# Data-driven methods

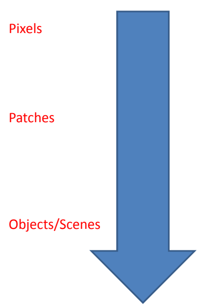
Lecture 8

## Admin

- Office hours straight after class today
- Assignment 3 out, due in 2 weeks
- Projects…..

## Overview

- Texture synthesis
- Quilting
- Image Analogies
- Super-resolution
- Scene completion

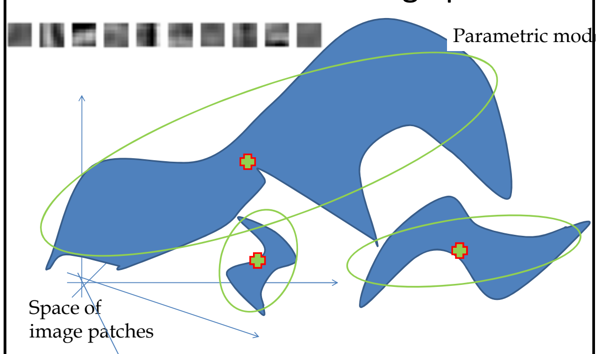



## Overview

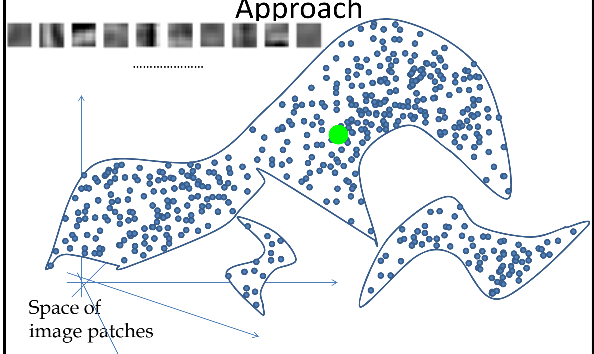
- Texture synthesis [Efros & Leung, ICCV'99]
- Quilting [Efros & Freeman 2001]
- Image Analogies [Hertzmann et al. 2001]
- Super-resolution [Freeman et al. 2002]
- Scene completion [Hays & Efros 2007]

Slides from: Alyosha Efros, Bill Freeman, James Hayes

## Parametric model of image patches




## Data-driven (Non-parametric) Approach



## Overview

- Texture synthesis
- Quilting
- Image Analogies
- Super-resolution
- Scene completion

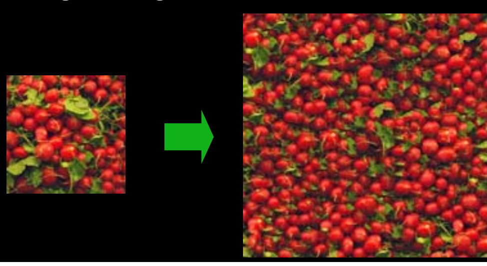
## Texture

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

radishes rocks yogurt

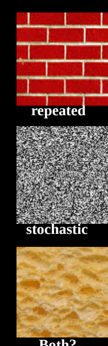
## Texture Synthesis

- Goal of Texture Synthesis: create new samples of a given texture
- Many applications: virtual environments, hole-filling, texturing surfaces

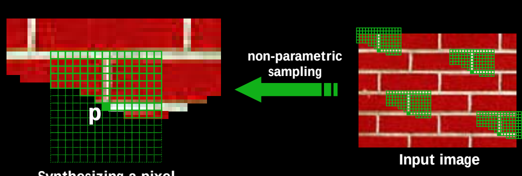


## The Challenge

- Need to model the whole spectrum: from repeated to stochastic texture



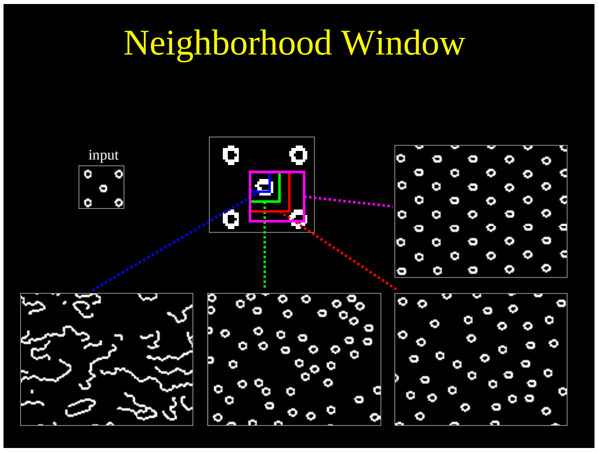

## Efros & Leung Algorithm

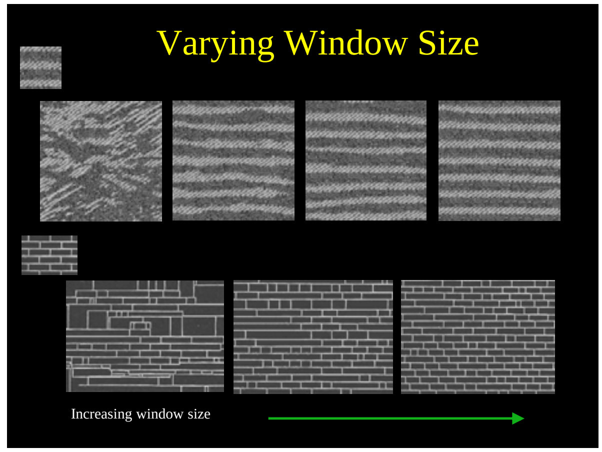


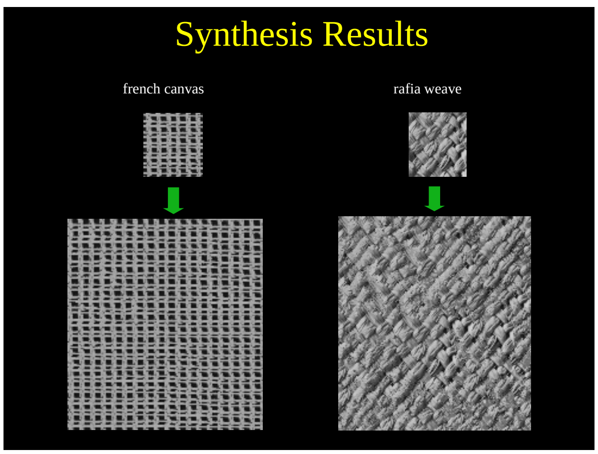

- Assuming Markov property, compute P(**p**|N(**p**))
  - Building explicit probability tables infeasible
  - Instead, we *search the input image* for all similar neighborhoods — that's our pdf for **p**
  - To sample from this pdf, just pick one match at random

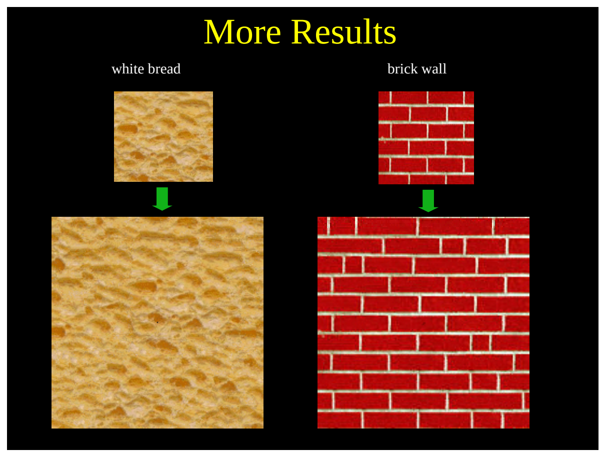
## Some Details

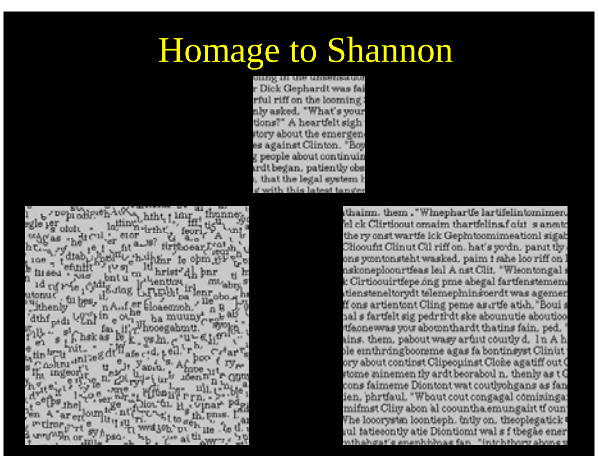
- Growing is in "onion skin" order
  - Within each "layer", pixels with most neighbors are synthesized first
  - If no close match can be found, the pixel is not synthesized until the end
- Using *Gaussian-weighted* SSD is very important
  - to make sure the new pixel agrees with its closest neighbors
  - Approximates reduction to a smaller neighborhood window if data is too sparse

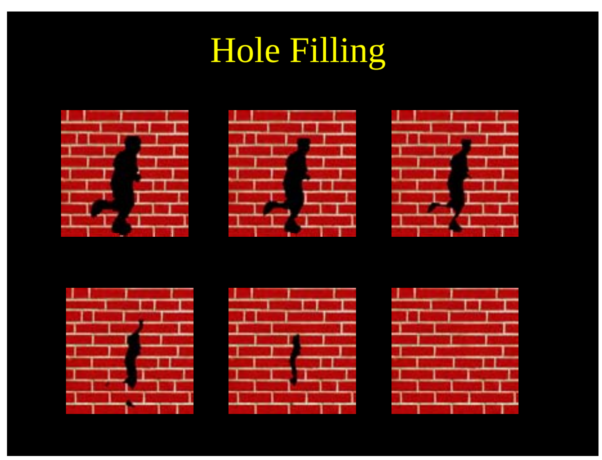

Neighborhood Window
input


Varying Window Size
Increasing window size

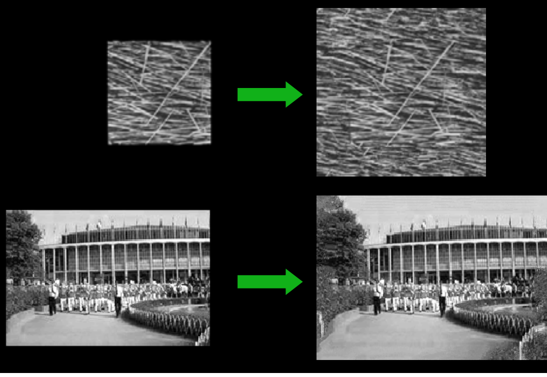

Synthesis Results
french canvas
rafia weave


More Results
white bread
brick wall


Homage to Shannon


Hole Filling

## Extrapolation



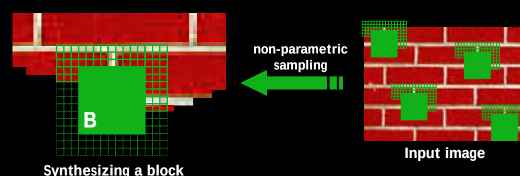
## Summary

- The Efros & Leung algorithm
  - Very simple
  - Surprisingly good results
  - Synthesis is easier than analysis!
  - …but very slow

## Overview

- Texture synthesis
- Quilting
- Image Analogies
- Super-resolution
- Scene completion

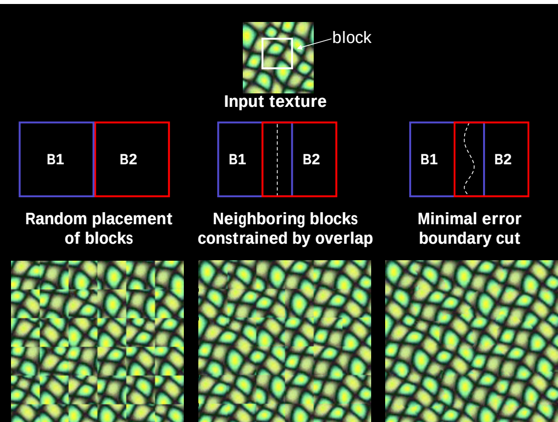
## Image Quilting [Efros & Freeman]



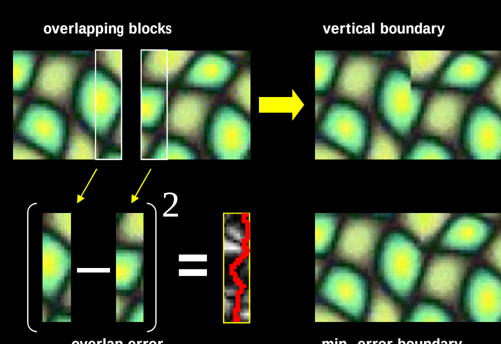

- Observation: neighbor pixels are highly correlated

**Idea: unit of synthesis = block**

- Exactly the same but now we want P(**B**|N(**B**))
- Much faster: synthesize all pixels in a block at once
- Not the same as multi-scale!

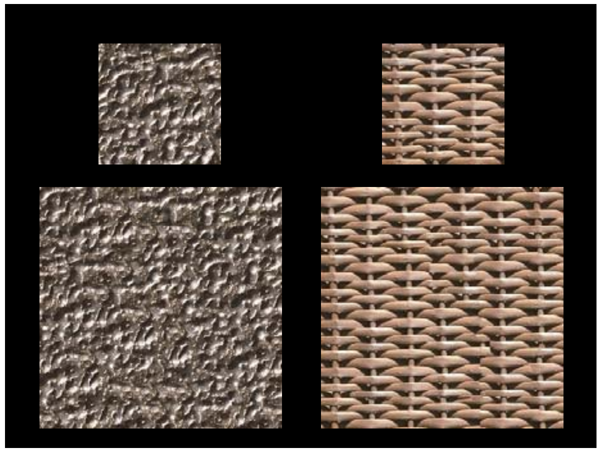


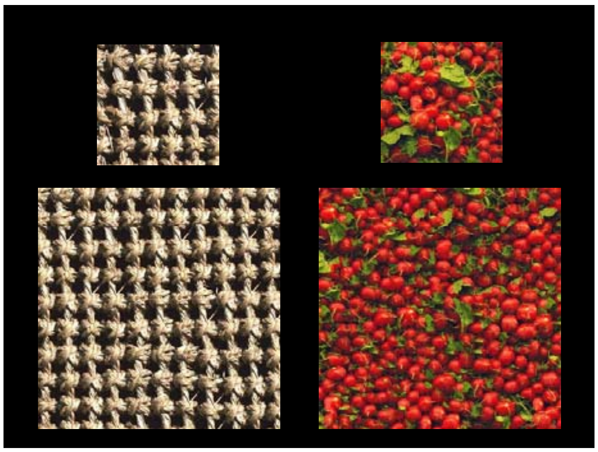

## Minimal error boundary

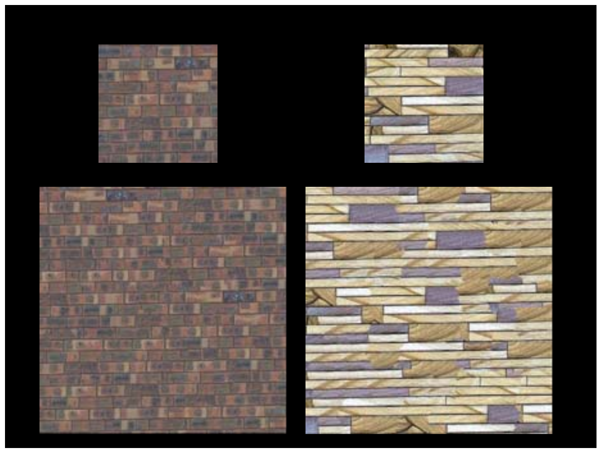


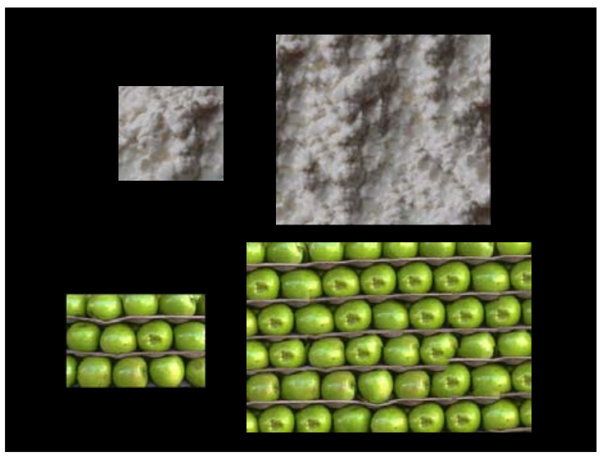
## Our Philosophy

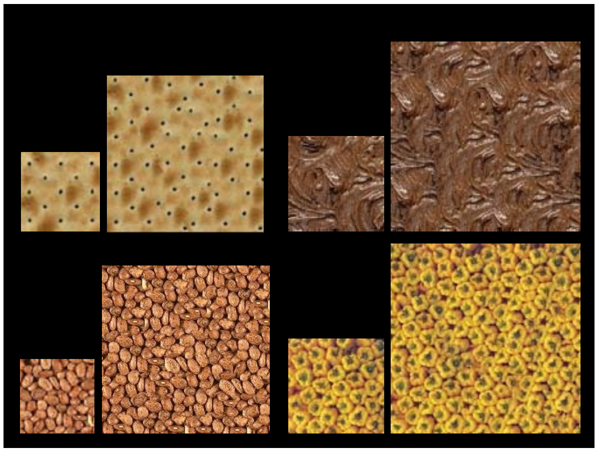
- The "Corrupt Professor's Algorithm":
  - Plagiarize as much of the source image as you can
  - Then try to cover up the evidence
- Rationale:
  - Texture blocks are by definition correct samples of texture so problem only connecting them together



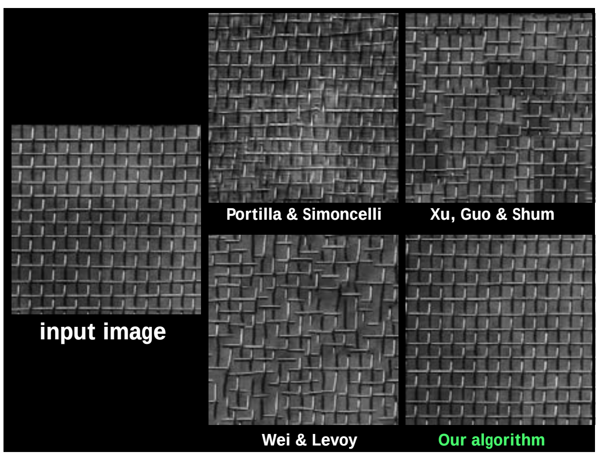


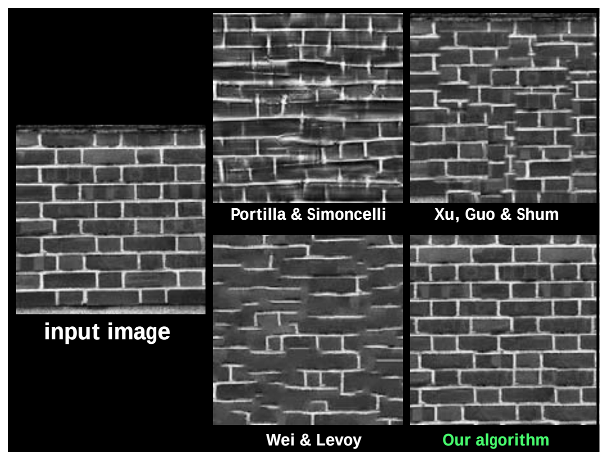


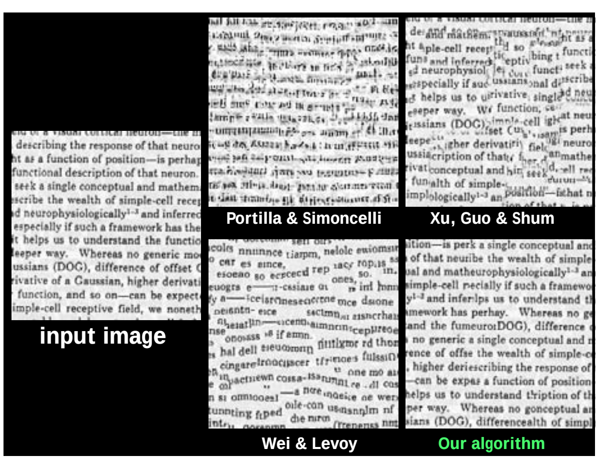


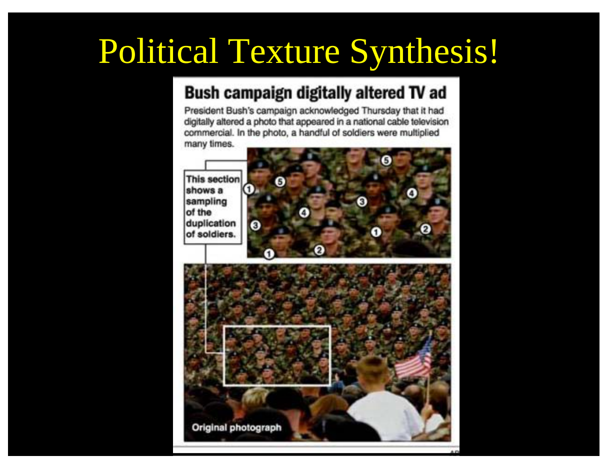



Failures
(Chernobyl Harvest)


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


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

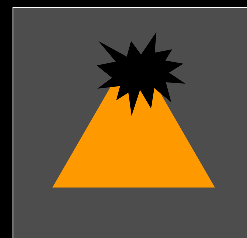

input image
Portilla & Simoncelli
Xu, Guo & Shum
Wei & Levoy
Our algorithm


Political Texture Synthesis!
Bush campaign digitally altered TV ad
President Bush's campaign acknowledged Thursday that it had digitally altered a photo that appeared in a national cable television commercial. In the photo, a handful of soldiers were multiplied many times.
This section shows a sampling of the duplication of soldiers.
Original photograph

## Fill Order

- In what order should we fill the pixels?

## Fill Order



- In what order should we fill the pixels?
  - choose pixels that have more neighbors filled
  - choose pixels that are continuations of

Criminisi, Pérez, and Toyama. "Object Removal by Exemplar-based Inpainting," Proc. CVPR, 2003.

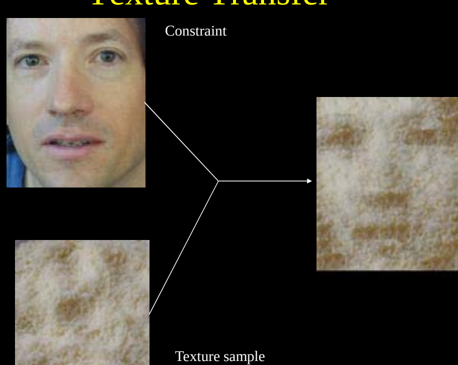
## Exemplar-based Inpainting demo

http://research.microsoft.com/vision/cambridge/i3l/patchworks.htm

## Application: Texture Transfer

- Try to explain one object with bits and pieces of another object:

## Texture Transfer

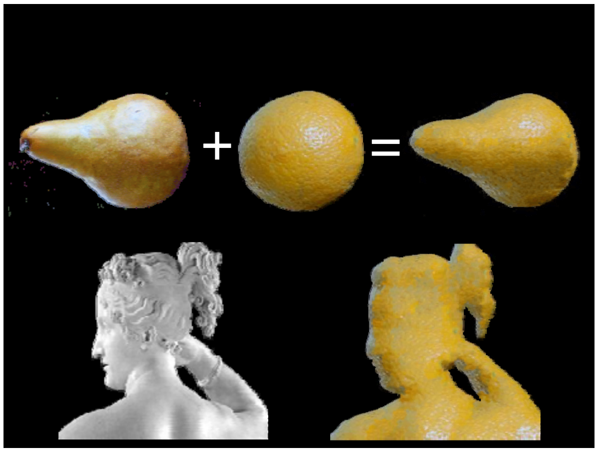



## Texture Transfer

- Take the texture from one image and "paint" it onto another object

Same as texture synthesis, except an additional constraint:

1. Consistency of texture
2. Similarity to the image being "explained"



## Overview

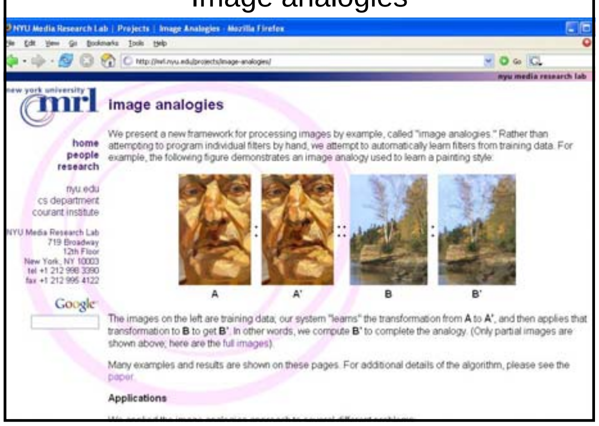
- Texture synthesis
- Quilting
- Image Analogies
- Super-resolution
- Scene completion

## Image Analogies

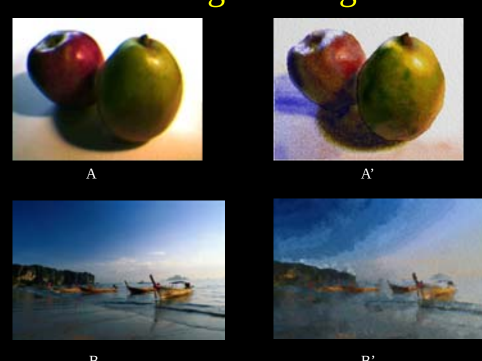

Aaron Hertzmann[1,2]
Chuck Jacobs[2]
Nuria Oliver[2]
Brian Curless[3]
David Salesin[2,3]

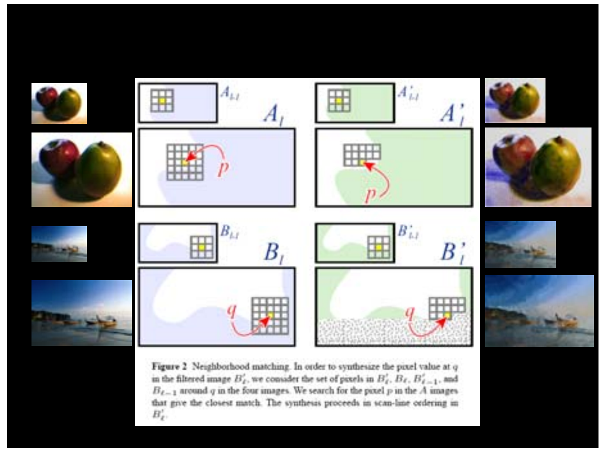
[1]New York University
[2]Microsoft Research
[3]University of Washington

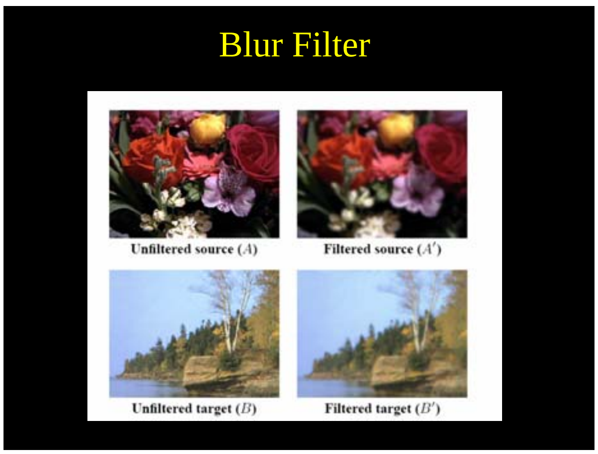

## Image analogies

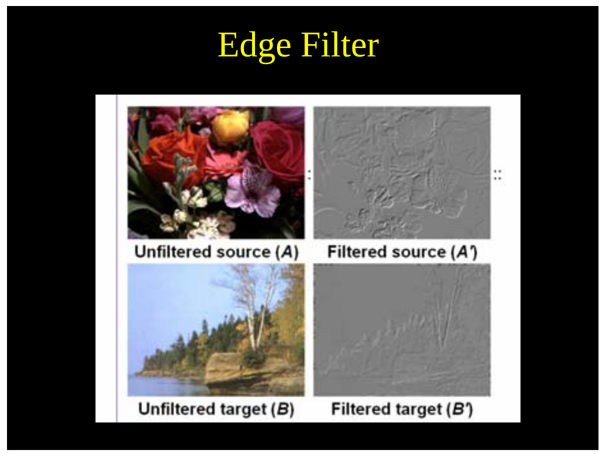


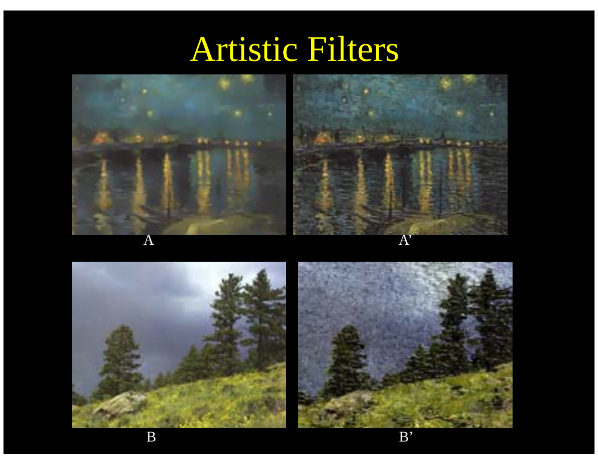

## Image Analogies

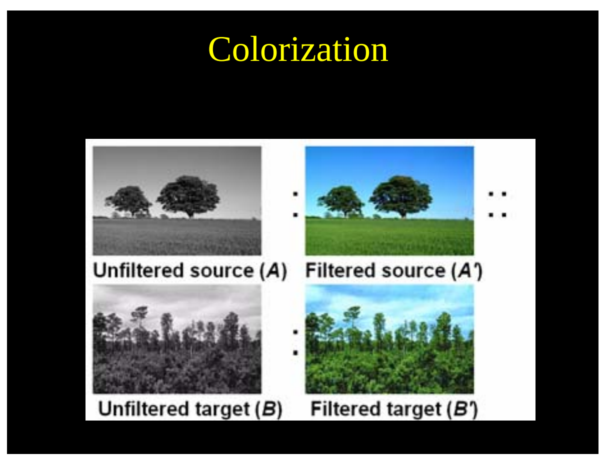



A'
B'

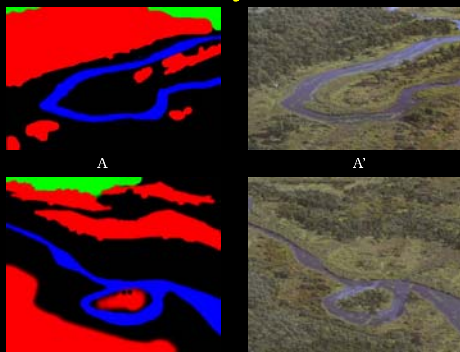

Blur Filter
Unfiltered source (A)
Filtered source (A')
Unfiltered target (B)
Filtered target (B')


Edge Filter
Unfiltered source (A)
Filtered source (A')
Unfiltered target (B)
Filtered target (B')


Artistic Filters
A
A'
B
B'


Colorization
Unfiltered source (A)
Filtered source (A')
Unfiltered target (B)
Filtered target (B')

## Texture-by-numbers




## Super-resolution


## Super-resolution (result!)


## Overview

- Texture synthesis
- Quilting
- Image Analogies
- Super-resolution
- Scene completion

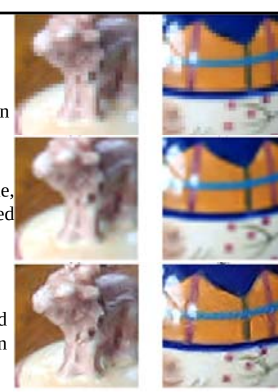
## Super-resolution

- Image: low resolution image
- Scene: high resolution image

ultimate goal...


Slides from Bill Freeman

Pixel-based images are not resolution independent



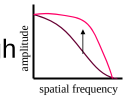
Pixel replication

Cubic spline, sharpened

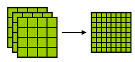
Training-based super-resolution

## 3 approaches to perceptual sharpening

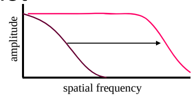
(1) Sharpening; boost existing high frequencies.




(2) Use multiple frames to obtain higher sampling rate in a still frame.



(3) Estimate high frequencies not present in image, although implicitly defined.




**In this talk, we focus on (3), which we'll call "super-resolution".**

## Super-resolution: other approaches

- Schultz and Stevenson, 1994
- Pentland and Horowitz, 1993
- fractal image compression (Polvere, 1998; Iterated Systems)
- astronomical image processing (eg. Gull and Daniell, 1978; "pixons" http://casswww.ucsd.edu/puetter.html)

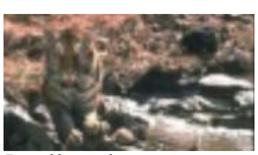
## Training images, ~100,000 image/scene patch pairs

Images from two Corel database categories: "giraffes" and "urban skyline".

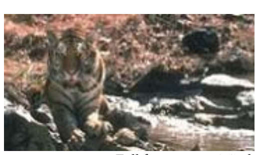
## Do a first interpolation

Zoomed low-resolution

Low-resolution



Zoomed low-resolution



Full frequency original

Low-resolution

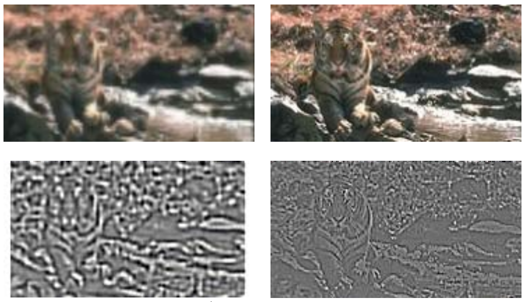
## Representation

Zoomed low-freq.

Full freq. original

# Representation

Zoomed low-freq. Full freq. original

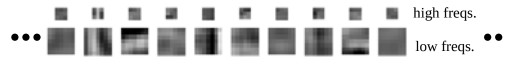


Low-band input (contrast normalized, PCA fitted)

True high freqs

(to minimize the complexity of the relationships we have to learn, we remove the lowest frequencies from the input image, and normalize the local contrast level).

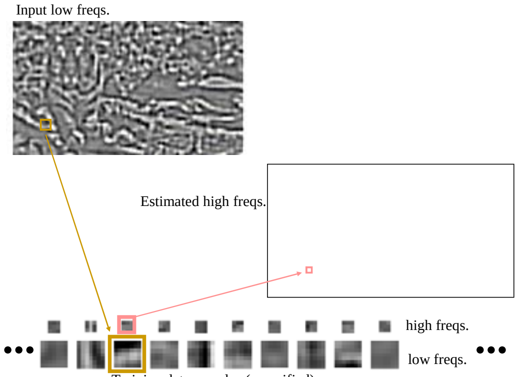
# Gather ~100,000 patches




Training data samples (magnified)

# Nearest neighbor estimate

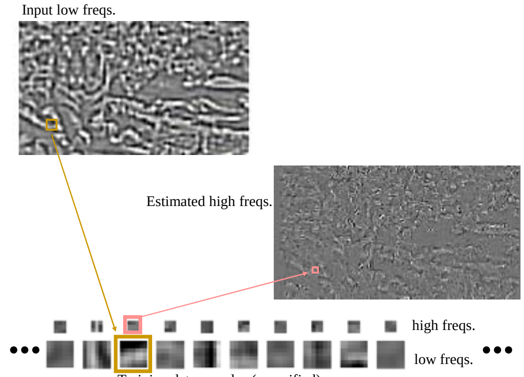
Input low freqs.




Training data samples (magnified)

# Nearest neighbor estimate

Input low freqs.

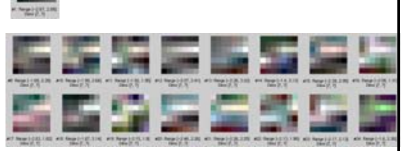



Training data samples (magnified)

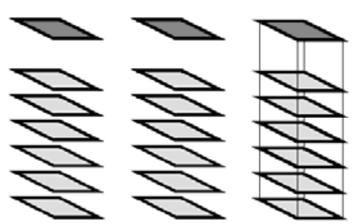
# Example: input image patch, and closest matches from database

Input patch

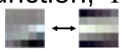
Closest image patches from database



Corresponding high-resolution patches from database

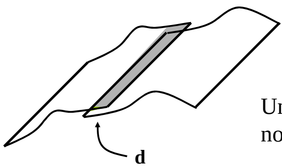


## Scene-scene compatibility function, $\Psi(x_i, x_j)$



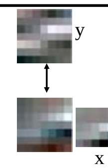
Assume overlapped regions, d, of hi-res. patches differ by Gaussian observation noise:

$$\Psi(x_i, x_j) = \exp^{-|d_i - d_j|^2 / 2\sigma^2}$$




Uniqueness constraint, not smoothness.

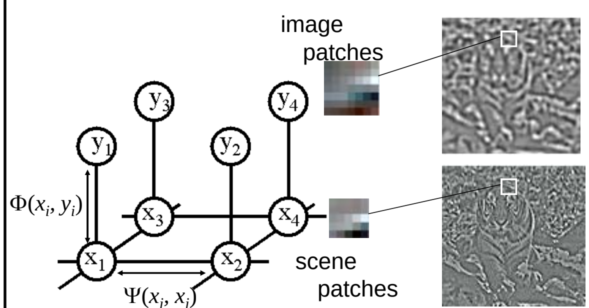
## Image-scene compatibility function, $\Phi(x_i, y_i)$



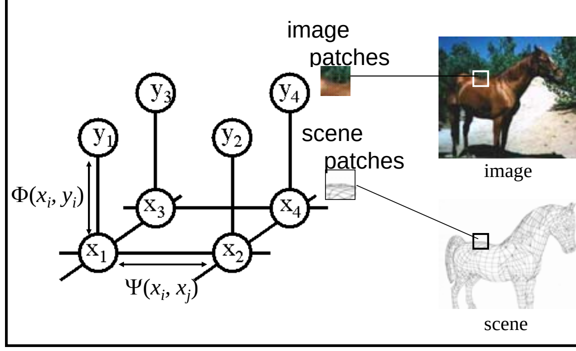

Assume Gaussian noise takes you from observed image patch to synthetic sample:

$$\Phi(x_i, y_i) = \exp^{-|y_i - y(x_i)|^2 / 2\sigma^2}$$

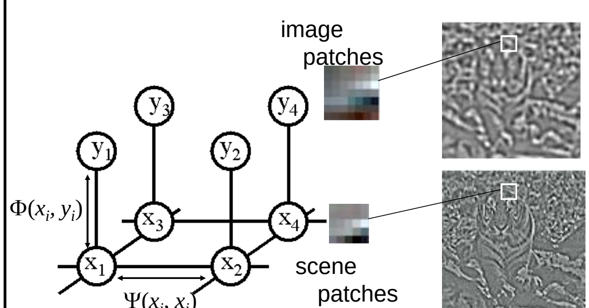
## Markov network




## VISTA-- Vision by Image-Scene TrAining

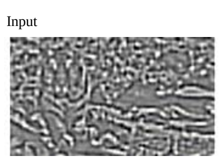


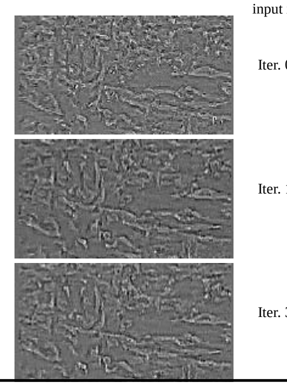

## Super-resolution application




## Belief Propagation

After a few iterations of belief propagation, the algorithm selects spatially consistent high resolution interpretations for each low-resolution patch of the input image.



## Zooming 2 octaves

We apply the super-resolution algorithm recursively, zooming up 2 powers of 2, or a factor of 4 in each dimension.

85 x 51 input

Cubic spline zoom to 340x204

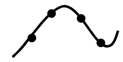
Max. likelihood zoom to 340x204

---

Now we examine the effect of the prior assumptions made about images on the high resolution reconstruction. First, cubic spline interpolation.

Original 50x58

(cubic spline implies thin plate prior)

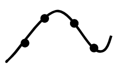


True 200x232

---

Original 50x58

(cubic spline implies thin plate prior)



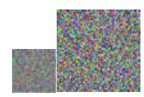
Cubic spline

True 200x232

---

Next, train the Markov network algorithm on a world of random noise images.

Original 50x58



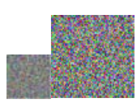
Training images

True

---

The algorithm learns that, in such a world, we add random noise when zoom to a higher resolution.

Original 50x58



Training images

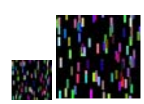
Markov network

True

---

Next, train on a world of vertically oriented rectangles.

Original 50x58

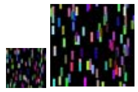


Training images

True

The Markov network algorithm hallucinates those vertical rectangles that it was trained on.

Original 50x58



Training images

Markov network

True

---

Now train on a generic collection of images.

Original 50x58

Training images

True

---

The algorithm makes a reasonable guess at the high resolution image, based on its training images.

Original 50x58

Training images

Markov network

True

---

## Generic training images

Next, train on a generic set of training images. Using the same camera as for the test image, but a random collection of photographs.

---

Original 70x70

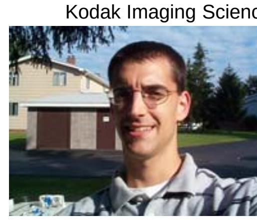
Cubic Spline

Markov net, training: generic

True 280x280

---

Kodak Imaging Science Technology Lab test.

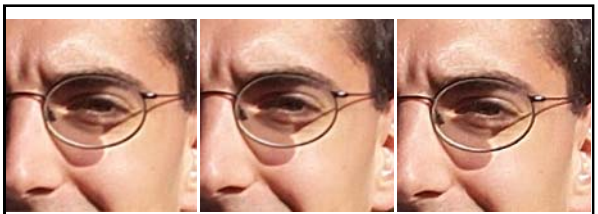


3 test images, 640x480, to be zoomed up by 4 in each dimension.

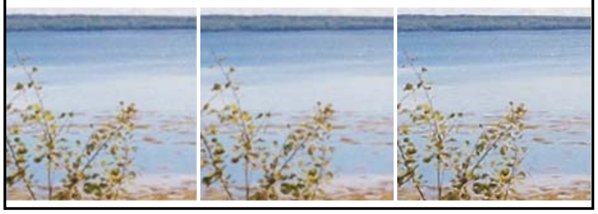
8 judges, making 2-alternative, forced-choice comparisons.

## Algorithms compared

- Bicubic Interpolation
- Mitra's Directional Filter
- Fuzzy Logic Filter
- Vector Quantization
- VISTA



Bicubic spline　Altamira　VISTA



Bicubic spline　Altamira　VISTA

## User preference test results

"The observer data indicates that six of the observers ranked Freeman's algorithm as the most preferred of the five tested algorithms. However the other two observers rank Freeman's algorithm as the least preferred of all the algorithms….

Freeman's algorithm produces prints which are by far the sharpest out of the five algorithms. However, this sharpness comes at a price of artifacts (spurious detail that is not present in the original scene). Apparently the two observers who did not prefer Freeman's algorithm had strong objections to the artifacts. The other observers apparently placed high priority on the high level of sharpness in the images created by Freeman's algorithm."

## Conclusions

- Exemplars (local, non-parametric image representations) are useful, fun, easy-to-use.
- Requirement: find ways to get by with too few exemplars.

## Overview

- Texture synthesis
- Quilting
- Image Analogies
- Super-resolution
- Scene completion

# Scene Completion Using Millions of Photographs

James Hays and Alexei A. Efros
Carnegie Mellon University

Efros and Leung result

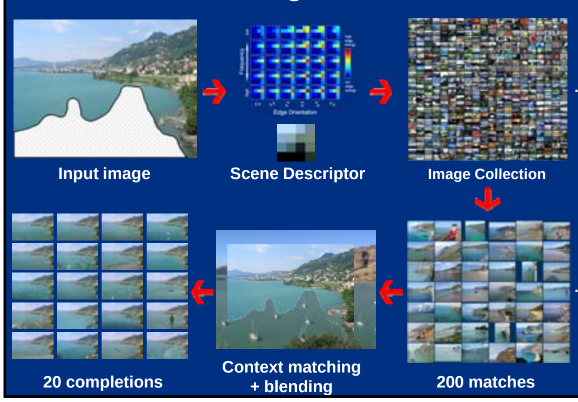
Criminisi et al. result

Criminisi et al. result

## Scene Matching for Image Completion


Scene Completion Result

## The Algorithm

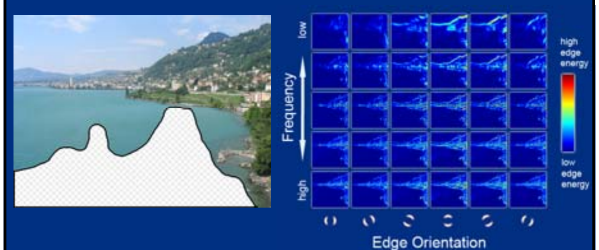


## Data

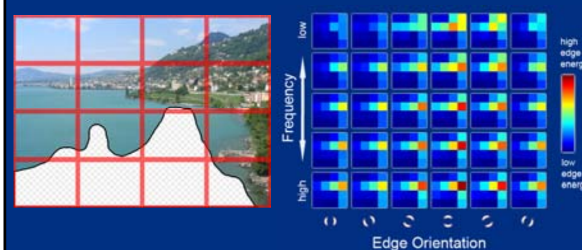
We downloaded **2.3 Million** unique images from Flickr groups and keyword searches.

## Scene Matching

## Scene Descriptor

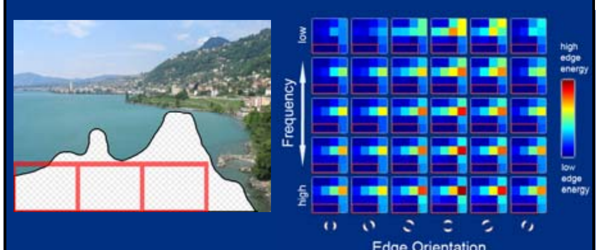



## Scene Descriptor



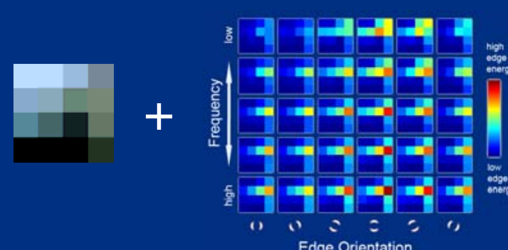

Gist scene descriptor
(Oliva and Torralba 2001)

## Scene Descriptor




Gist scene descriptor
(Oliva and Torralba 2001)

## Scene Descriptor




Gist scene descriptor
(Oliva and Torralba 2001)

… 200 total

## Context Matching

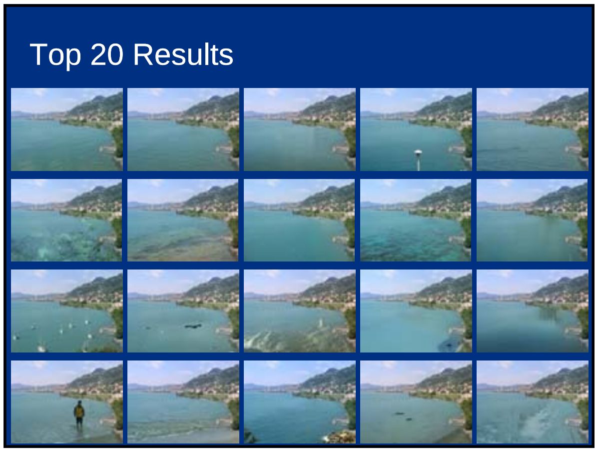
## Result Ranking

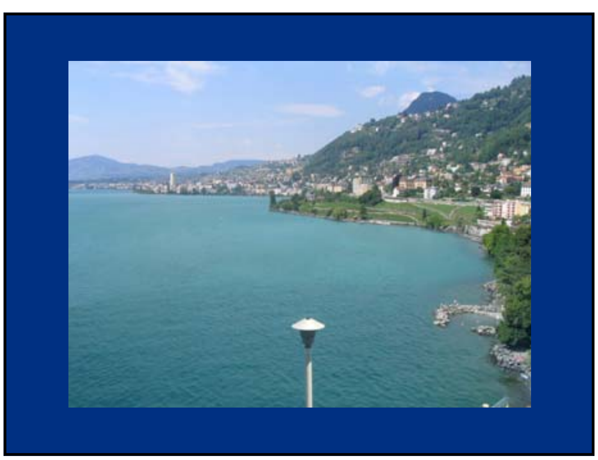
We assign each of the 200 results a score which is the sum of:

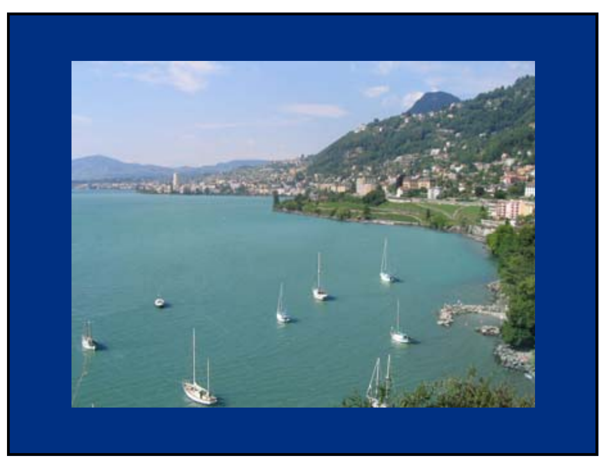
The scene matching distance

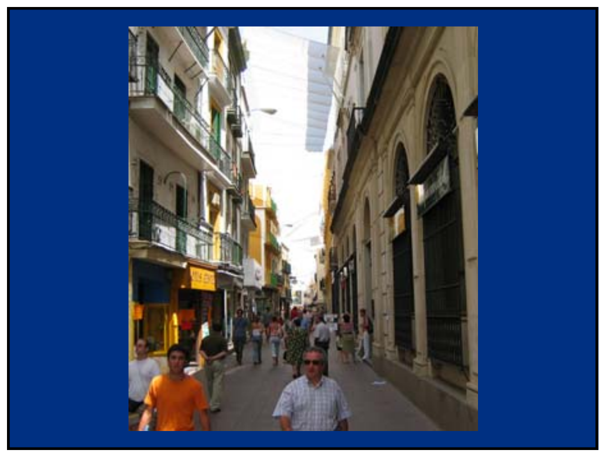
The context matching distance (color + texture)

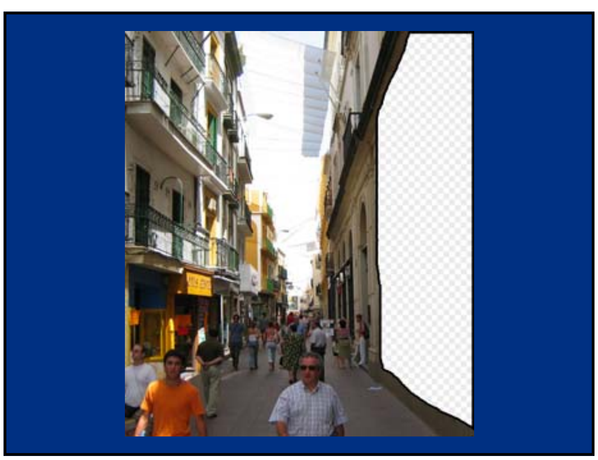
The graph cut cost

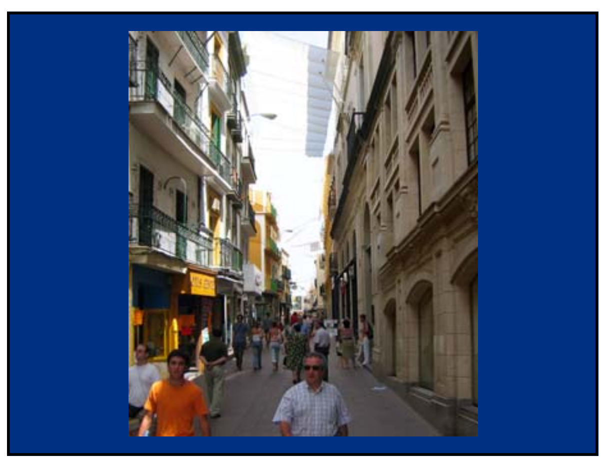

Top 20 Results





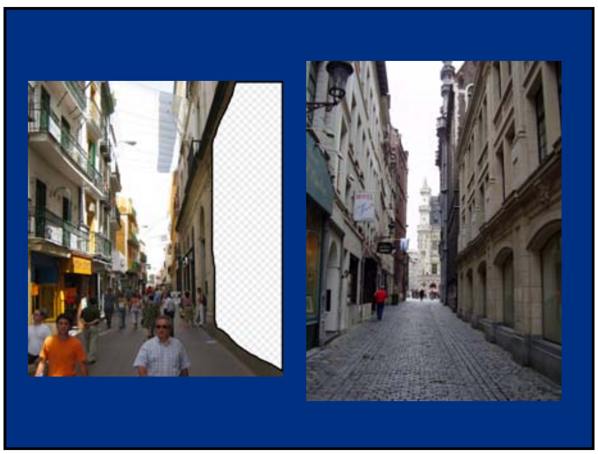


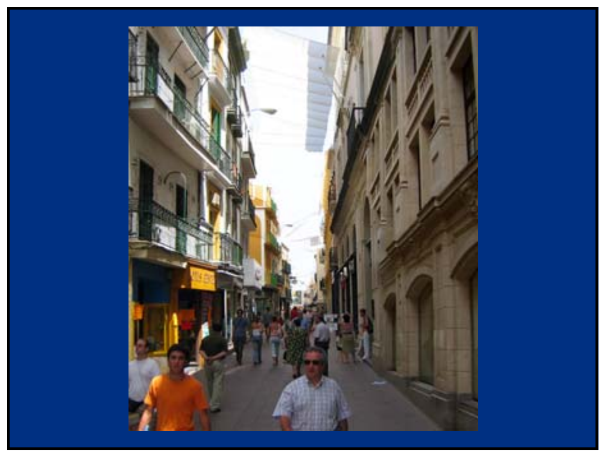


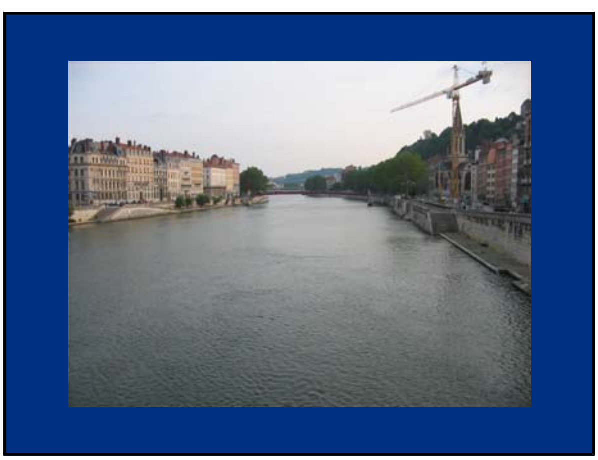


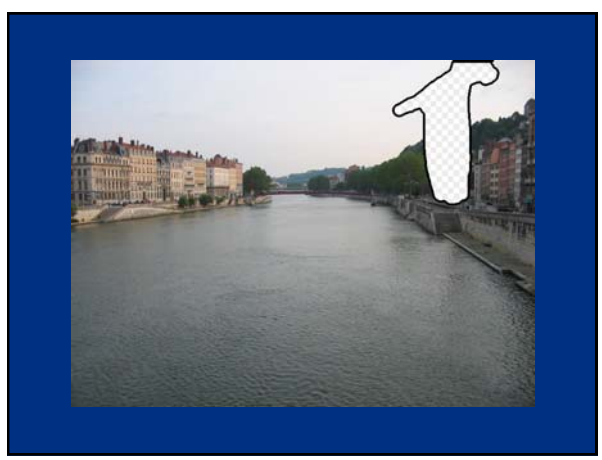

… 200 scene matches

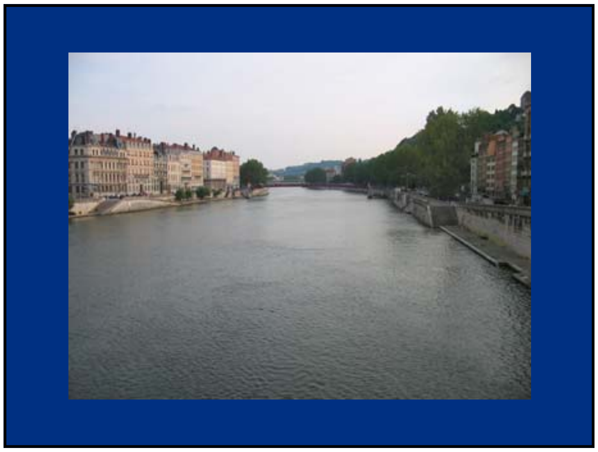

… 200 scene matches

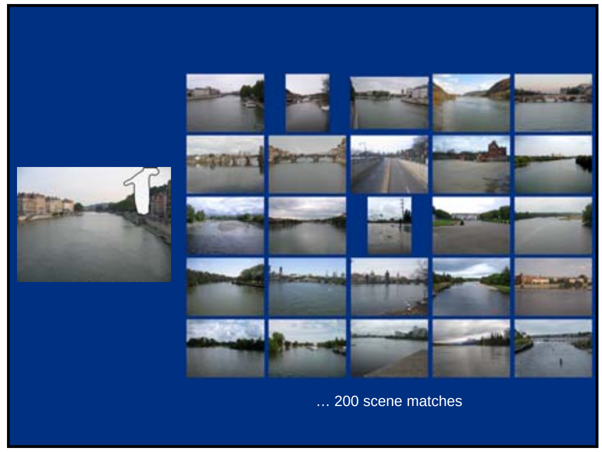


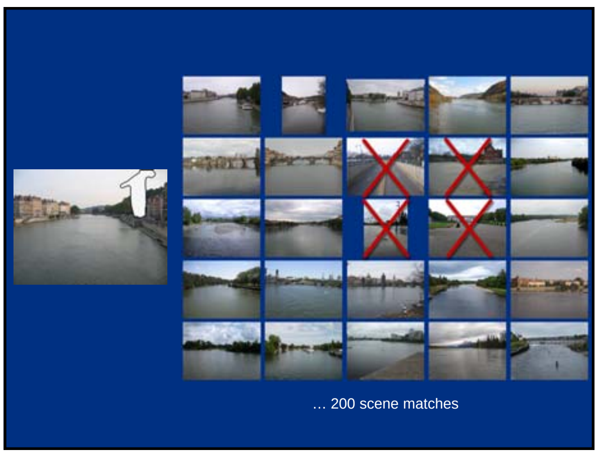


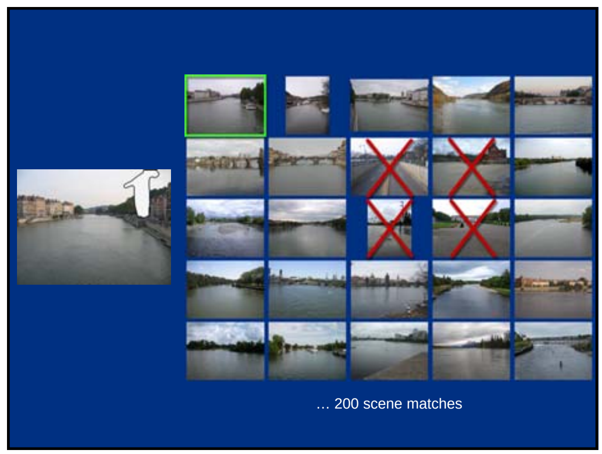


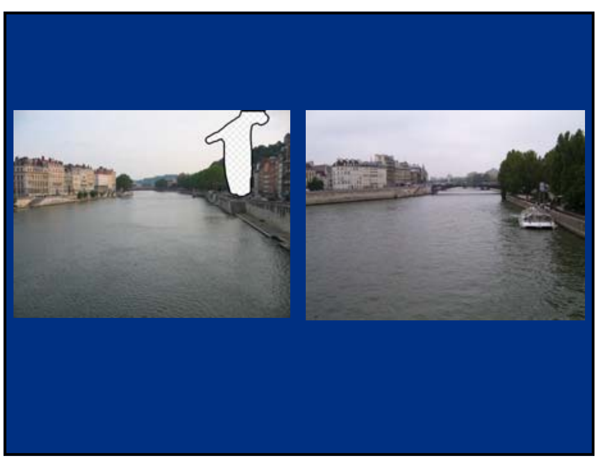


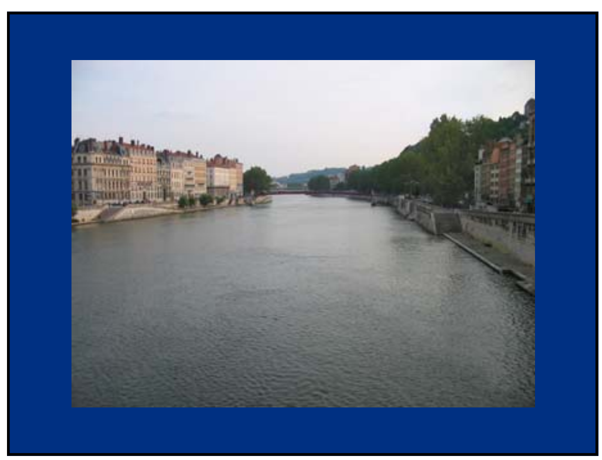


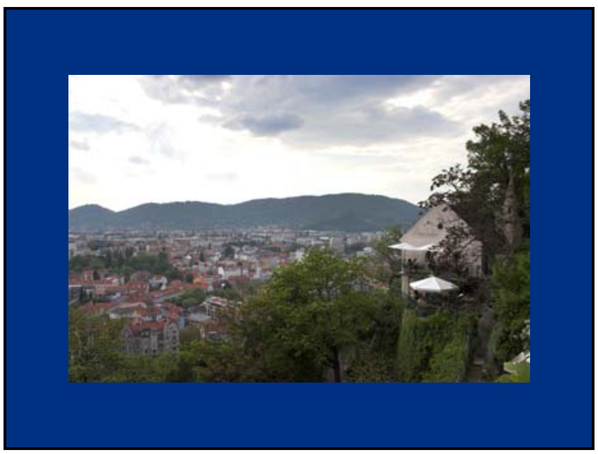

… 200 scene matches

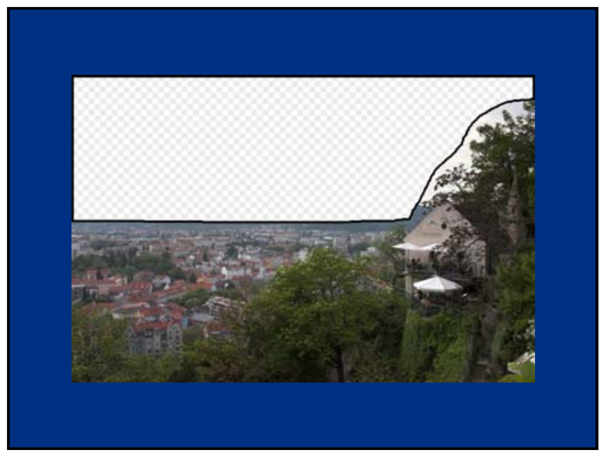

… 200 scene matches

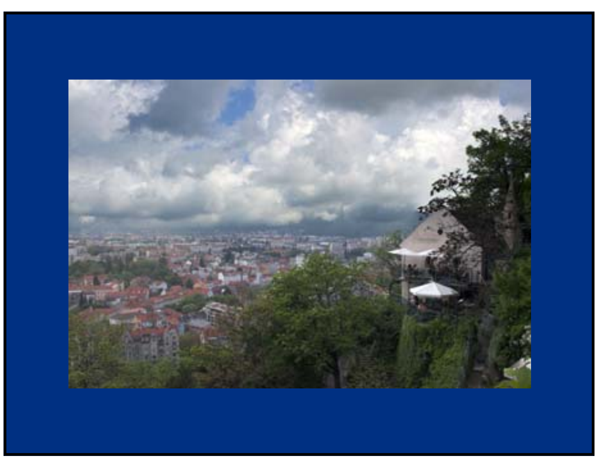

… 200 scene matches

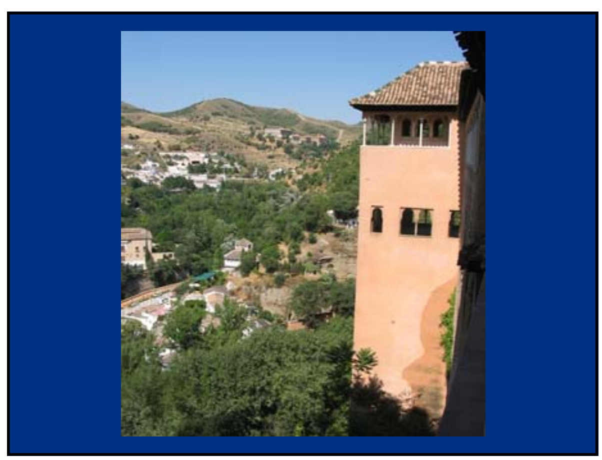


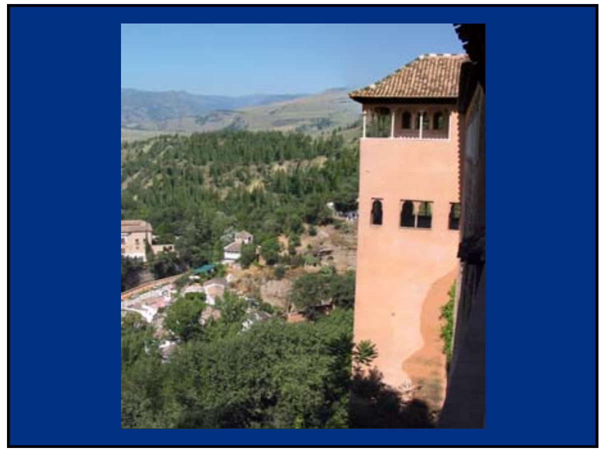


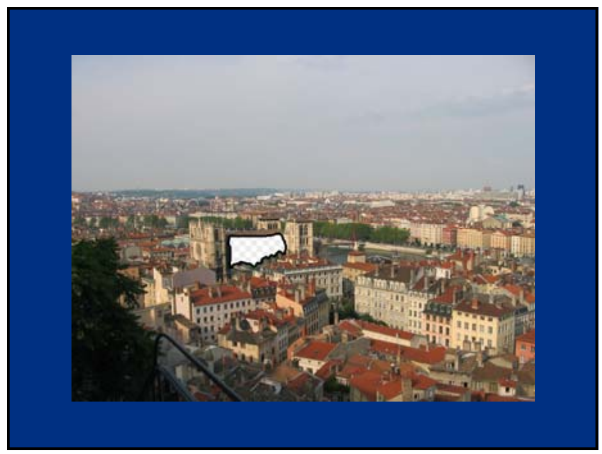


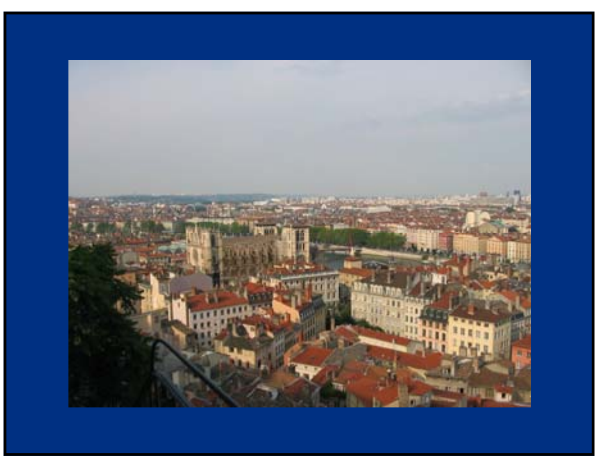






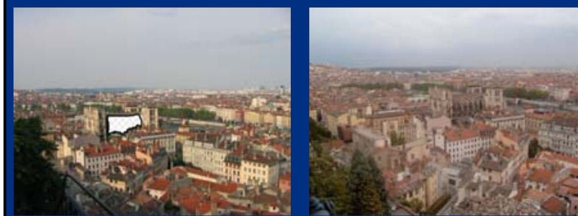







… 200 scene matches

… 200 scene matches



## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Failures

## Evaluation

Original Images

Criminisi et al.

Single result

Scene Completion

Each result
selected from 20

Original Images

Criminisi et al.

Single result

Scene Completion

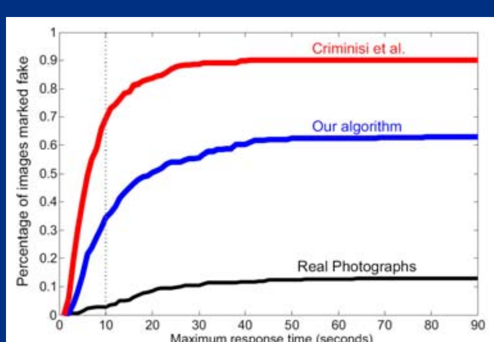
Each result selected from 20

Real Image. This image has not been manipulated

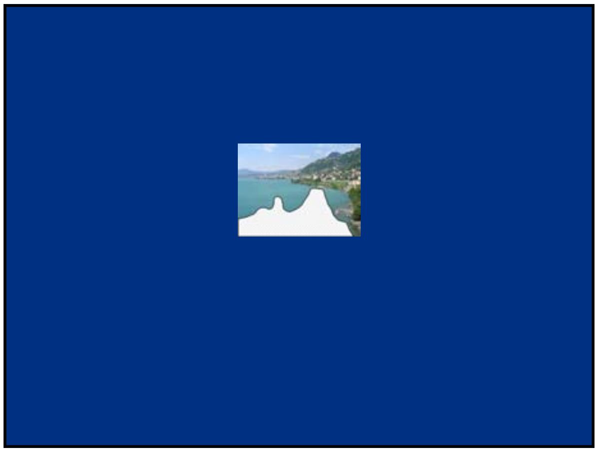
or

Fake Image. This image has been manipulated

## User Study Results - 20 Participants

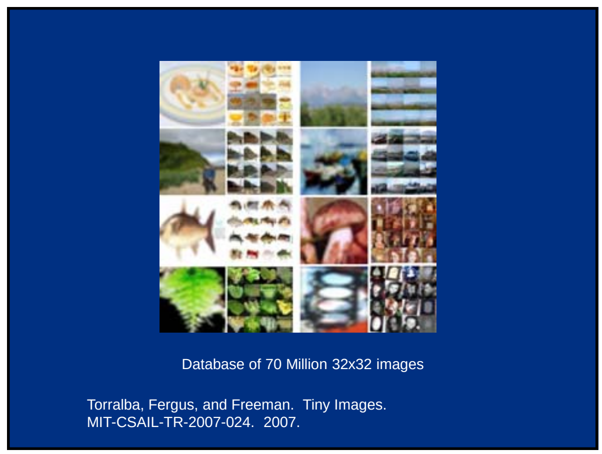



Why does it work?




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.