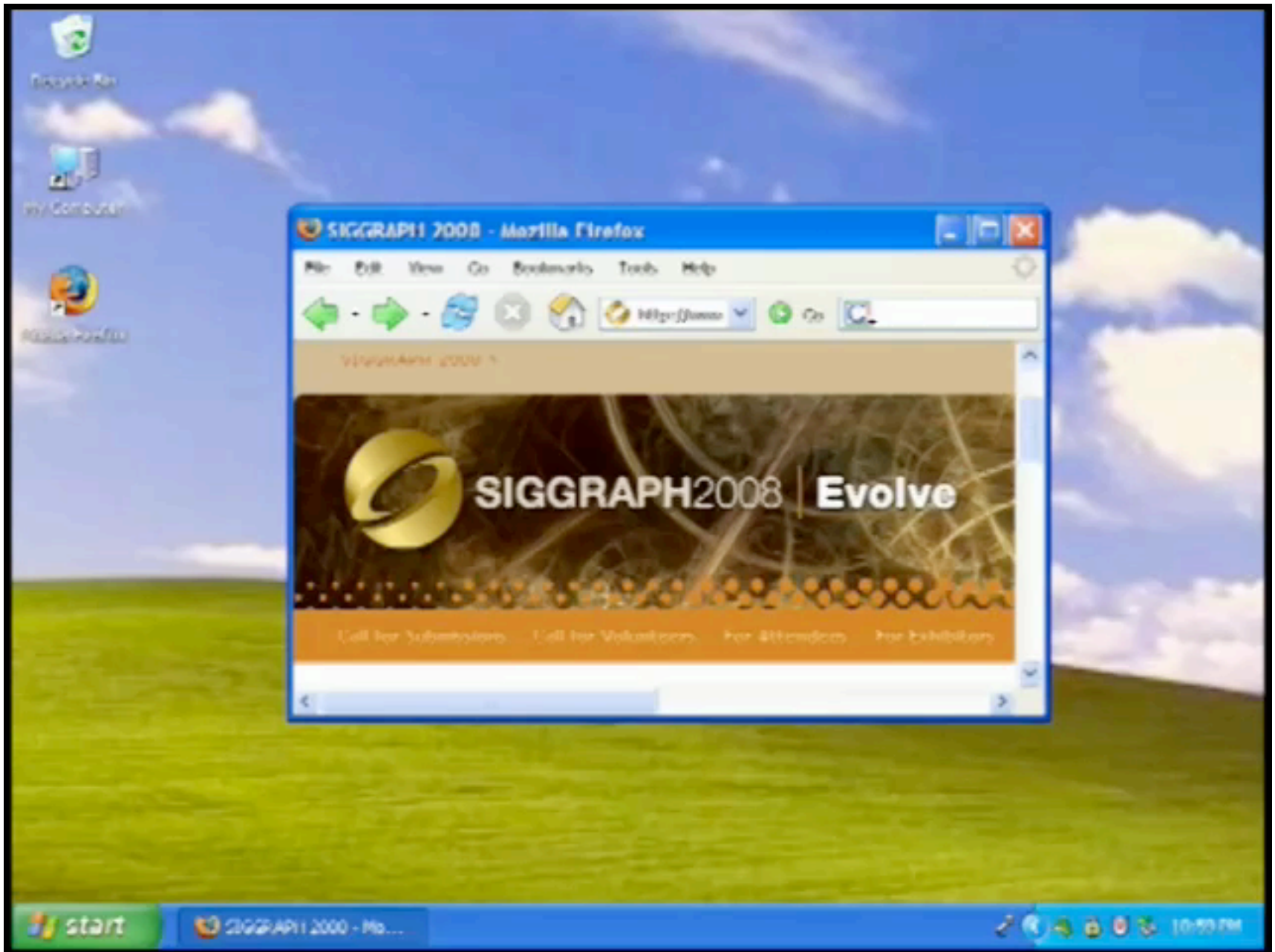






Wednesday, May 4, 2011



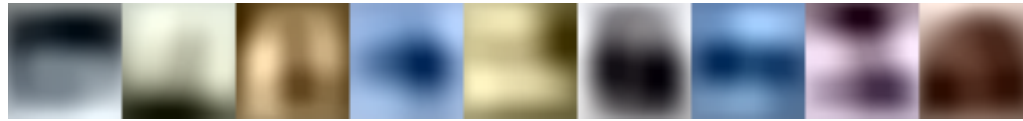






# tools for images on an internet scale

# How many bits do we need?



16 bits

32 bits

64 bits

128 bits

256 bits

512 bits

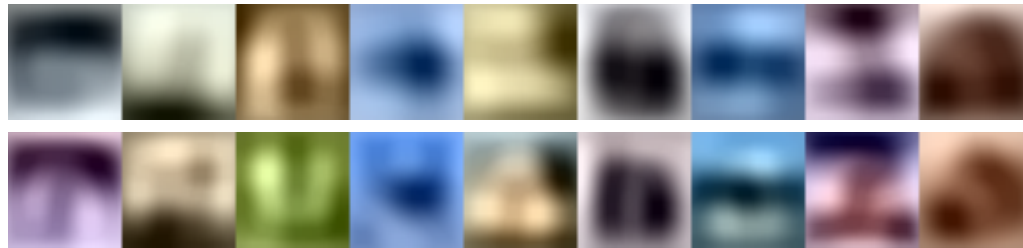
1024 bits

2048 bits

24576 bits



# How many bits do we need?



16 bits

32 bits

64 bits

128 bits

256 bits

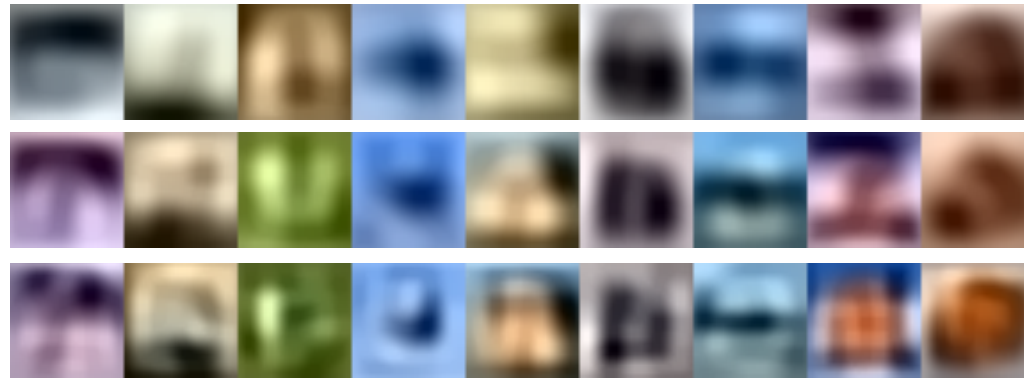
512 bits

1024 bits

2048 bits

24576 bits

# How many bits do we need?



16 bits

32 bits

64 bits

128 bits

256 bits

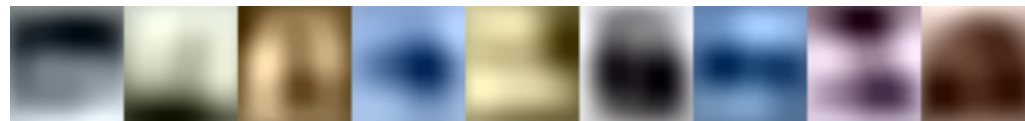
512 bits

1024 bits

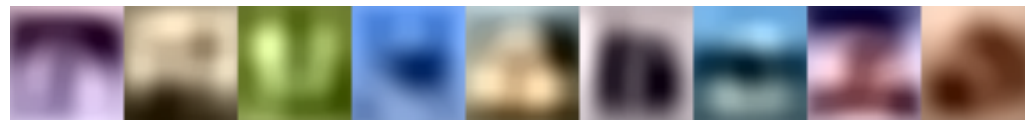
2048 bits

24576 bits

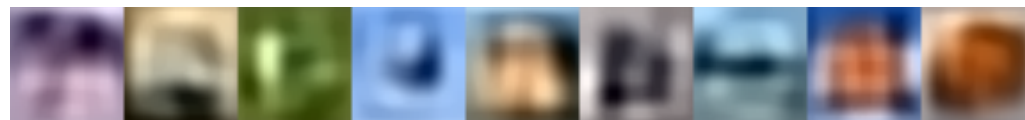
# How many bits do we need?



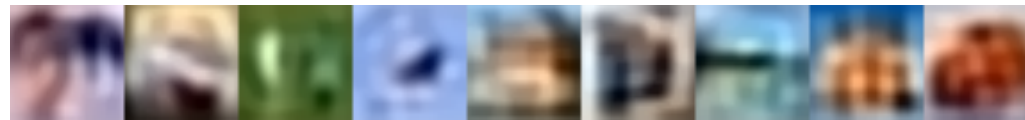
16 bits



32 bits



64 bits



128 bits

256 bits

512 bits

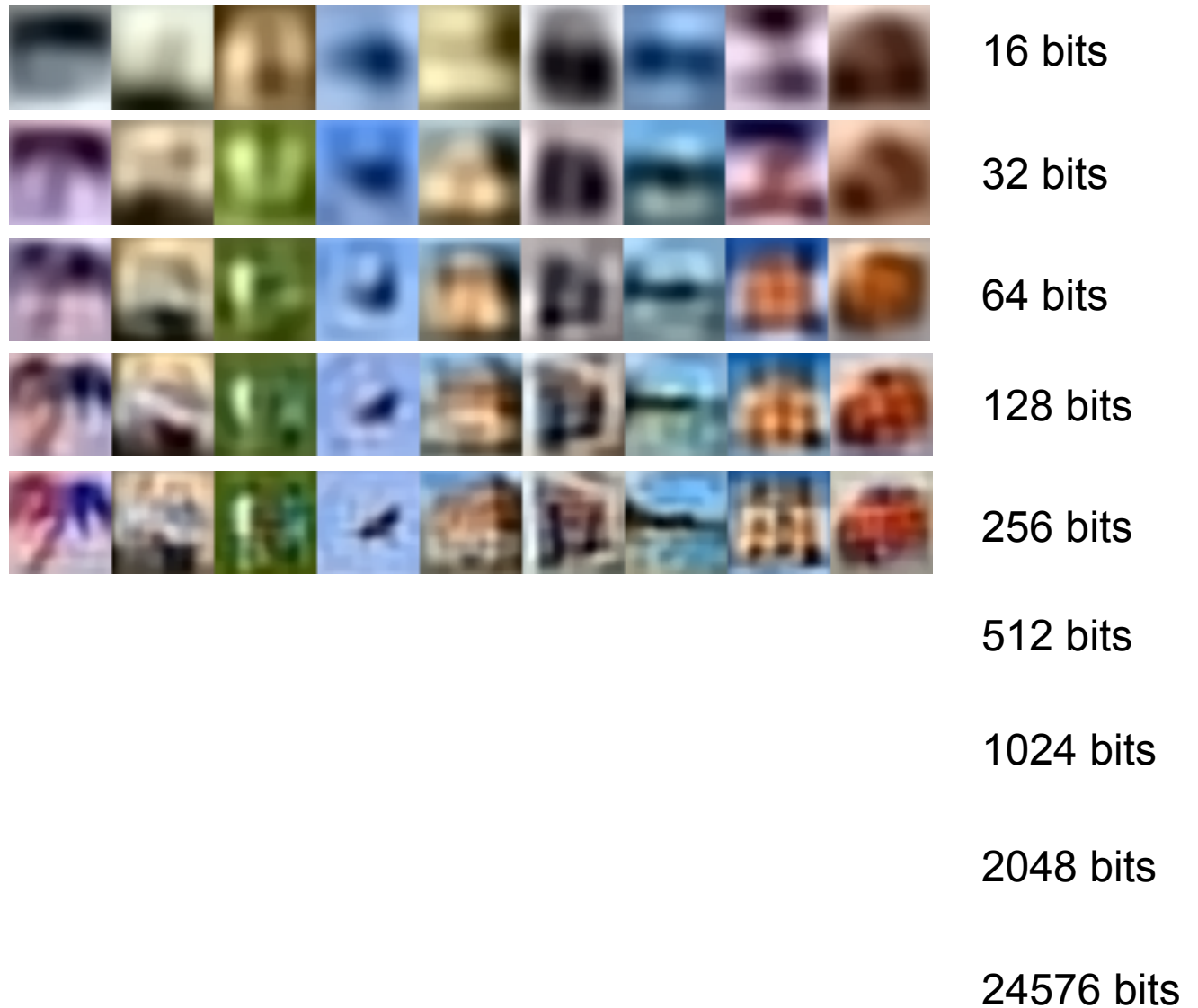
1024 bits

2048 bits

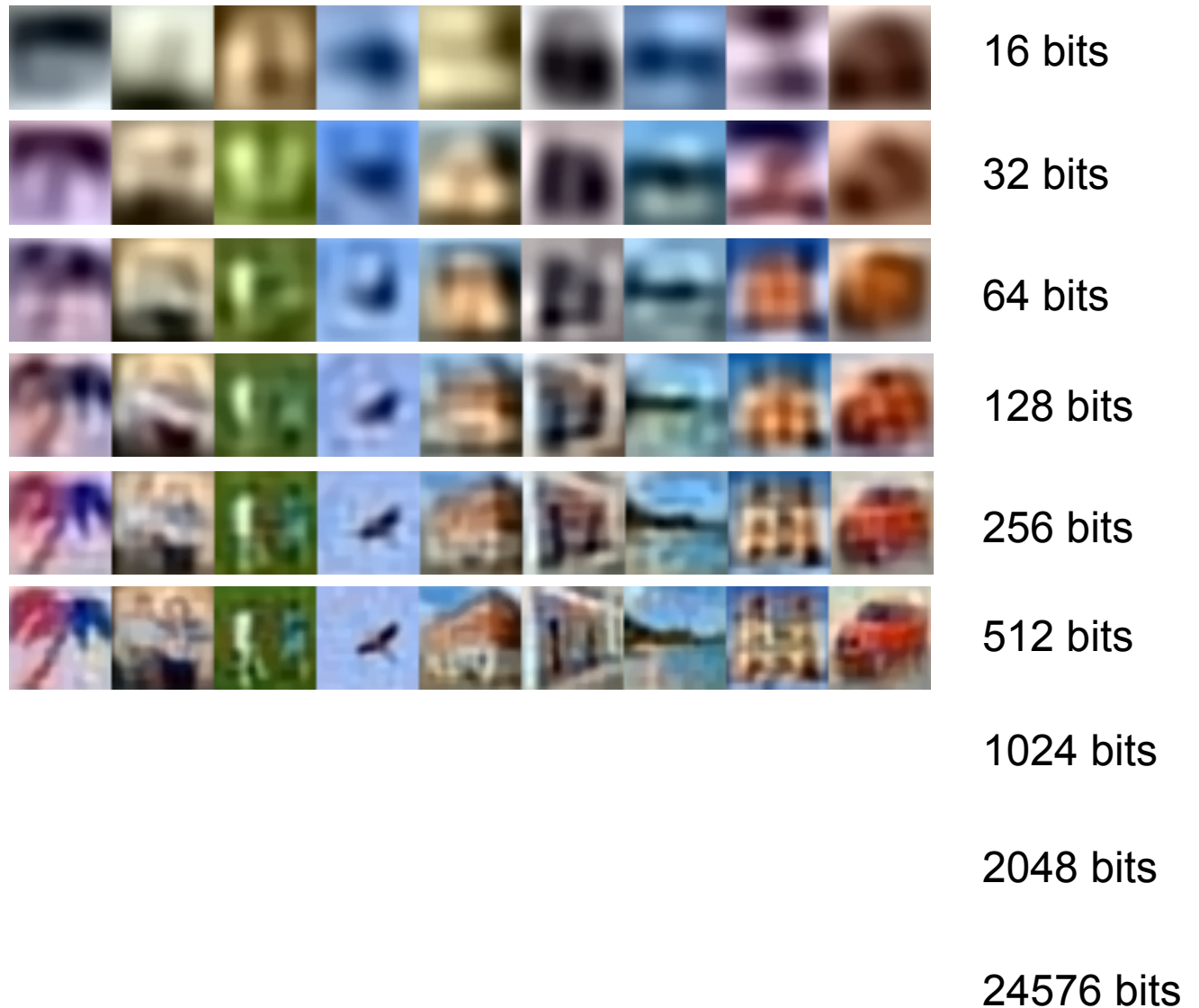
24576 bits



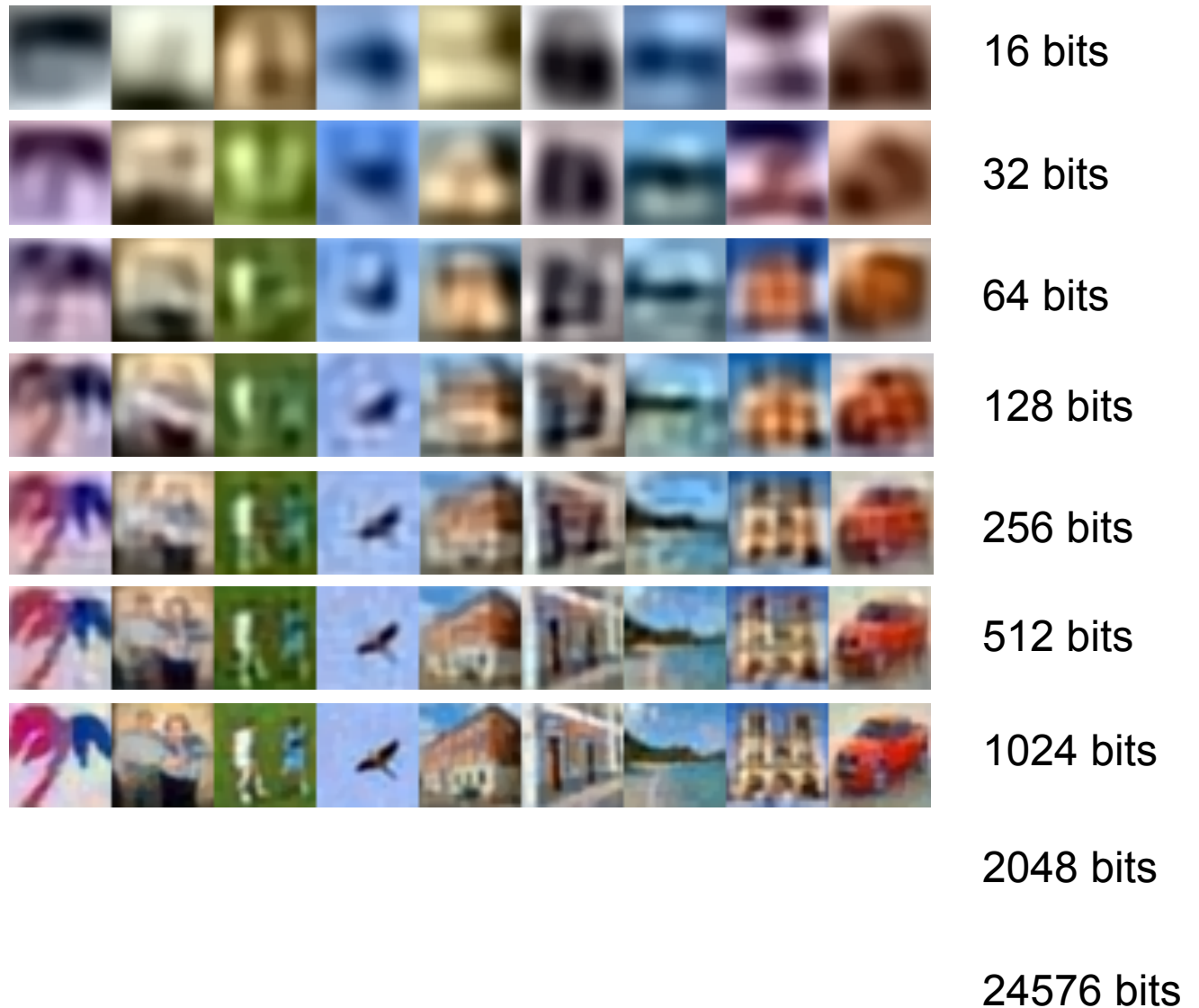
# How many bits do we need?



# How many bits do we need?

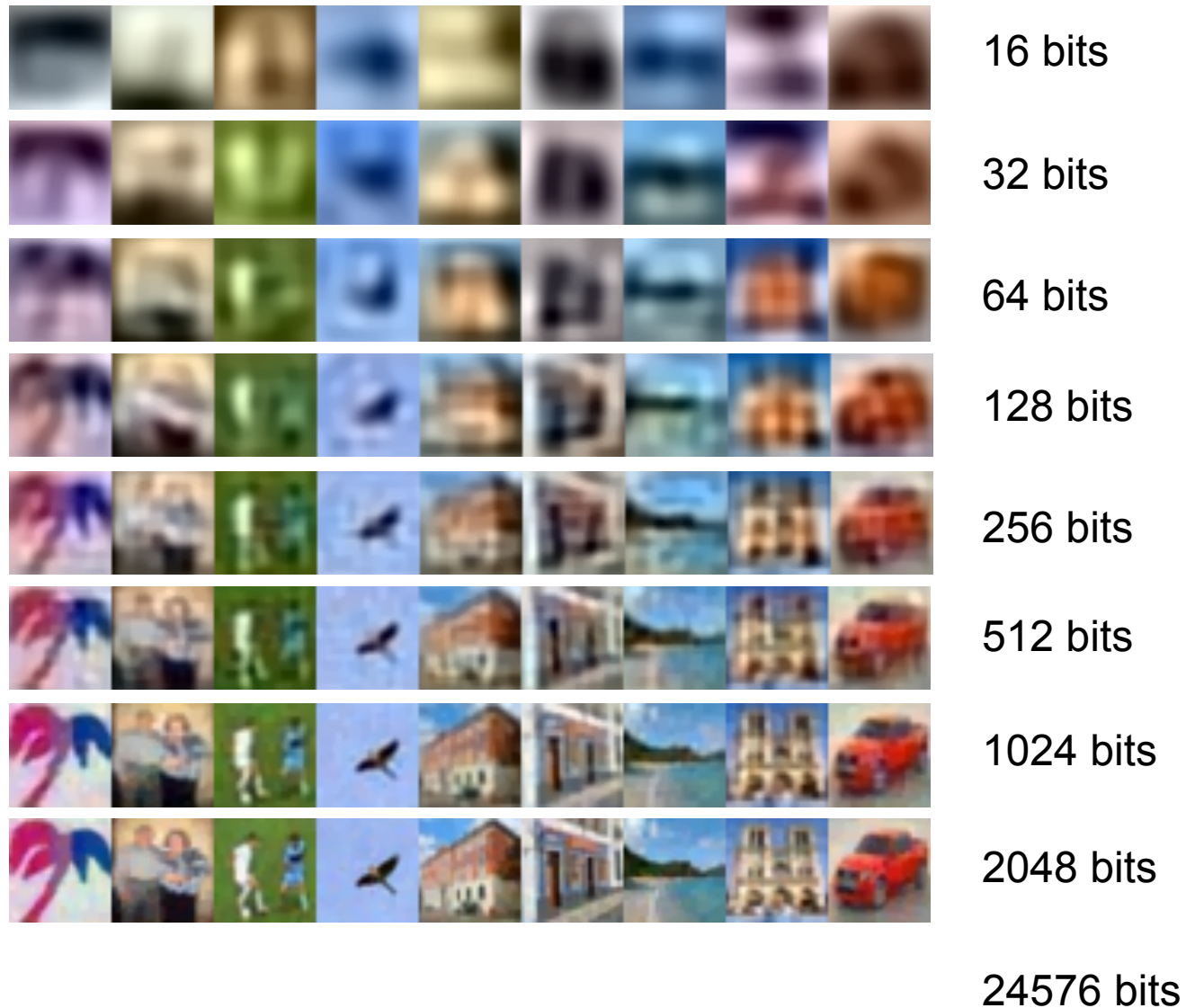


# How many bits do we need?

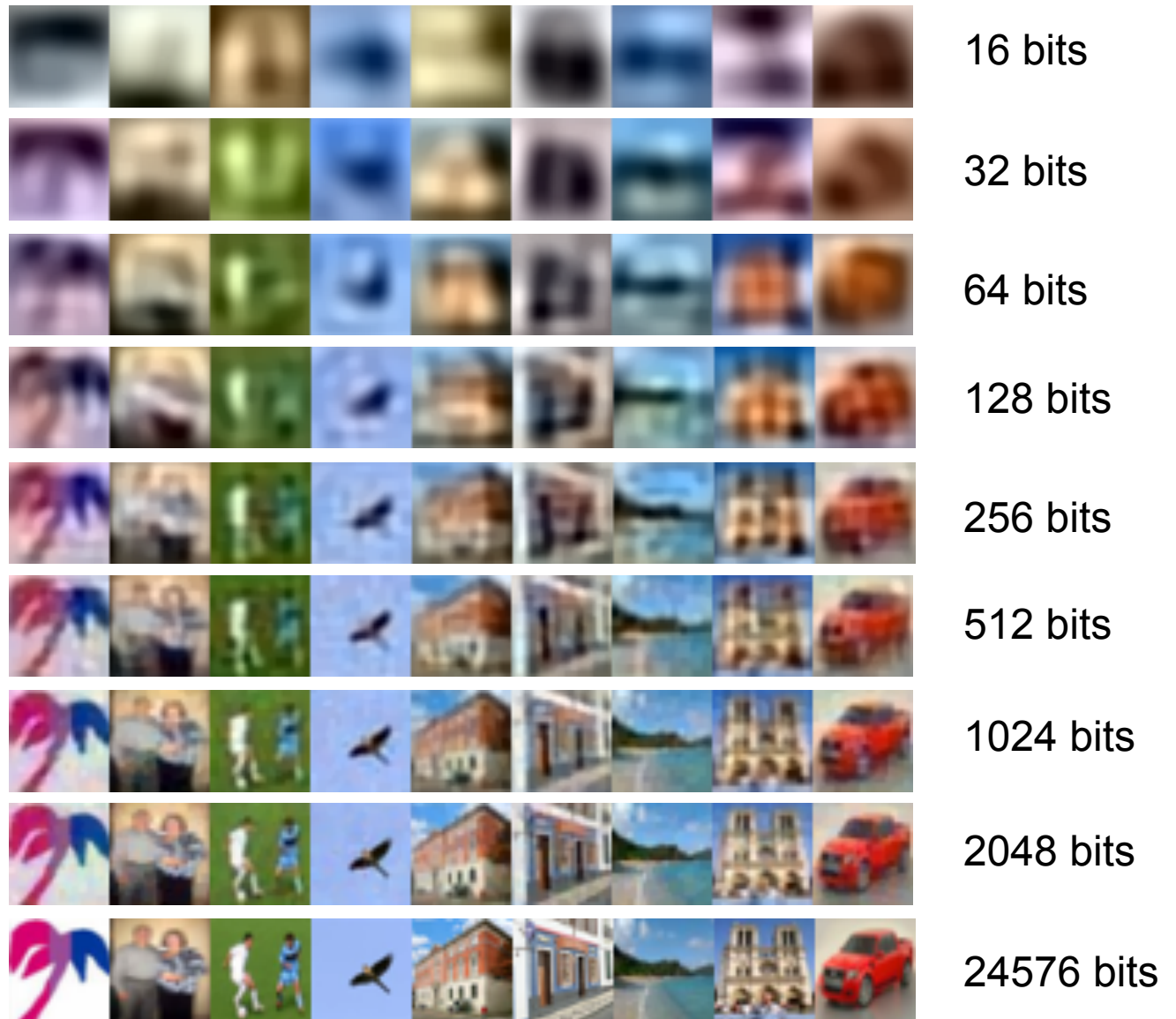




# How many bits do we need?

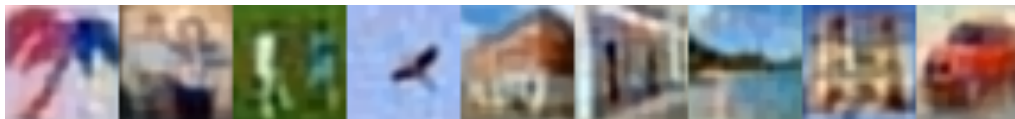


# How many bits do we need?



# Binary codes for global scene representation

- Short codes allow for storing millions of images
- Efficient search: hamming distance (search millions of images in few microseconds)
- Internet scale experiments: compute nearest neighbors between all images in the internet



512 bits

A. Torralba, R. Fergus, and Y. Weiss. *Small codes and large databases*  
[people.csail.mit.edu/torralba/publications/spectralhashing.pdf](http://people.csail.mit.edu/torralba/publications/spectralhashing.pdf)



end

# Binary codes for images

- Want images with similar content to have similar binary codes
- Use Hamming distance between codes
  - Number of bit flips
  - E.g.:  
 $\text{Ham\_Dist}(10001010, 10001110) = 1$   
 $\text{Ham\_Dist}(10001010, 11101110) = 3$
- Semantic Hashing [Salakhutdinov & Hinton, 2007]
  - Text documents

# Compact Binary Codes

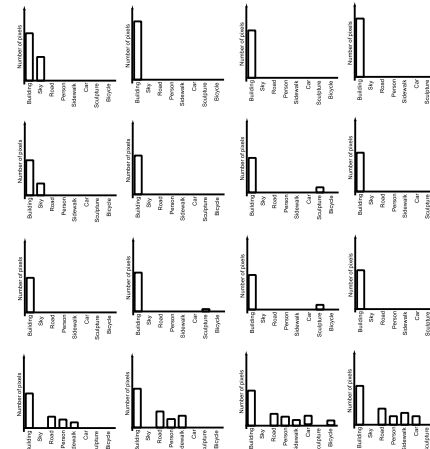
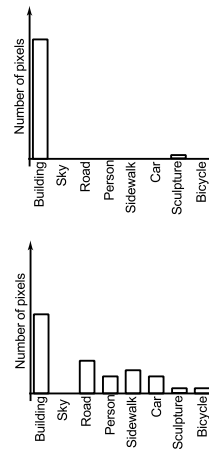
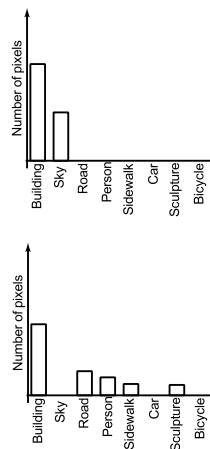
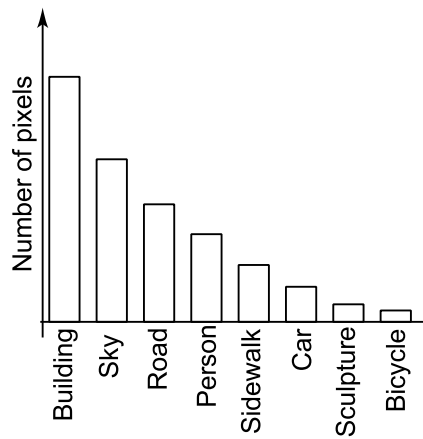
- Google has few billion images ( $10^9$ )
  - Big PC has  $\sim 10$  Gbytes ( $10^{11}$  bits)
  - Codes must fit in memory (disk too slow)
- Budget of  $10^2$  bits/image



# Compact Binary Codes

- Google has few billion images ( $10^9$ )
- Big PC has  $\sim 10$  Gbytes ( $10^{11}$  bits)
- Codes must fit in memory (disk too slow)  
→ Budget of  $10^2$  bits/image
  
- 1 Megapixel image is  $10^7$  bits
- 32x32 color image is  $10^4$  bits  
→ Semantic hash function must also reduce dimensionality

# Measuring image similarity with annotated data



$$S(h_1, h_2) = \text{sum}(\min(h_1, h_2))$$

Spatial pyramid matching [Lazebnik06, Grauman07]

# Hashing

We consider the following learning problem - given a database of images  $\{x_i\}$  and a distance function  $D(i, j)$  we seek a binary feature vector  $y_i = f(x_i)$  that preserves the nearest neighbor relationships using a Hamming distance.

Salakhutdinov and Hinton [SIGIR 2007], Shakhnarovich et al [ICCV 2003], Athitsos et al. [ICDE 2008], Grauman et al [CVPR 2007], Nascimento et al [ACM Smp. App. Computing 2002], Wang [ICME 2006], Wang [PAMI 2008],

# Learning hamming distances with boosting

Each image is represented by a binary vector with  $M$  bits

$$y = [h_1(x), h_2(x), \dots, h_M(x)] \quad \left\{ \begin{array}{l} x = \text{vector of image features} \\ h_i = \text{function with binary output} \\ y = \text{binary vector} \end{array} \right.$$

Distance between two images is given by a weighted Hamming distance

$$D(i, j) = \sum_{n=1}^M \alpha_n |h_n(x_i) - h_n(x_j)|$$

The weights  $\alpha_i$  and the functions  $h_n(x_i)$  that map the input vector  $x_i$  into binary features are learned.

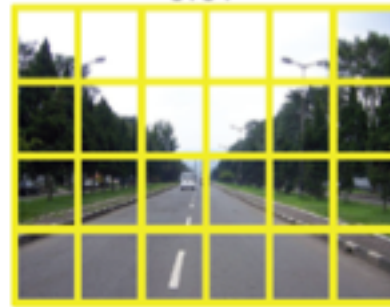


# Compressing the gist descriptor

Original image



GIST  
[Oliva and Torralba'01]



Input image

Ground truth neighbors

Gist

Gist (32 - bits)

