# CS195: Computer Vision

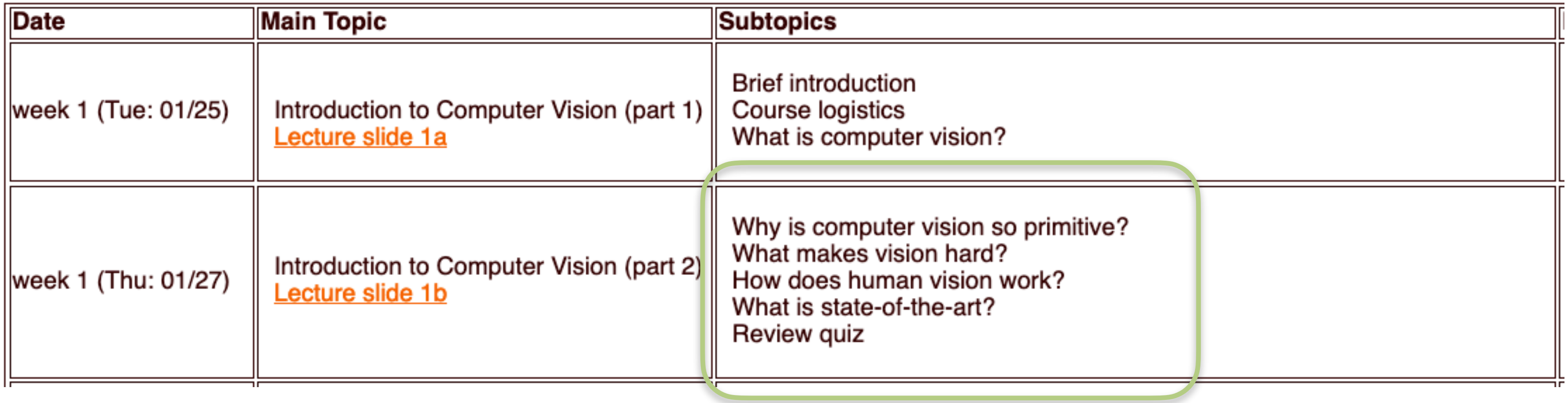
January 27, 2022

Md Alimoor Reza
Assistant Professor of Computer Science
Drake University

Drake
UNIVERSITY

# Road Map

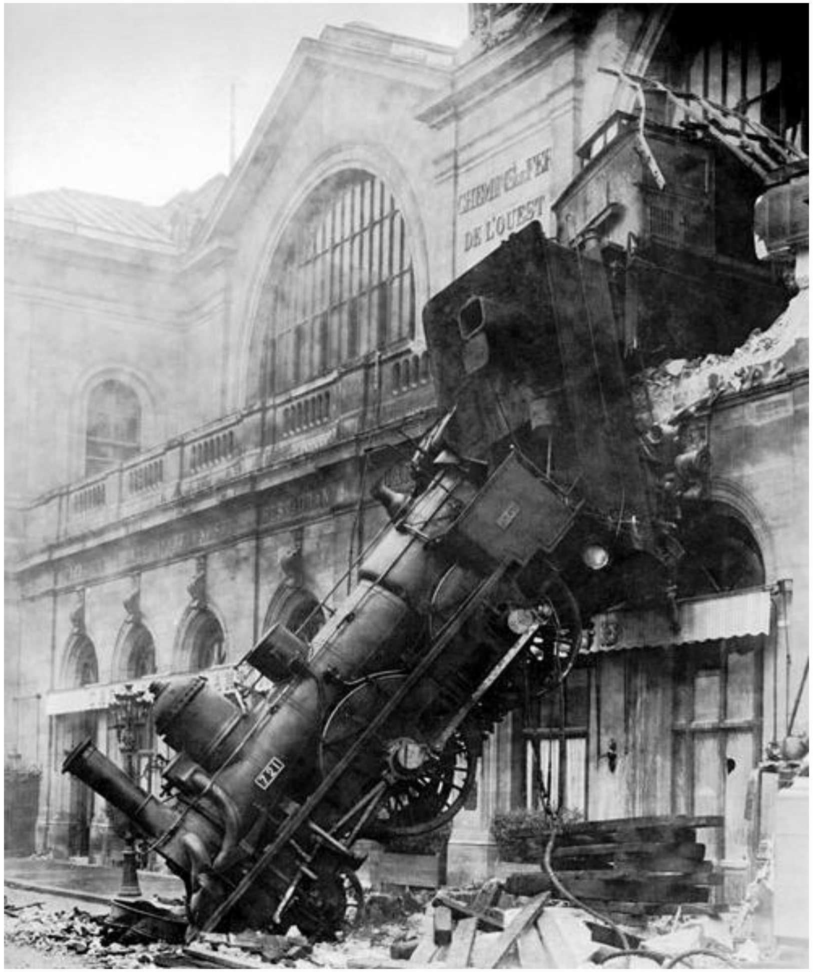
- Topics
  - Why is computer vision so primitive?
  - What makes vision hard?
  - How does human vision work?
  - Recent progress



| Date | Main Topic | Subtopics |
|---|---|---|
| week 1 (Tue: 01/25) | Introduction to Computer Vision (part 1)<br>Lecture slide 1a | Brief introduction<br>Course logistics<br>What is computer vision? |
| week 1 (Thu: 01/27) | Introduction to Computer Vision (part 2)<br>Lecture slide 1b | Why is computer vision so primitive?<br>What makes vision hard?<br>How does human vision work?<br>What is state-of-the-art?<br>Review quiz |

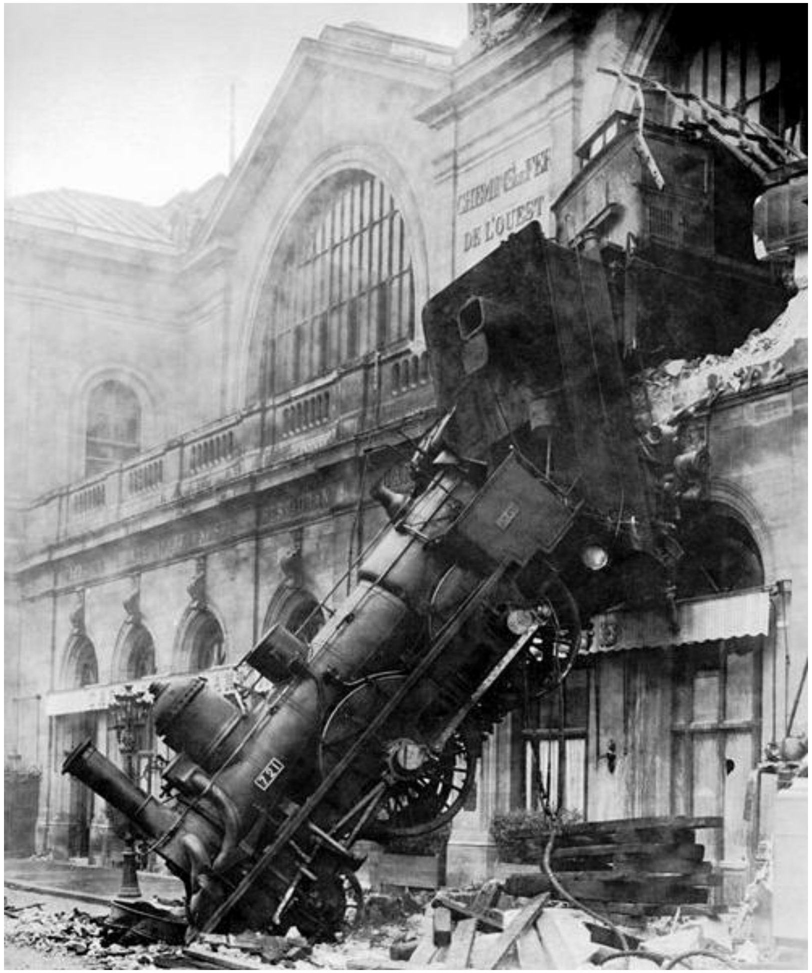
# What is computer vision?

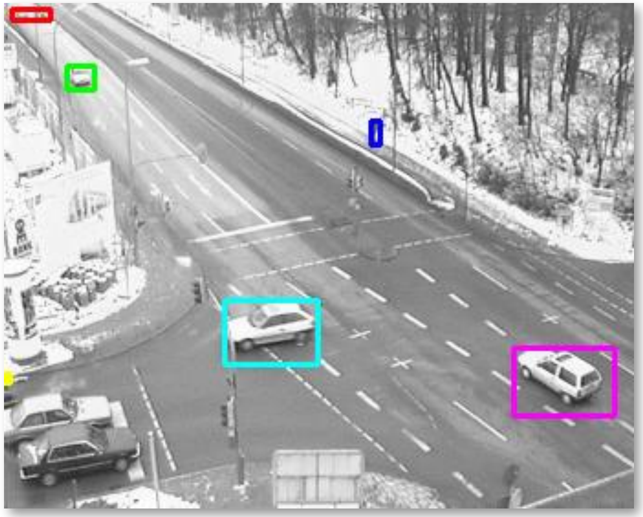
# Goal: from images to meaning



# Goal: from images to meaning

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 2 | 5 | 4 | 7 | 6 | 9 | 8 |
| 3 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 1 | 0 | 3 | 2 | 5 | 4 | 7 | 6 |
| 5 | 2 | 3 | 0 | 1 | 2 | 3 | 4 | 5 |
| 4 | 3 | 2 | 1 | 0 | 3 | 2 | 5 | 4 |
| 7 | 4 | 5 | 2 | 3 | 0 | 1 | 2 | 3 |
| 6 | 5 | 4 | 3 | 2 | 1 | 0 | 3 | 2 |
| 9 | 6 | 7 | 4 | 5 | 2 | 3 | 0 | 1 |
| 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |





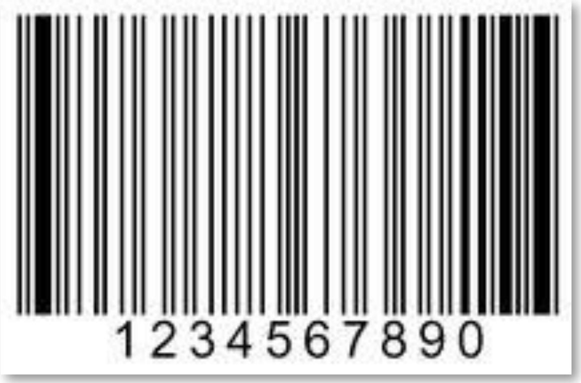
flickr® from YAHOO!

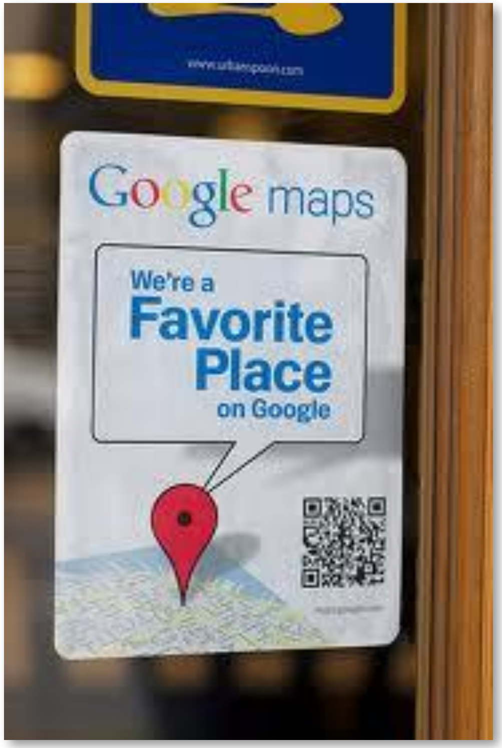
YouTube

# Can computers see as well as humans?

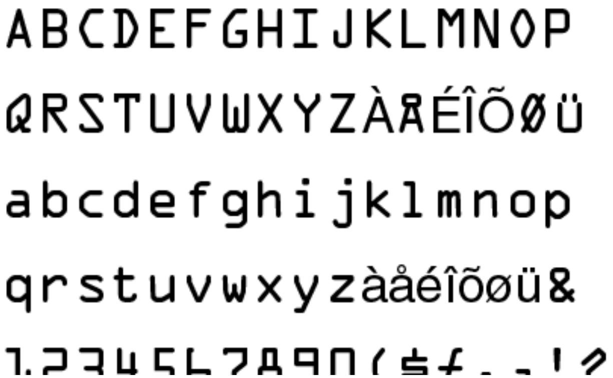
- Yes and no, but mostly no (so far).
- Current vision technology is useful in select applications, with:
  - Specific, constrained environments, and/or
  - High tolerance for errors

# The most successful and ubiquitous application of computer vision … ?

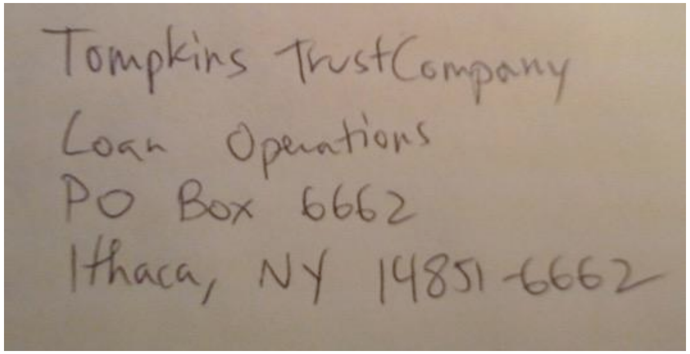




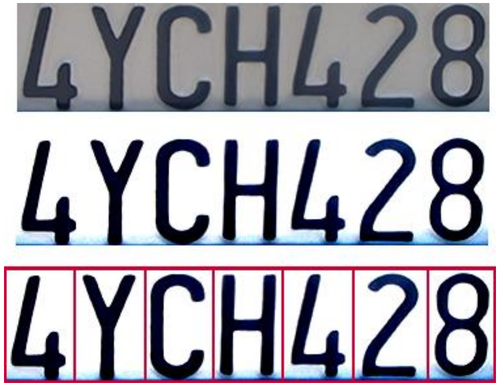
# Optical character recognition (OCR)


ABCDEFGHIJKLMNOP
QRSTUVWXYZÀÅÉÎÕØÜ
abcdefghijklmnop
qrstuvwxyzàåéîõøü&
1234567890($£.,-!?

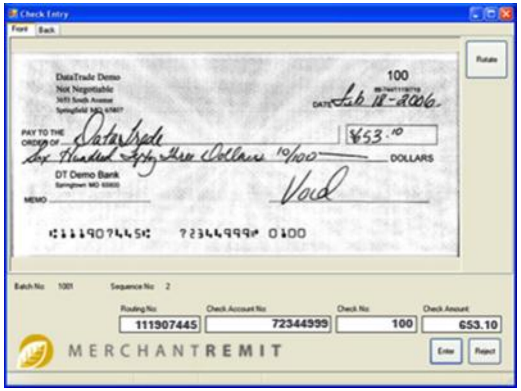
Document digitization



Postal address recognition



License plate readers



DataTrade Demo
Not Negotiable
3651 South Avenue
Springfield MO 65807

100

DATE Feb 18-2006

PAY TO THE ORDER OF DataTrade

$653.10

Six Hundred Fifty Three Dollars 10/100 DOLLARS

DT Demo Bank
Springtown MO 65800

MEMO

Void

⑆111907445⑆ 72344999⑈ 0100

Batch No 1001 Sequence No 2

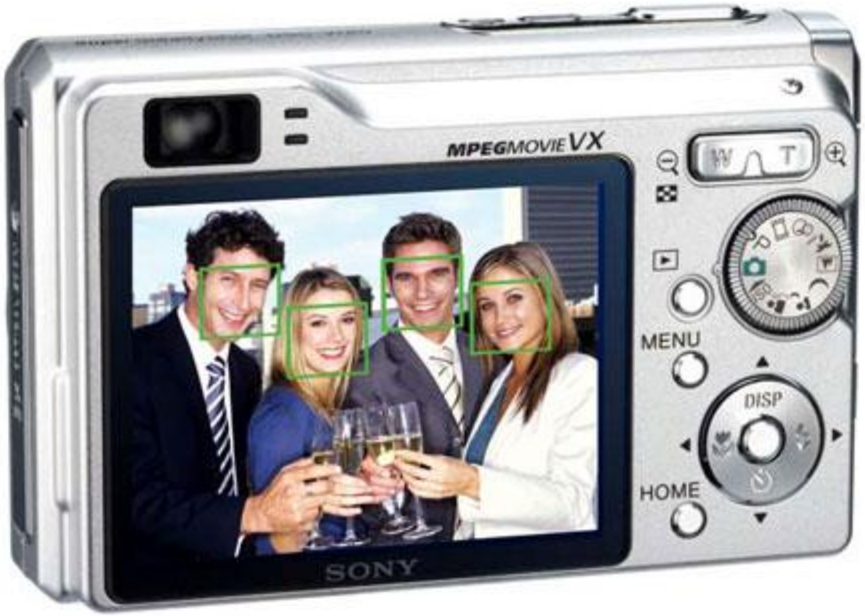
| Routing No | Check Account No | Check No | Check Amount |
|---|---|---|---|
| 111907445 | 72344999 | 100 | 653.10 |

Automatic check processing

Source: S. Seitz

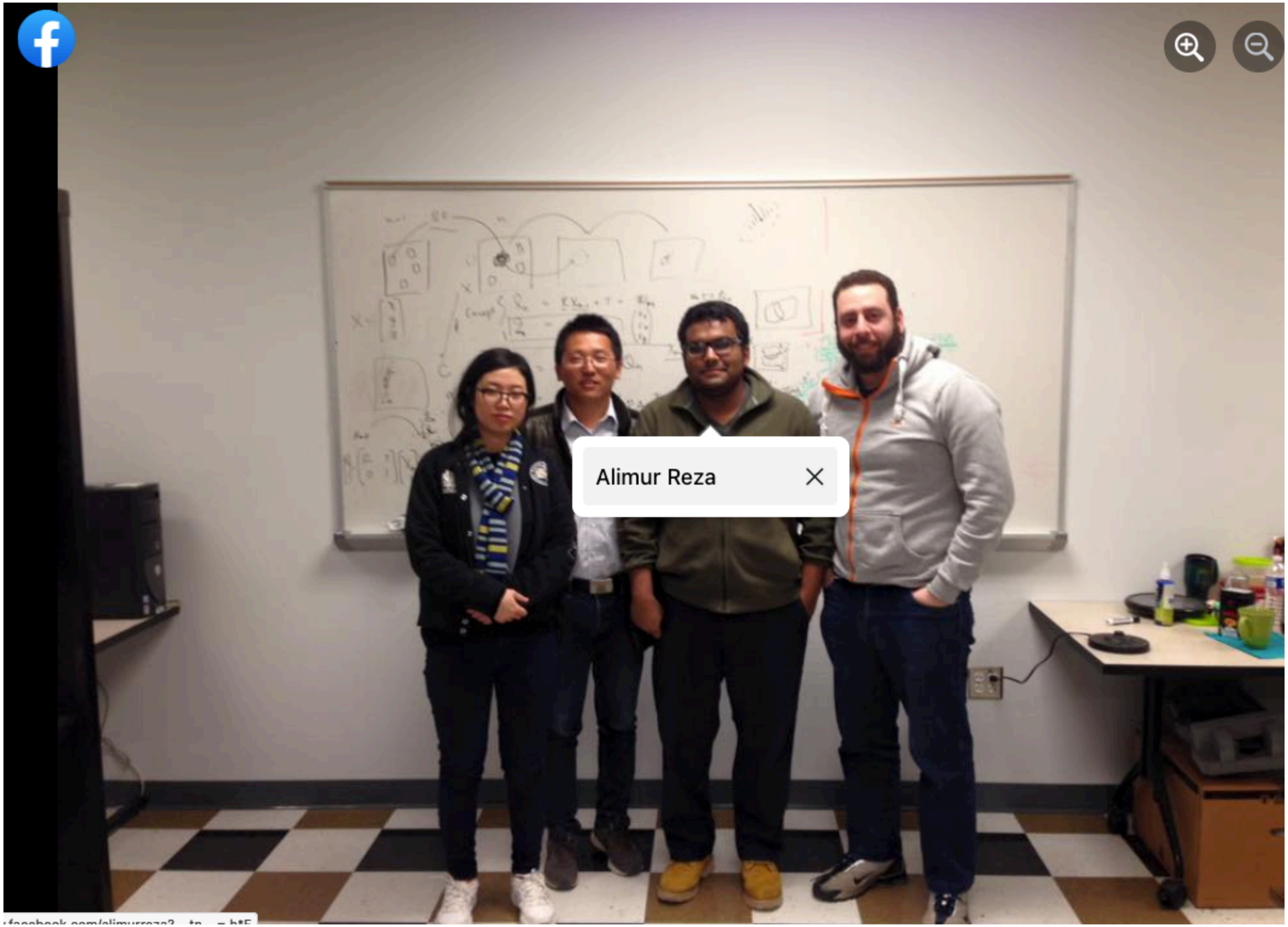
# Industrial inspection
# (aka Machine Vision)

# Face detection

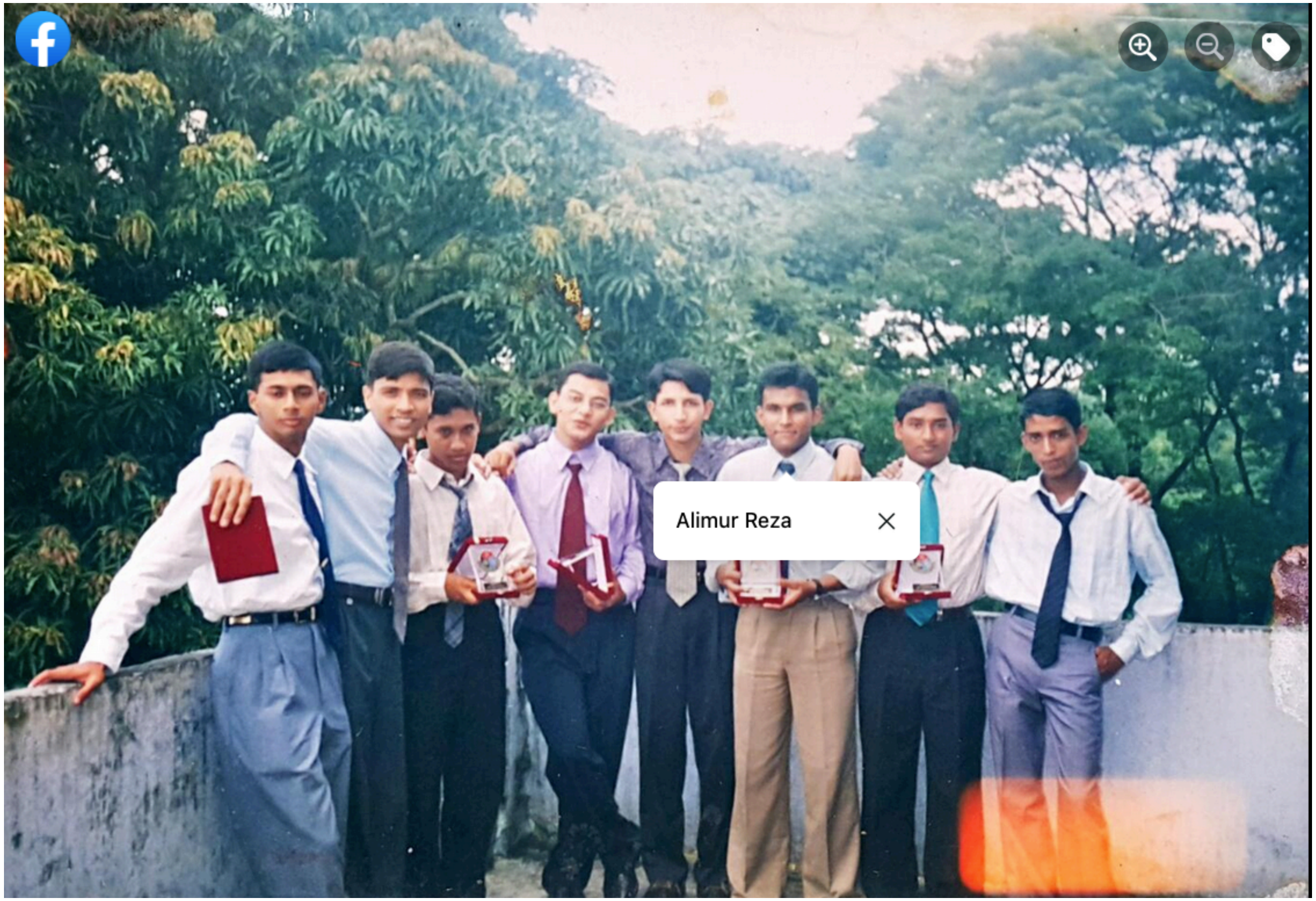


Source: S. Seitz

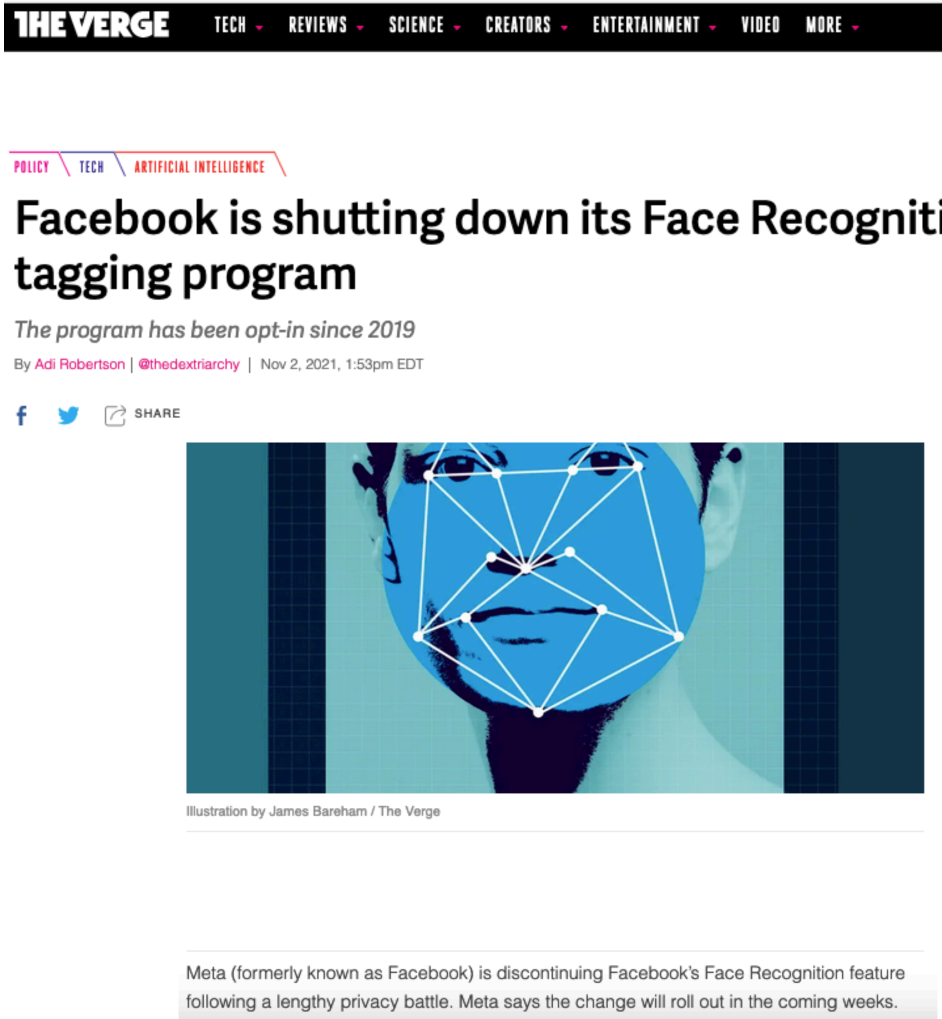
# Facebook's face detection



# Facebook’s face detection



# Facebook's face detection



## Facebook is shutting down its Face Recognition tagging program

*The program has been opt-in since 2019*

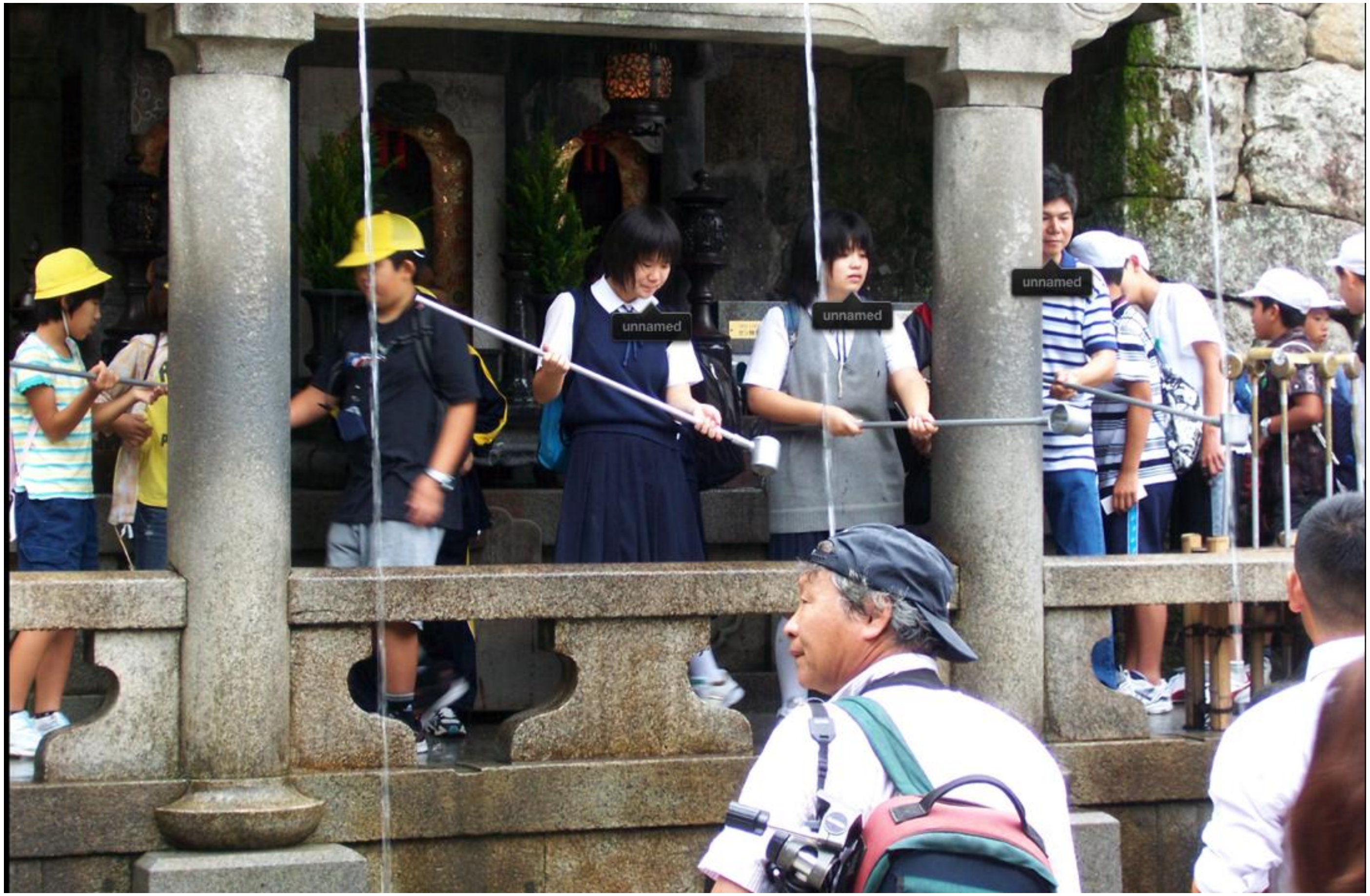
By Adi Robertson | @thedextriarchy | Nov 2, 2021, 1:53pm EDT

Illustration by James Bareham / The Verge

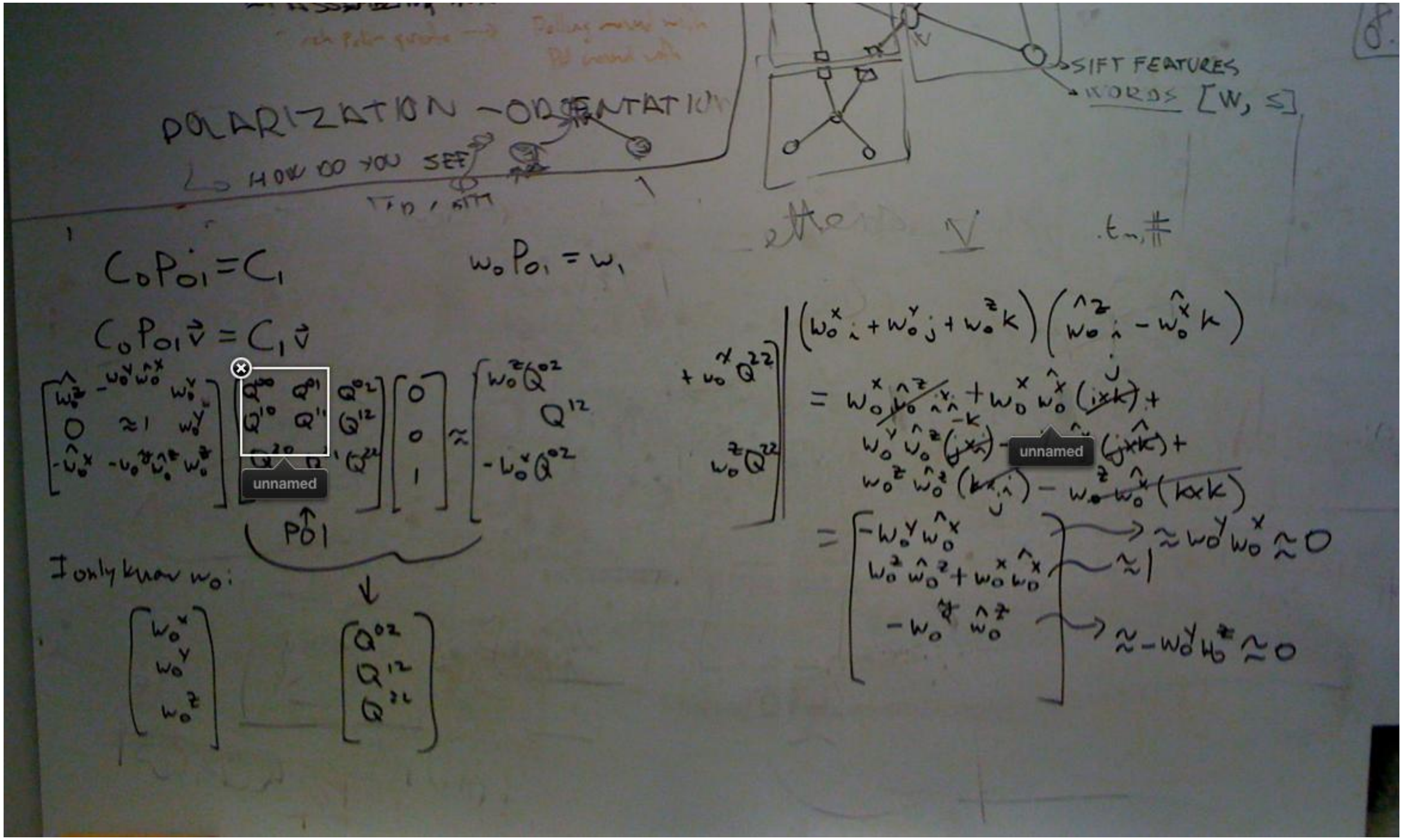
Meta (formerly known as Facebook) is discontinuing Facebook's Face Recognition feature following a lengthy privacy battle. Meta says the change will roll out in the coming weeks.

Reference: https://www.theverge.com/2021/11/2/22759613/meta-facebook-face-recognition-automatic-tagging-feature-shutdown

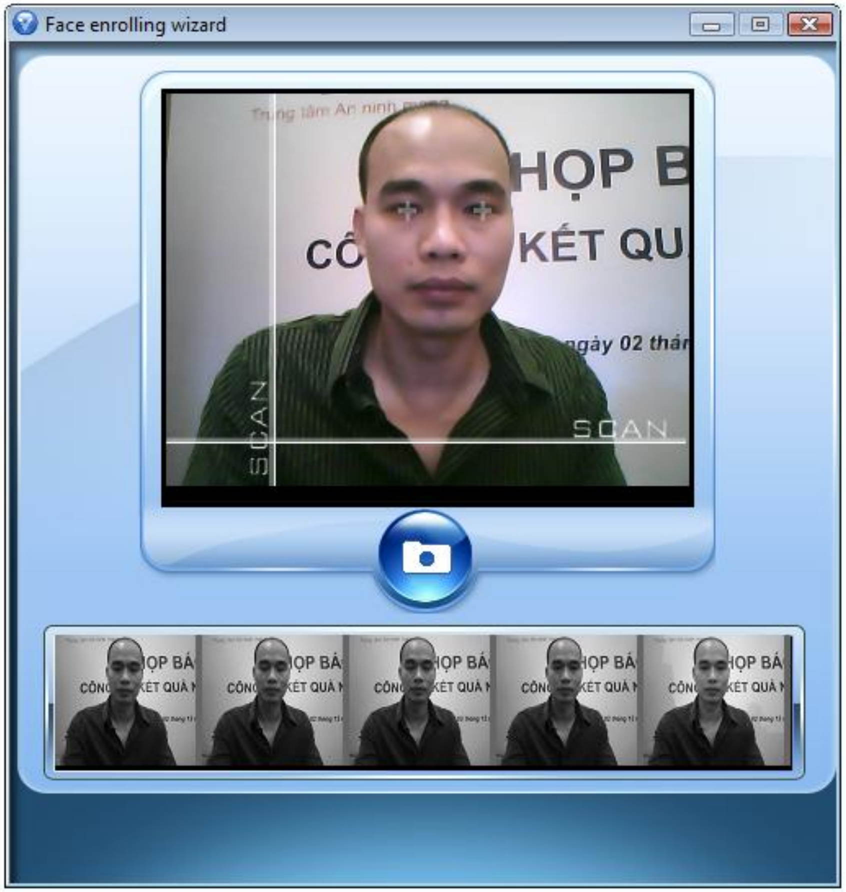
# iPhoto’s face detection

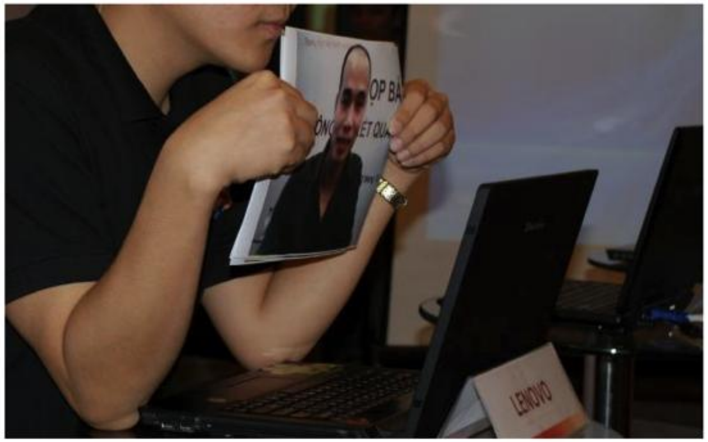


# iPhoto’s face detection



# Login without a password…





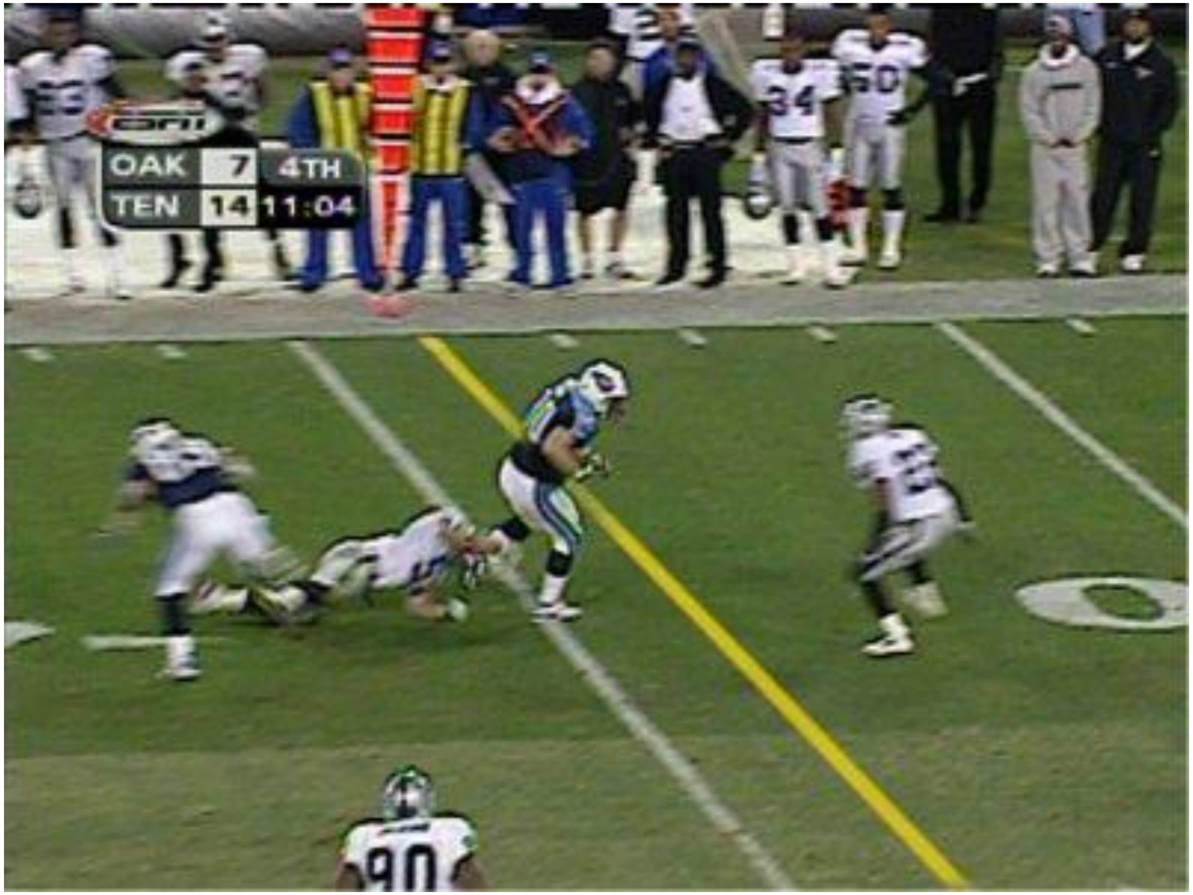
Dong Ngo / Cnet.com

# Vision-based interaction

Kinect

Source: S. Seitz

# Sports



*Sportvision* first down line

Nice explanation on www.howstuffworks.com

Source: S. Seitz

Why is computer vision so primitive?

# Why is computer vision so primitive?

- Vision is deceptively hard
- In 1966, Marvin Minsky at MIT asked an undergrad, Gerald Jay Sussman, to "spend the summer linking a camera to a computer and getting the computer to describe what it saw."

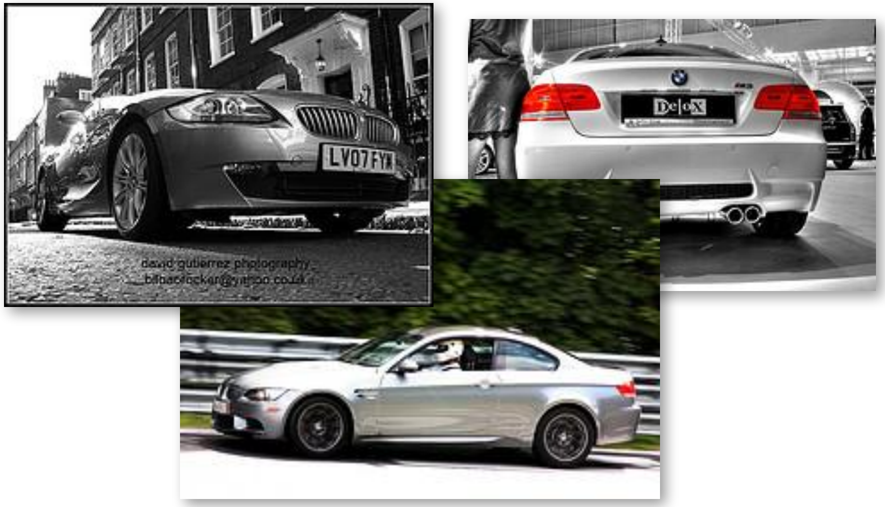
# Compare to NLP & speech recognition

- Speech recognition:
  - Well-defined atomic unit (phonemes, words)
  - Well-defined grammar
  - 1d sequence
  - Well-defined structure of documents (letters, words, sentences)

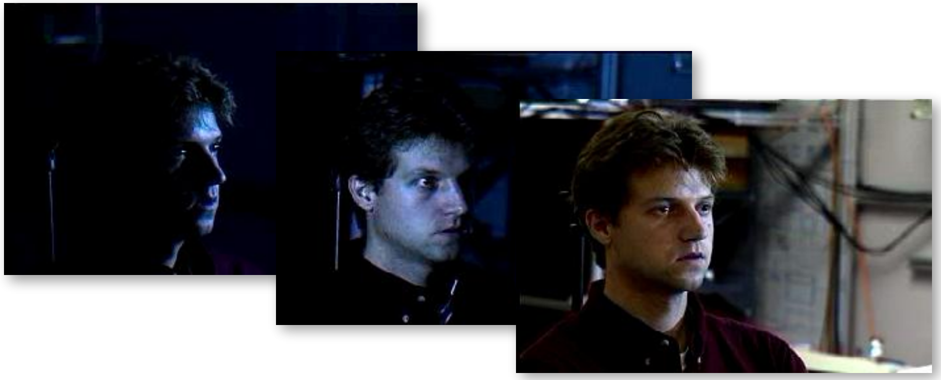
- Computer Vision:
  - Atomic unit: ?? (pixels? objects? “regions”?)
  - Grammar: ??
  - 2d image or 3D scene
  - Structure of images: ??

# Why is computer vision difficult?

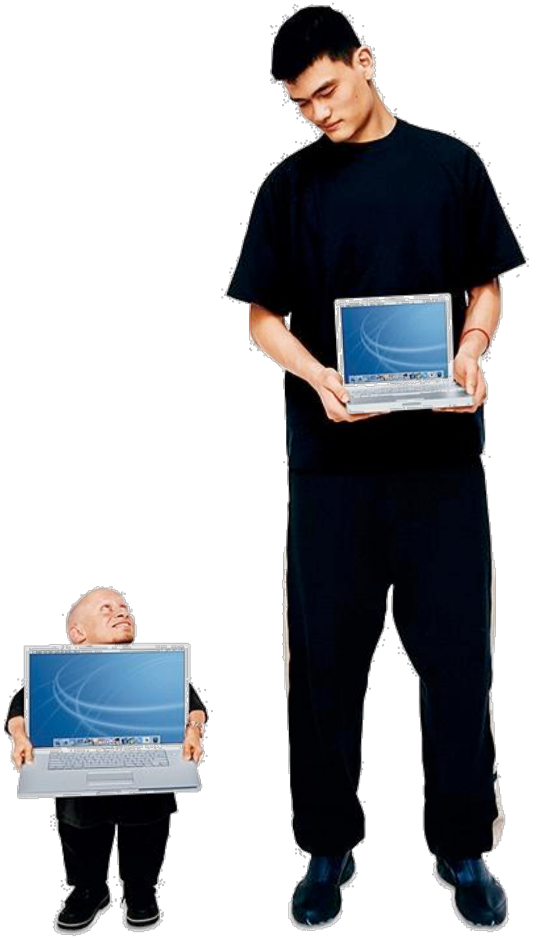
# Why is computer vision difficult?



Viewpoint variation



Illumination changes



Scale changes

# Why is computer vision difficult?

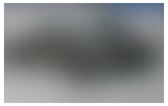
Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter

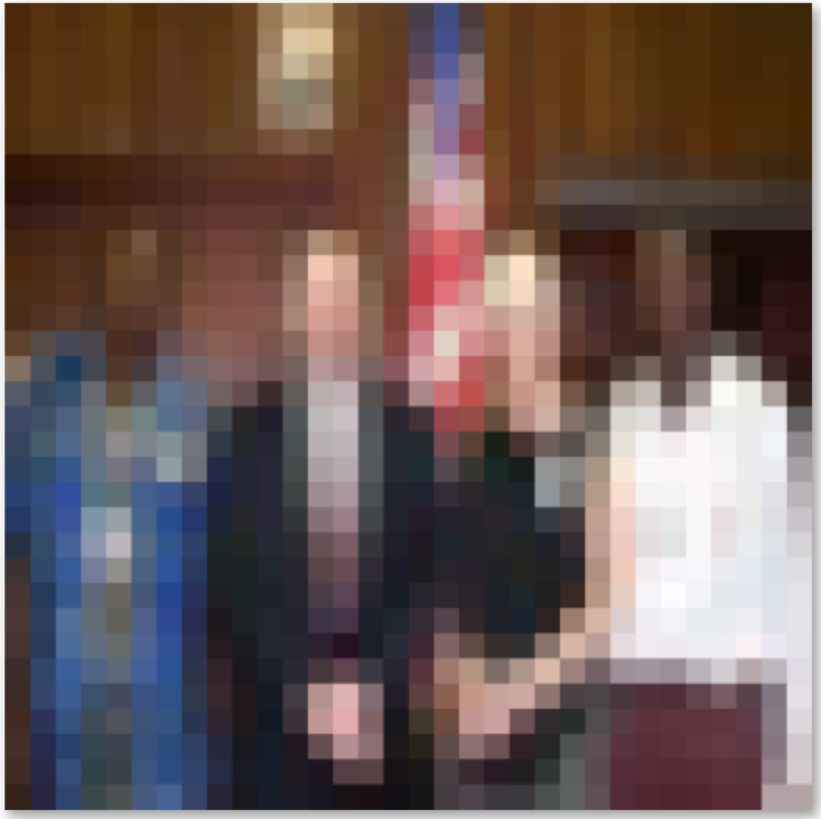
Occlusion

# Role of high-level reasoning



Fei-Fei, Fergus & Torralba

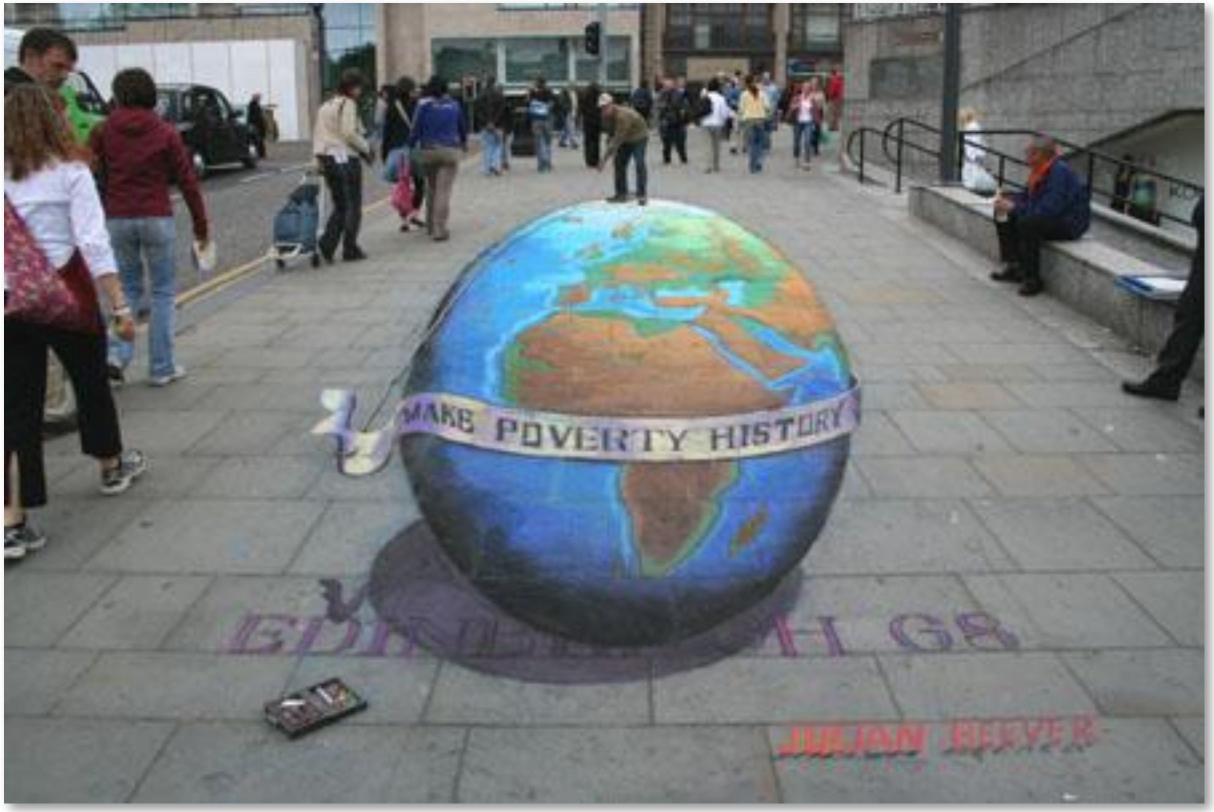
# Role of high-level reasoning



Source: “80 million tiny images” by Torralba, et al.

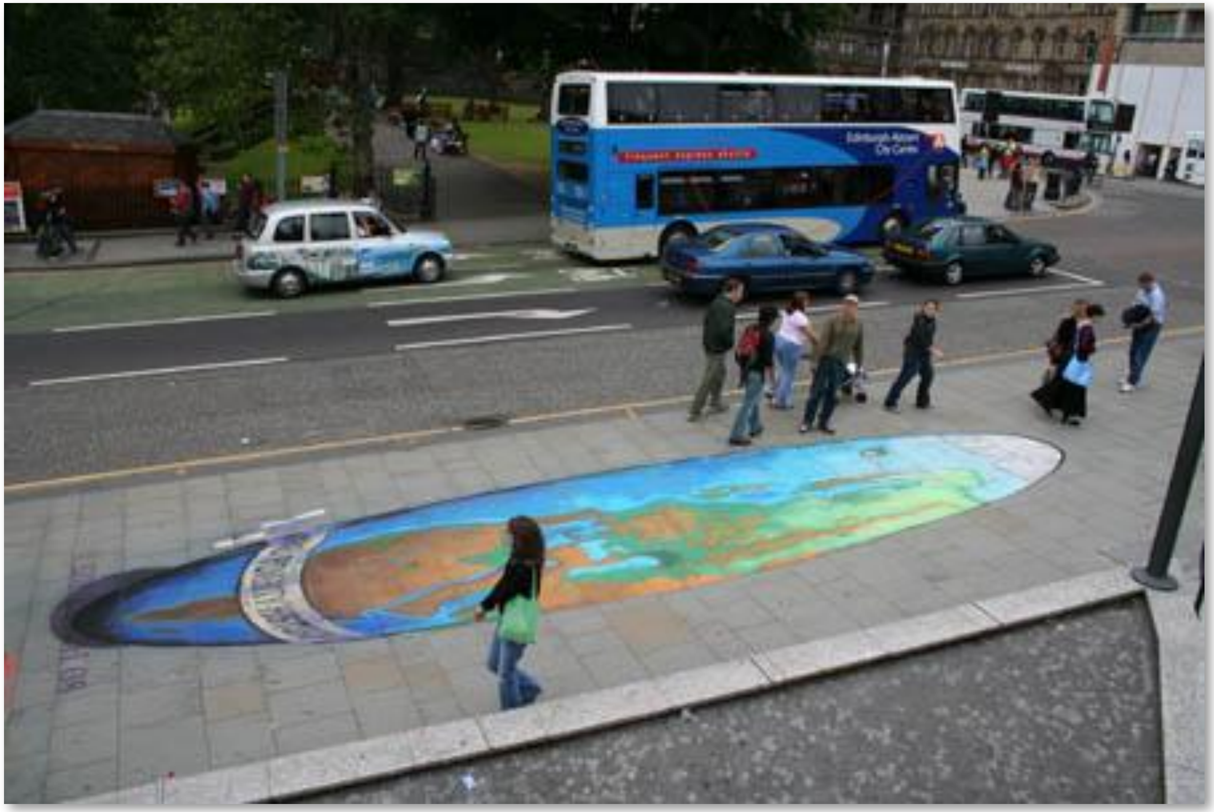
# Perception is inherently ambiguous

- Many scenes could have created a given 2D image
  - People figure out the "most likely" one based on experience, intuition, convention, … ?



Julian Beever

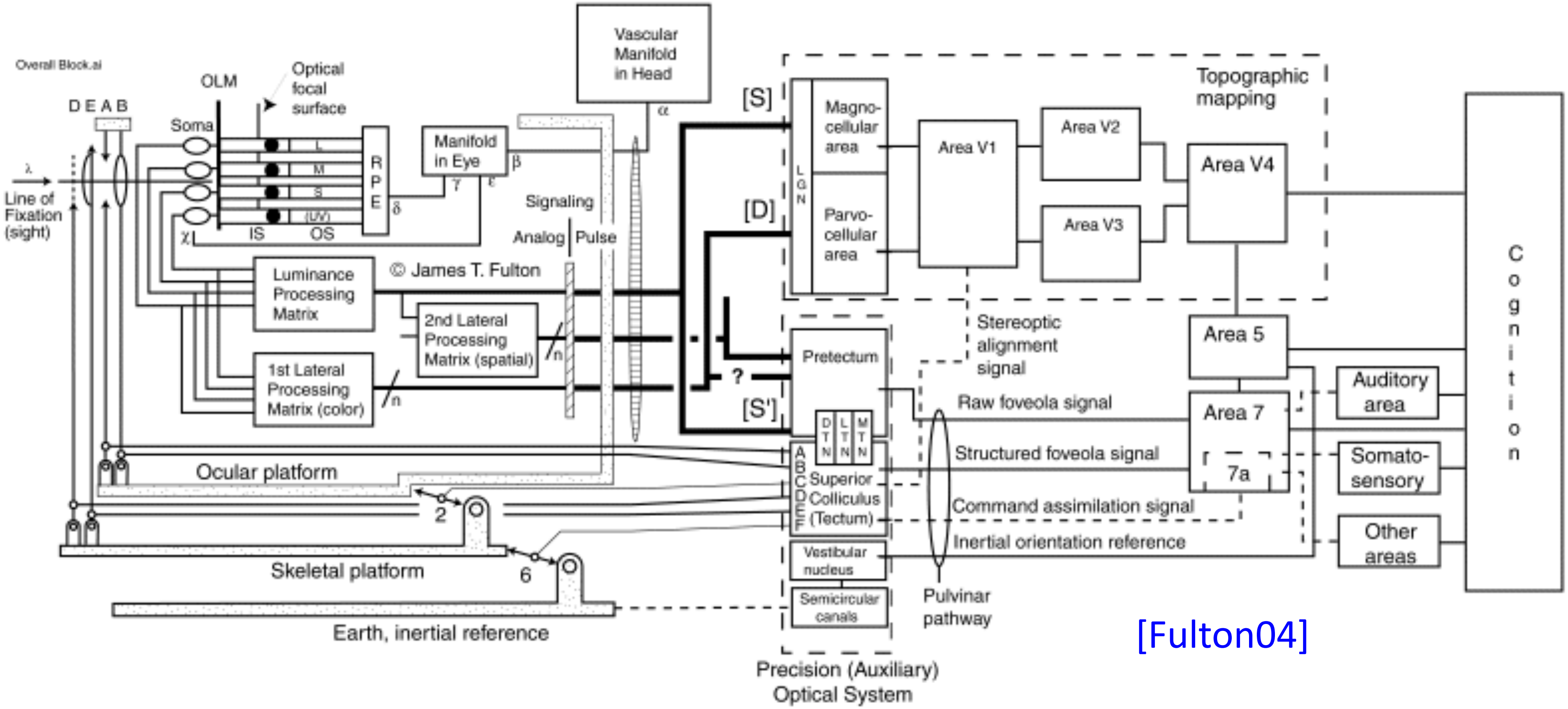
# Perception is inherently ambiguous



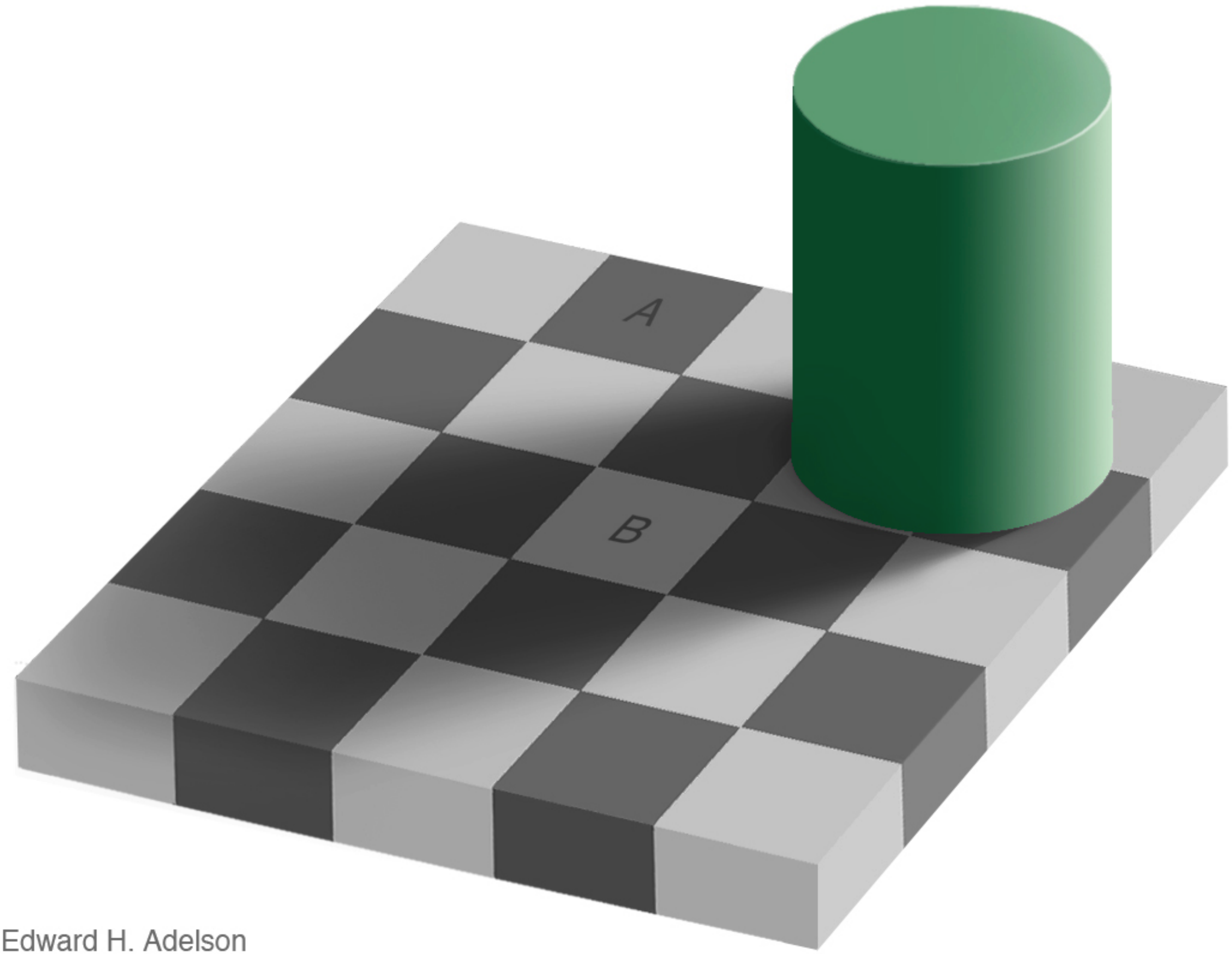
# How does human vision work?

# How do people (and animals) see?

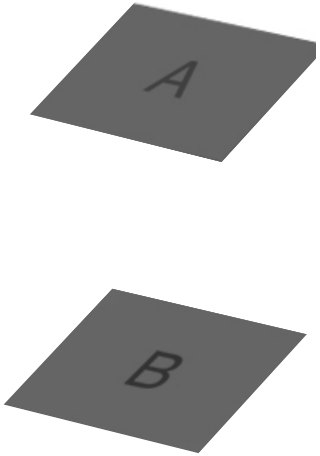
- We don’t really know.

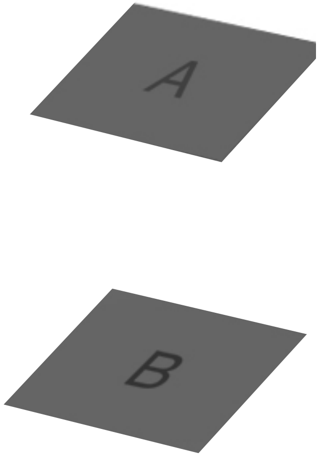


[Fulton04]

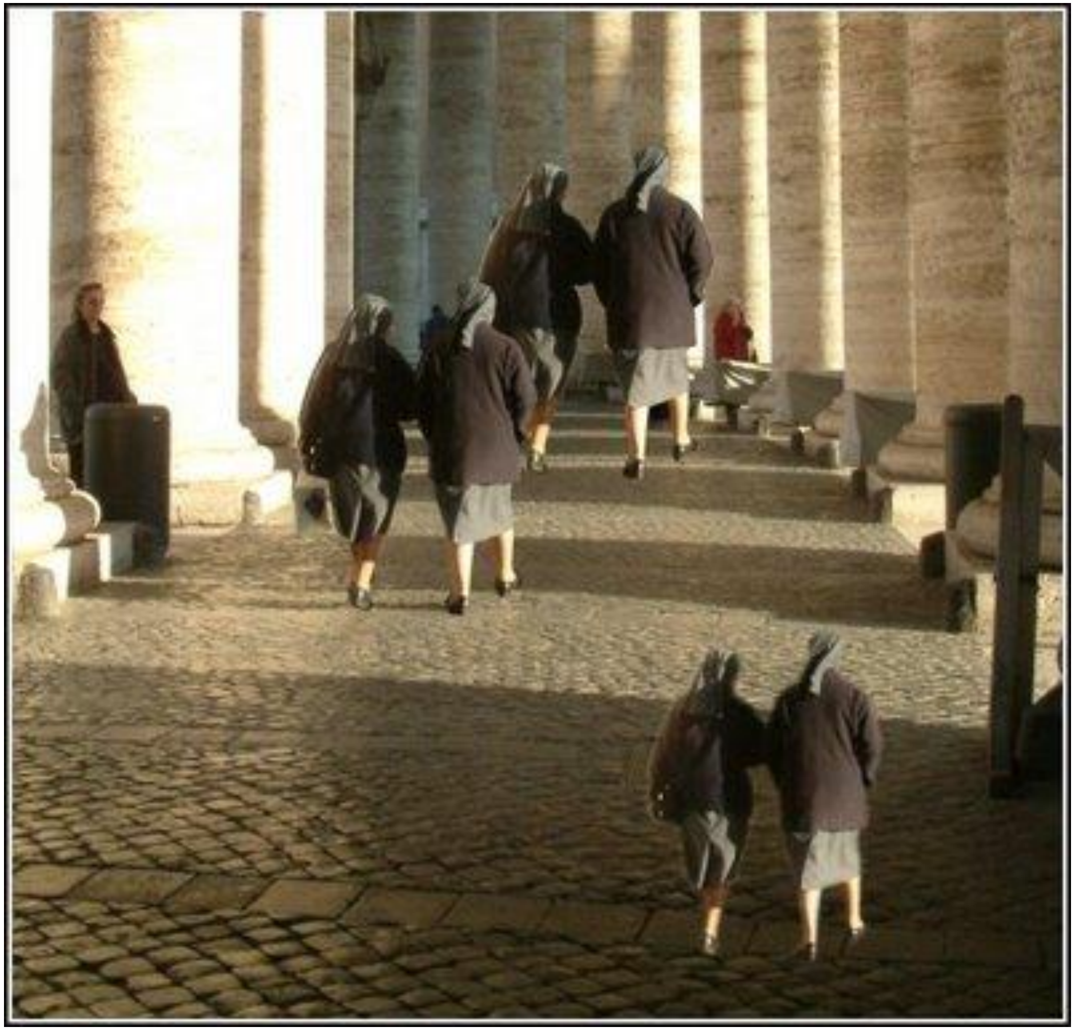


Edward H. Adelson


A
B


A
B

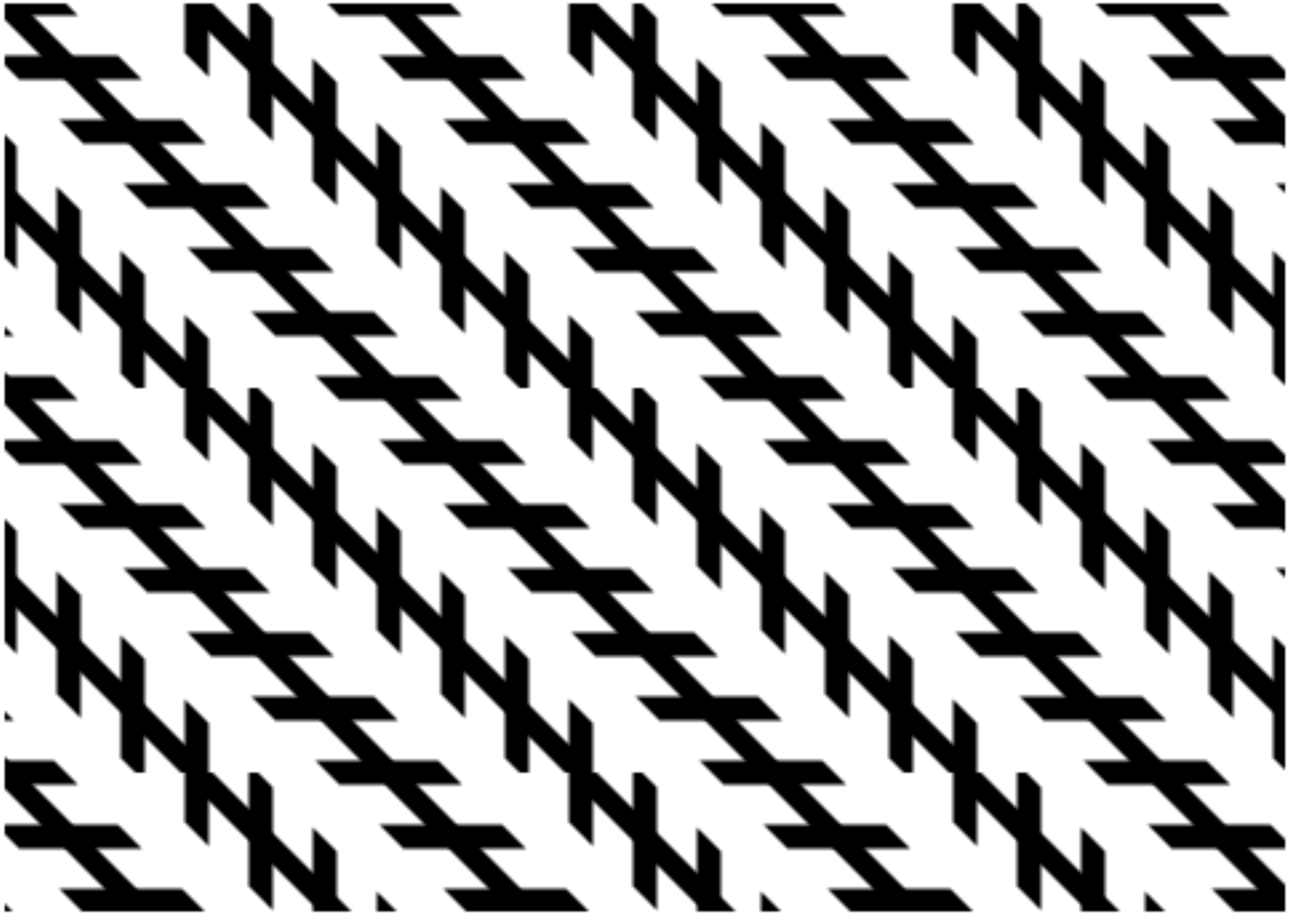
# How does human vision work?



# How does human vision work?

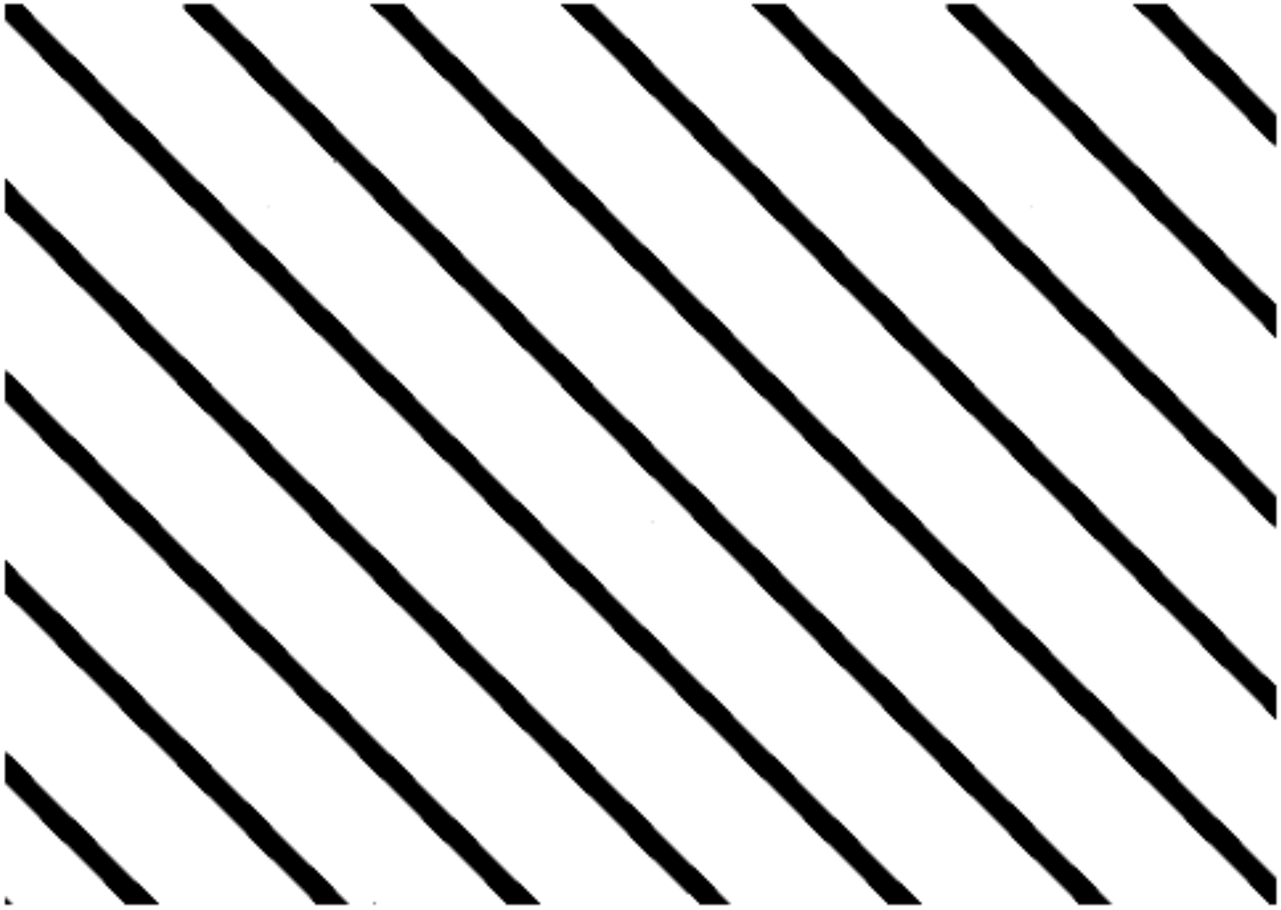
# Zollner illusion

- Are these lines parallel?



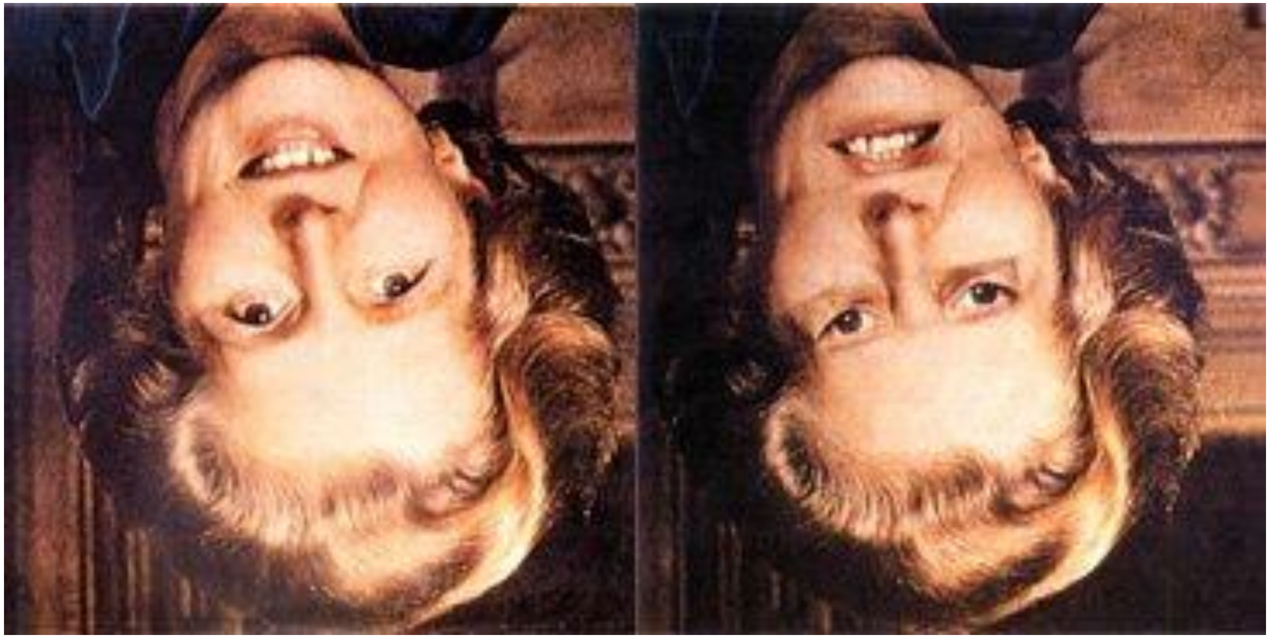
# Zollner illusion

- After removing the hatches on these lines, they look parallel

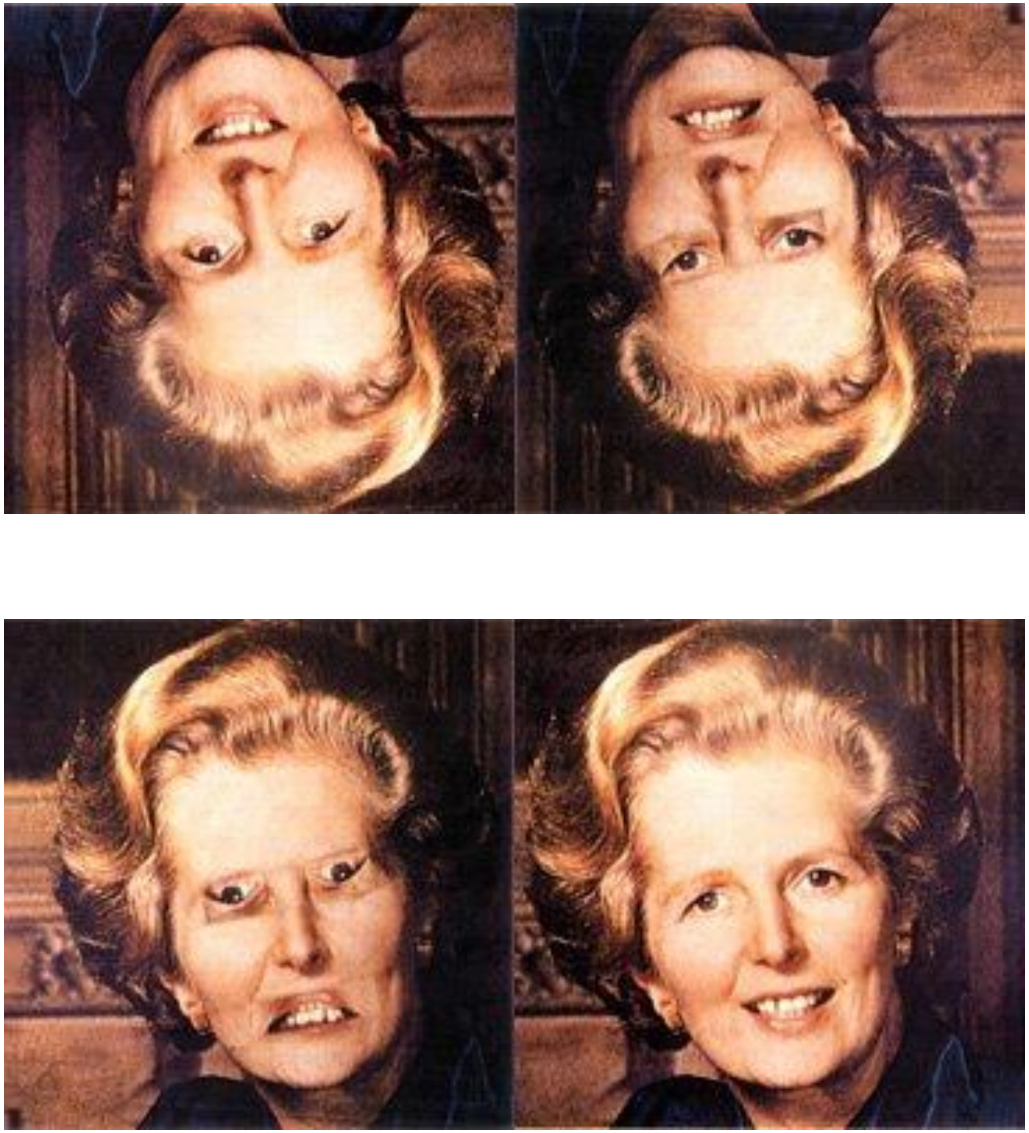


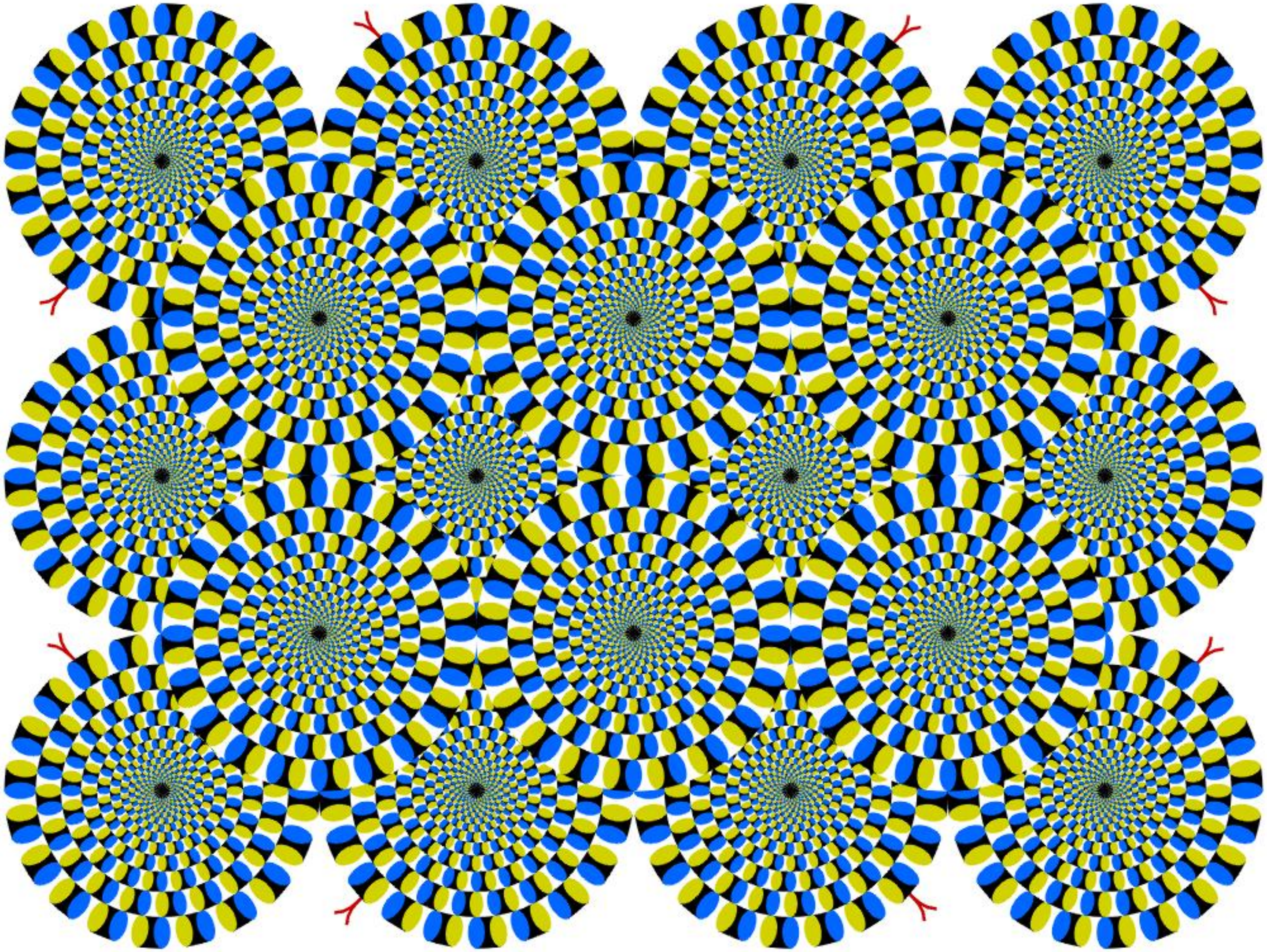
https://www.illusionsindex.org/ir/zoellner-illusion

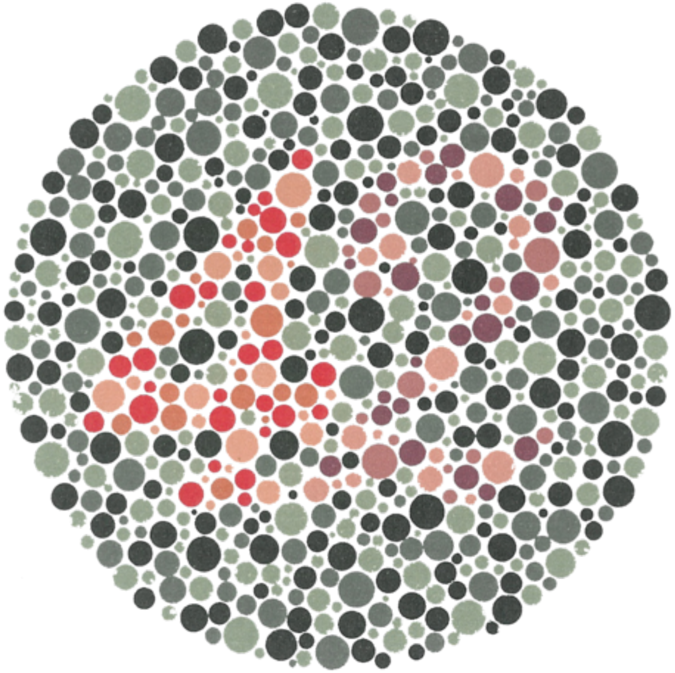
# The Thatcher effect

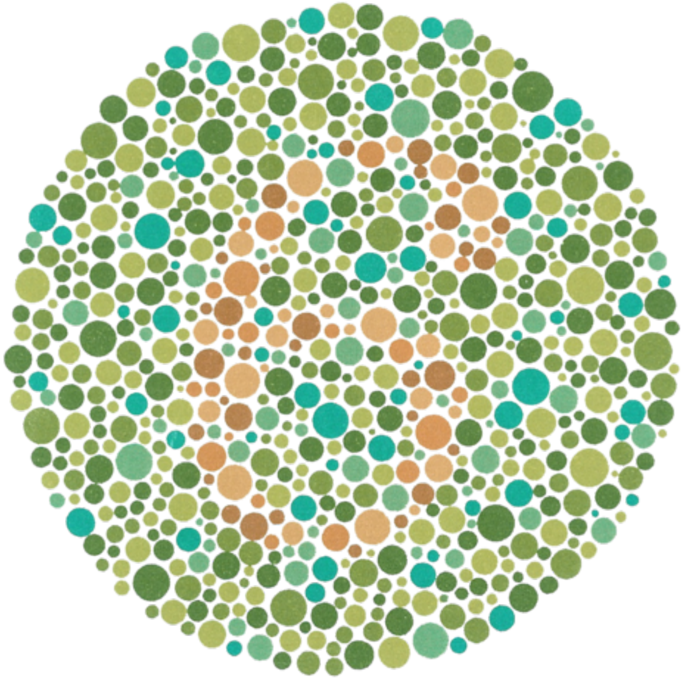


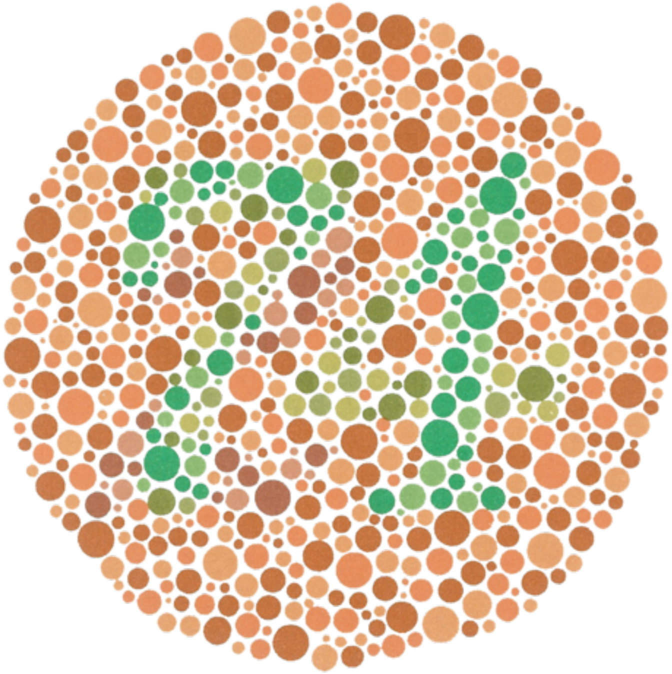
[Thompson 1980]

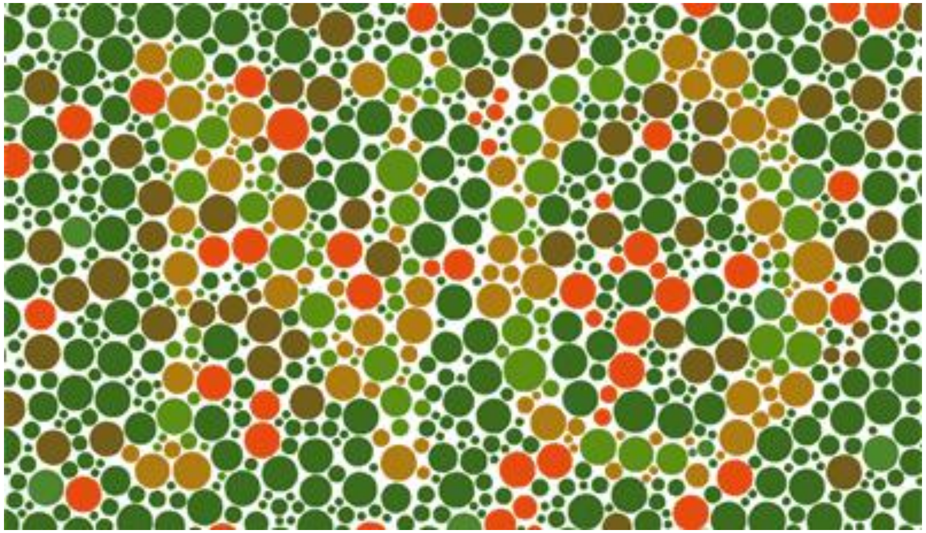












# Conclusion: why is computer vision so difficult?

Bad news:

- Computers lack higher-level prior knowledge
- Perception is inherently ambiguous
- We don't know how the human brain works
- Haven't found mathematical models that represent human vision well
- The models we do have require intensive (usually intractable) computation

Good news:

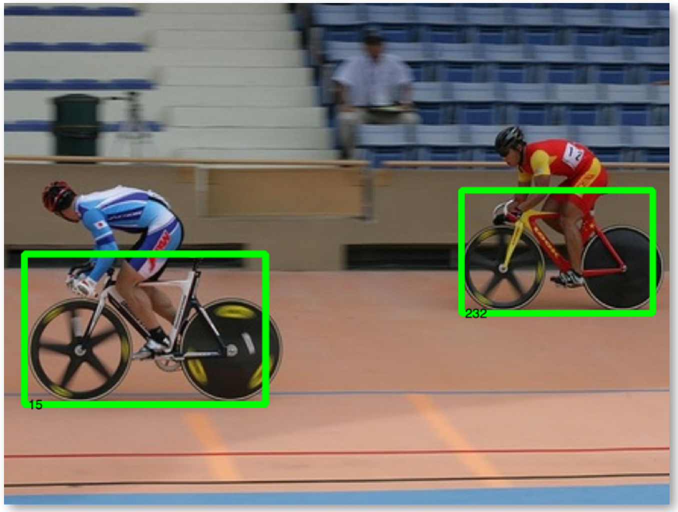
- So much progress is being made!  Especially in applications where perfect performance isn't needed.

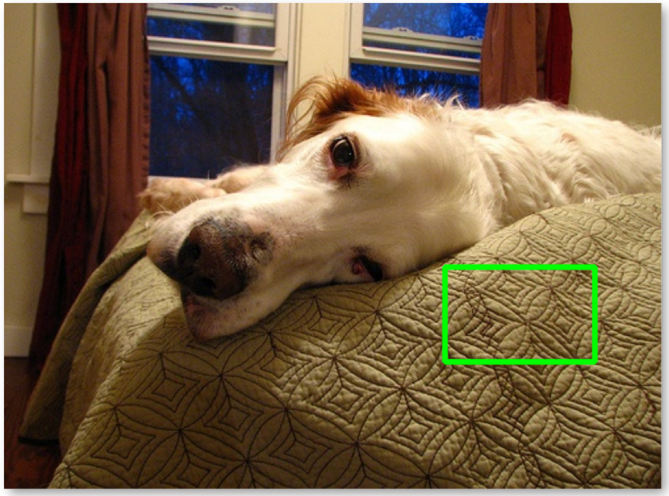
# Recent progress

# Computer vision

- We don’t understand the visual system well enough to model it, let alone replicate it
- For now, most successful computer vision systems are not inspired by biology
  - Instead use techniques and mathematical models that work well in practice, e.g. probabilistic models, machine learning, robust optimization, …
- A large amount of progress in the last ~10 years

# Object recognition





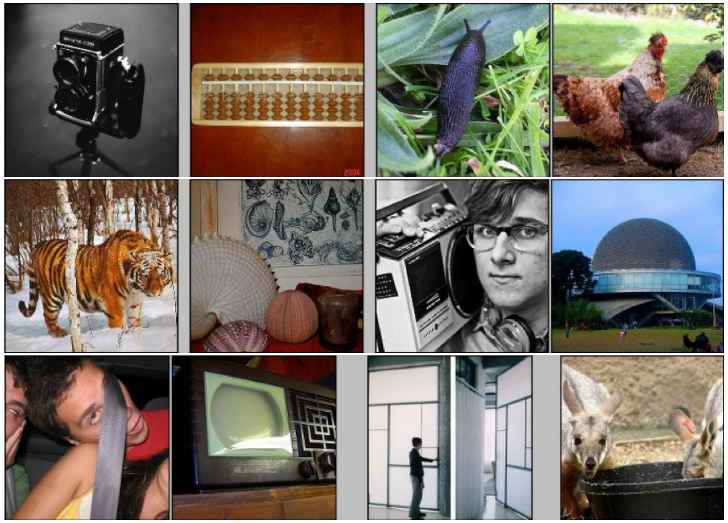
Facebook's Facial Recognition 'Approaching Human-Level Performance'

New computer program first to recognize sketches more accurately than a human

A Computer Can Recognize Emotions Better Than Most People

Microsoft, Google Beat Humans at Image Recognition

# ImageNet Challenge 2012

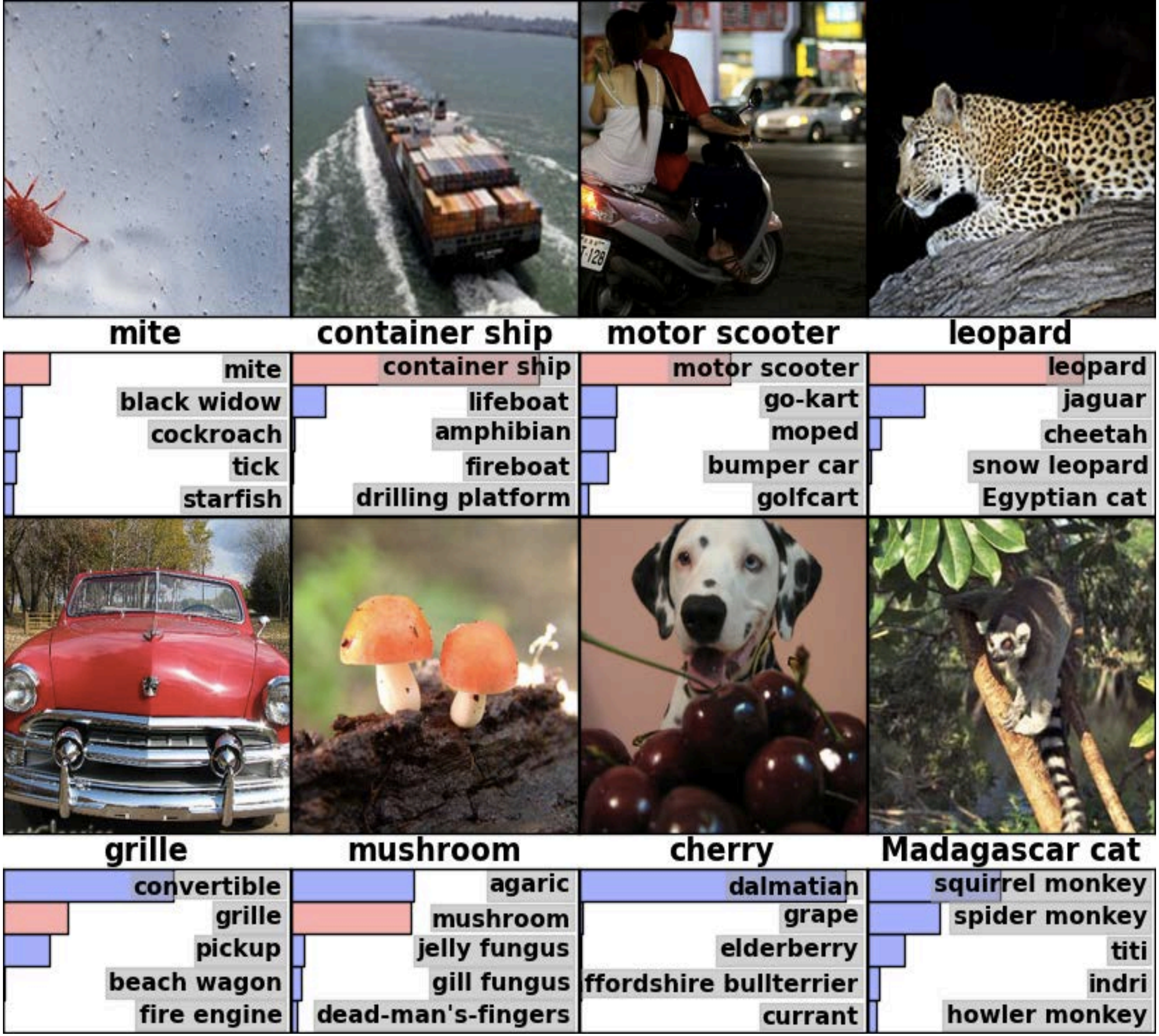


[Deng et al. CVPR 2009]

- ~14 million labeled images, 20k classes
- Images gathered from Internet
- Human labels via Amazon Turk
- Challenge: 1.2 million training images, 1000 classes

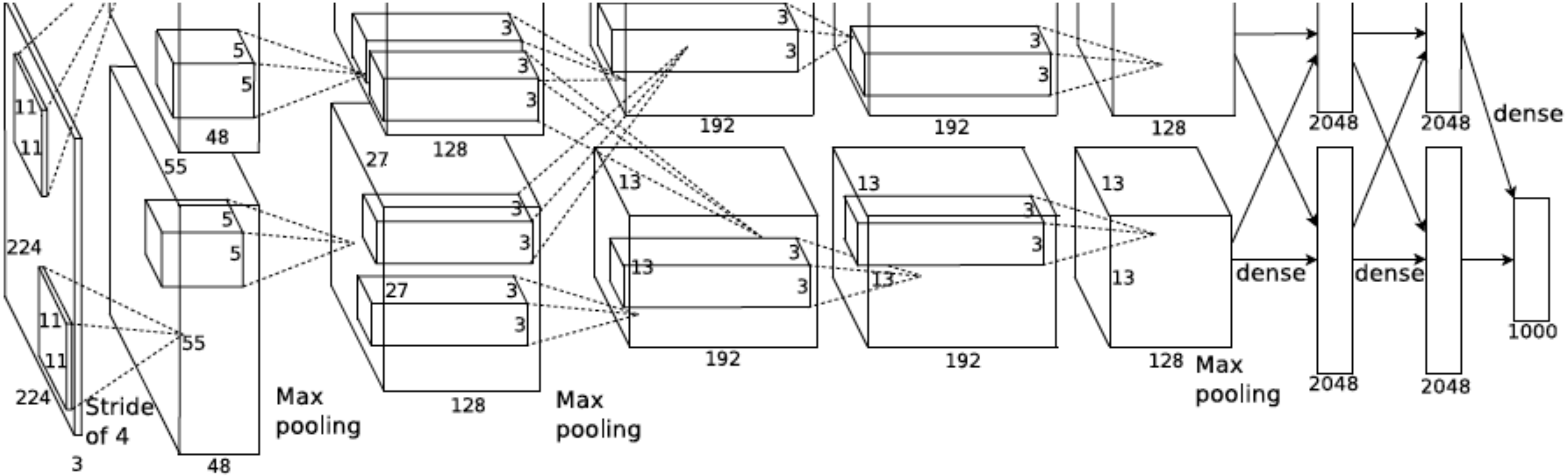
A. Krizhevsky, I. Sutskever, and G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

Slide credit: Rob Fergus


mite
mite
black widow
cockroach
tick
starfish
container ship
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
motor scooter
go-kart
moped
bumper car
golfcart
leopard
leopard
jaguar
cheetah
snow leopard
Egyptian cat
grille
convertible
grille
pickup
beach wagon
fire engine
mushroom
agaric
mushroom
jelly fungus
gill fungus
dead-man's-fingers
cherry
dalmatian
grape
elderberry
ffordshire bullterrier
currant
Madagascar cat
squirrel monkey
spider monkey
titi
indri
howler monkey

# ImageNet Challenge 2012

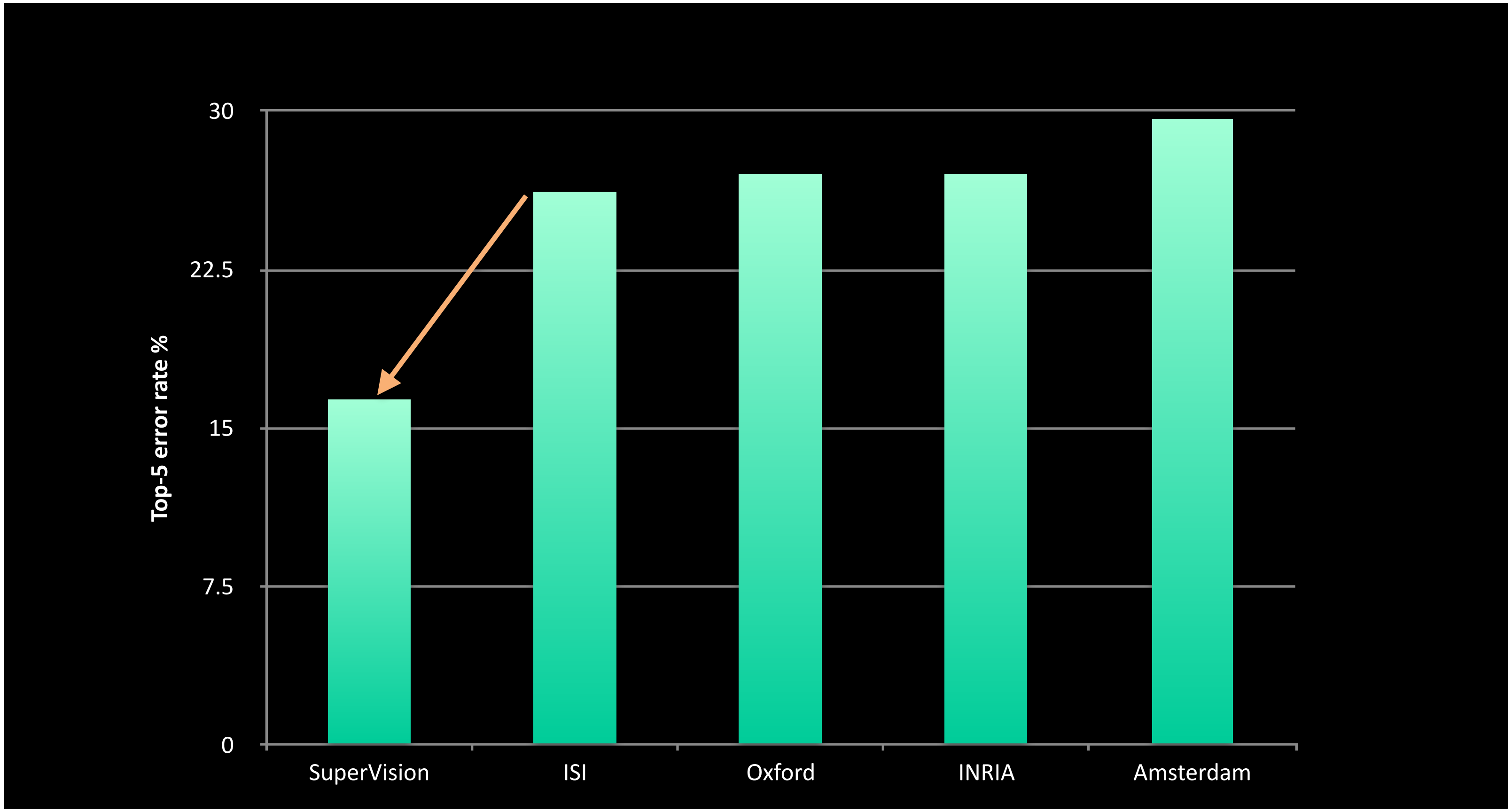
- Similar framework to LeCun'98 but:
  - Bigger model (7 hidden layers, 650,000 units, 60,000,000 params)
  - More data ($10^6$ vs. $10^3$ images)
  - GPU implementation (50x speedup over CPU)
    - Trained on two GPUs for a week
  - Better regularization for training (DropOut)



A. Krizhevsky, I. Sutskever, and G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

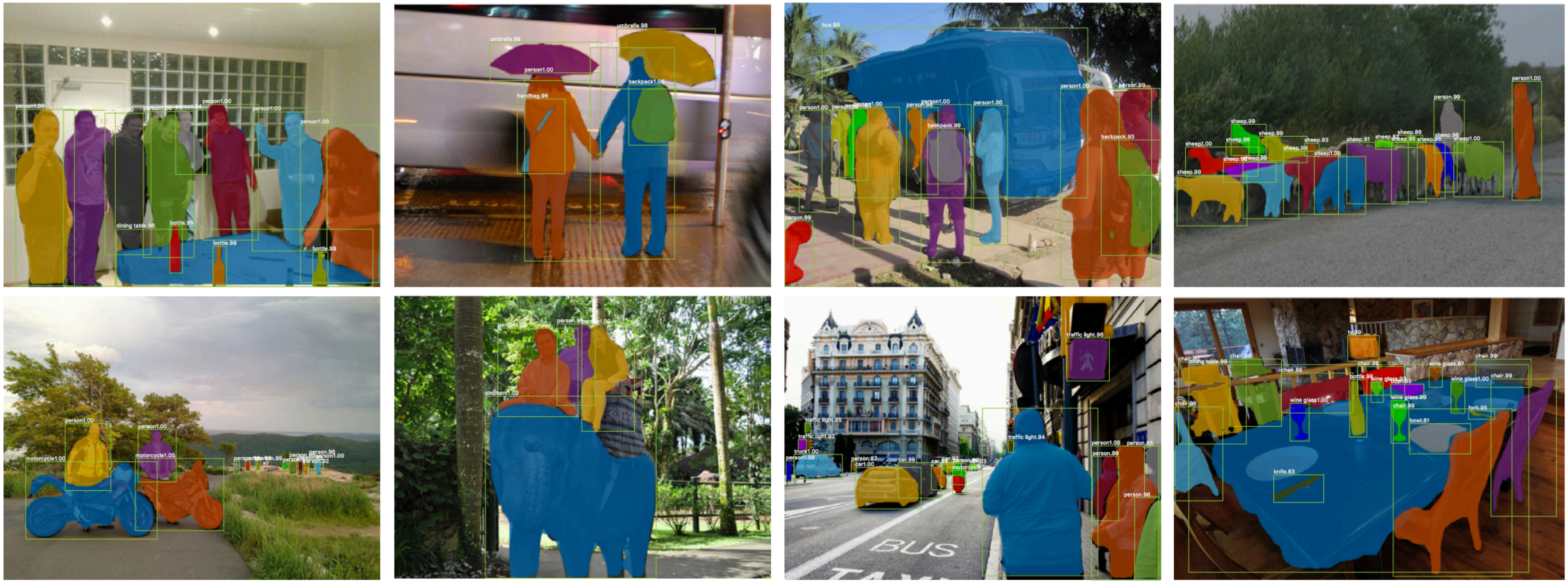
# ImageNet Challenge 2012

- A huge drop in error-rate with deep neural network-based model

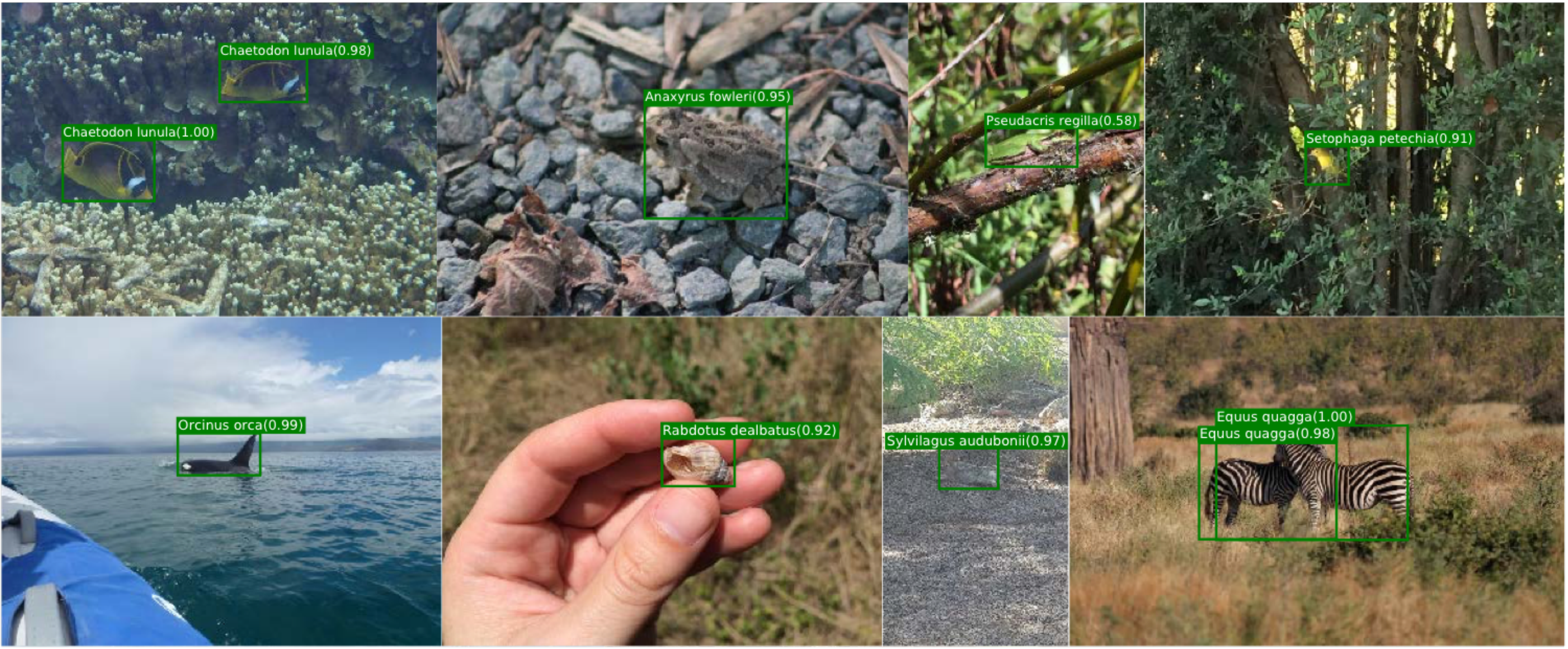


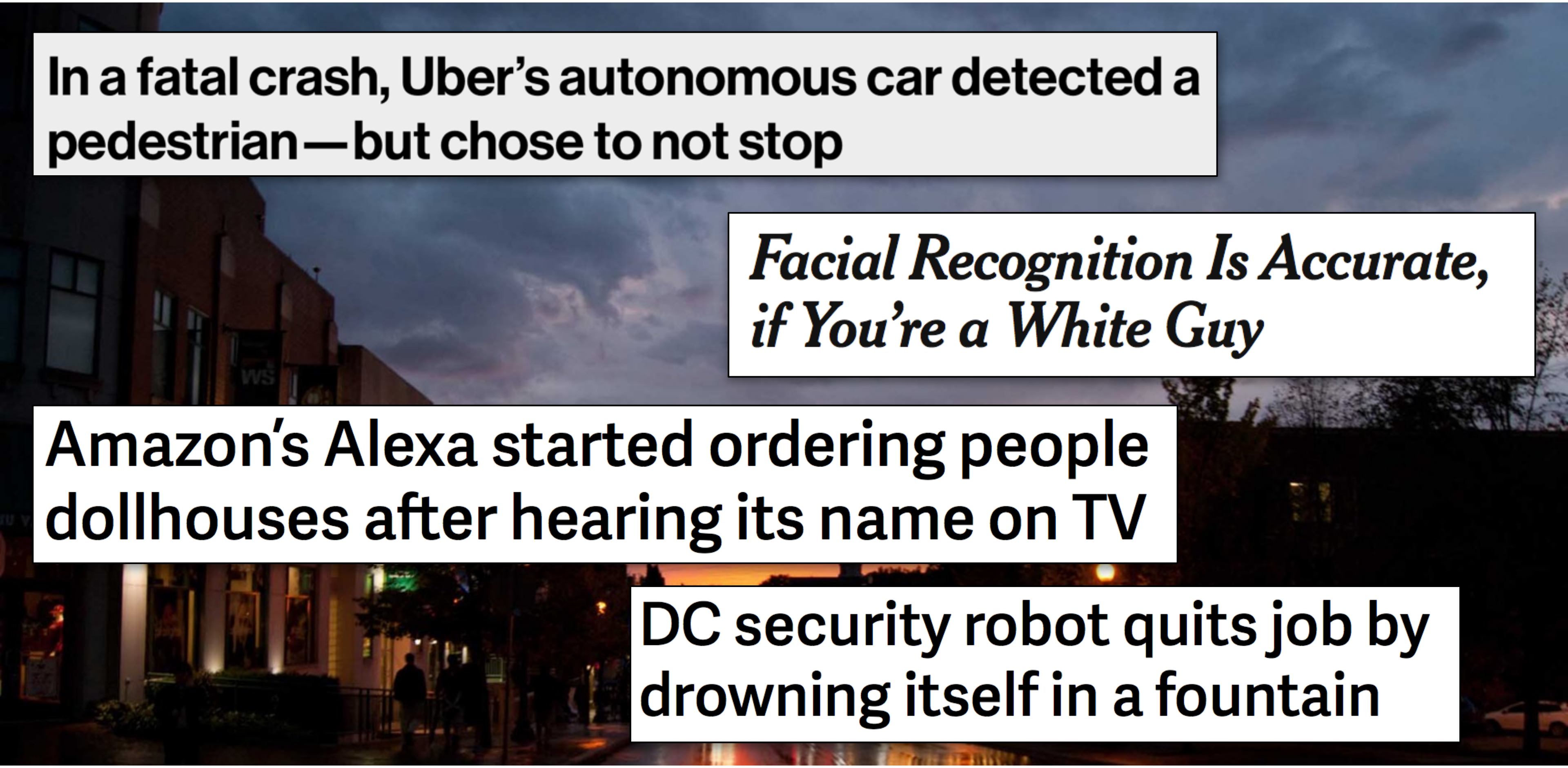
Slide credit: Rob Fergus

# Instance segmentation



He, Gkioxari, Dolar, Girschick, “Mask R-CNN,” CVPR 2017.


Chaetodon lunula(0.98)
Chaetodon lunula(1.00)
Anaxyrus fowleri(0.95)
Pseudacris regilla(0.58)
Setophaga petechia(0.91)
Orcinus orca(0.99)
Rabdotus dealbatus(0.92)
Sylvilagus audubonii(0.97)
Equus quagga(1.00)
Equus quagga(0.98)



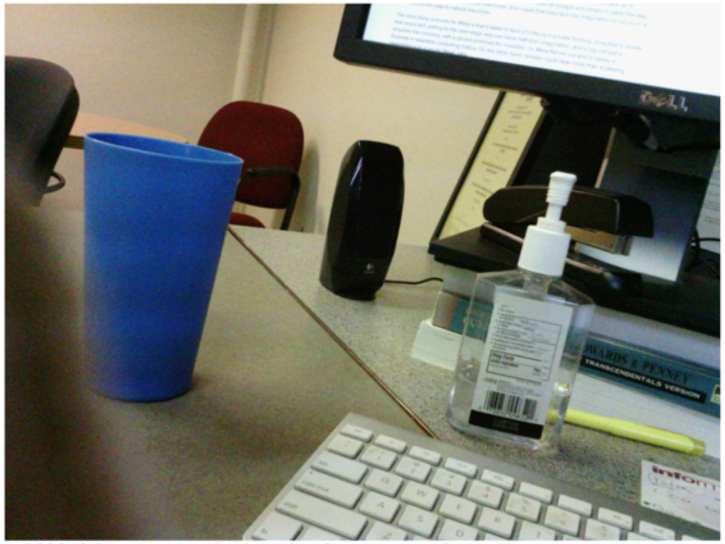
In a fatal crash, Uber's autonomous car detected a pedestrian—but chose to not stop

*Facial Recognition Is Accurate, if You're a White Guy*

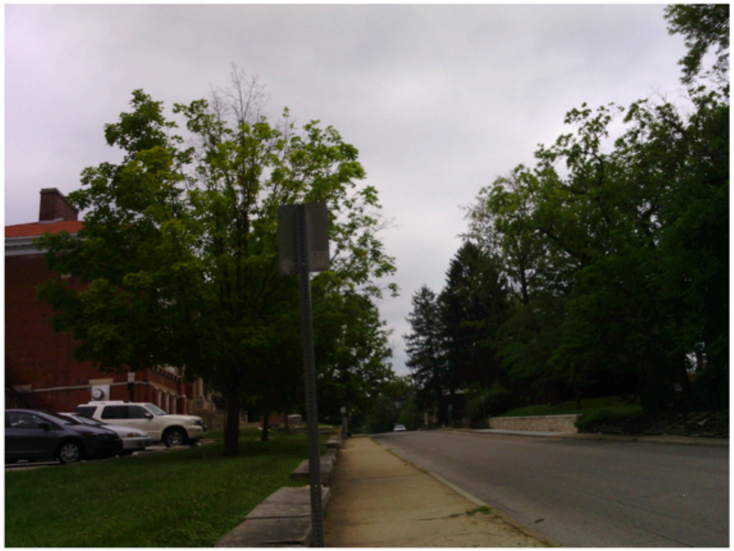
Amazon's Alexa started ordering people dollhouses after hearing its name on TV

DC security robot quits job by drowning itself in a fountain

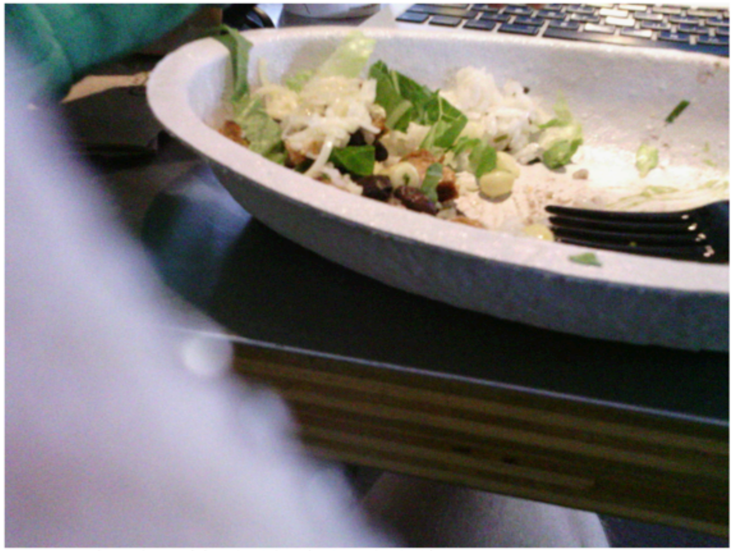
# Automatic image captioning by deep nets (success)



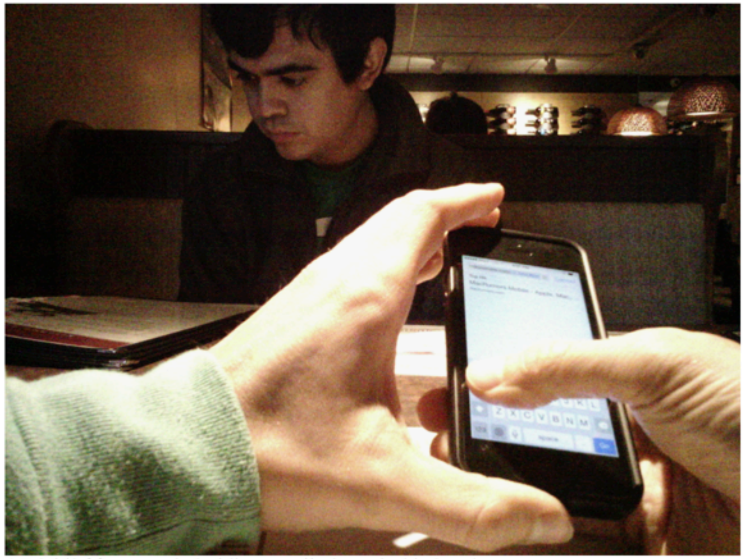
(-5.908705) a computer keyboard and mouse on a desk



(-11.431205) a street sign on a pole near a street

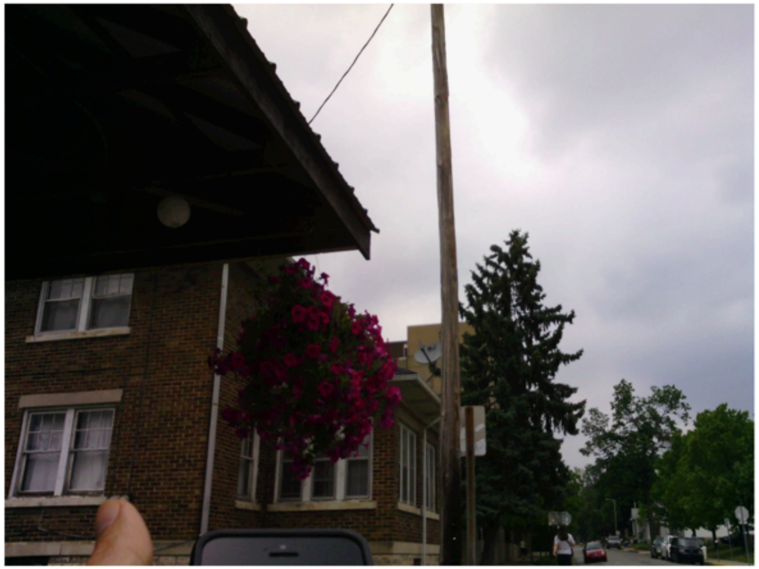


(-8.920025) a plate of food with a fork and a knife

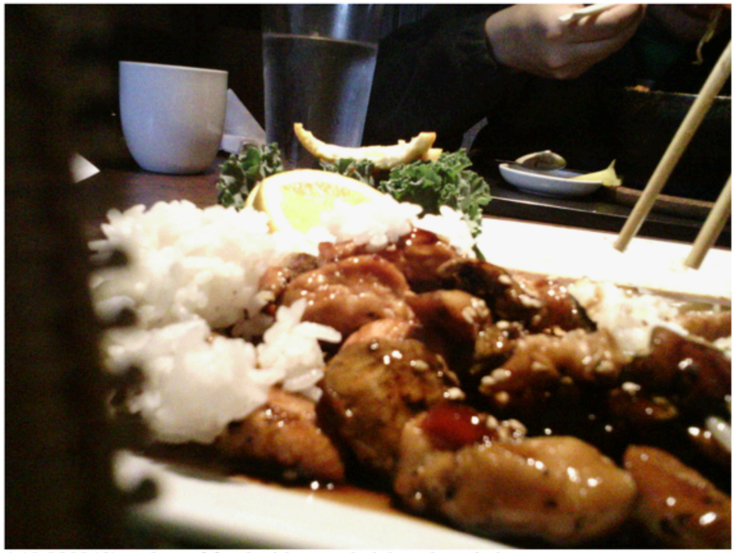


(-7.955366) a man is holding a cell phone in his hand

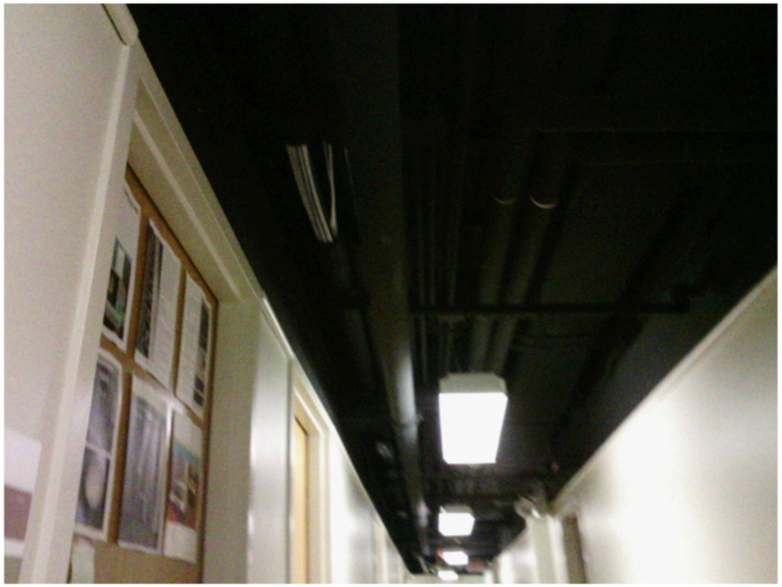
# Automatic image captioning by deep nets (failure)



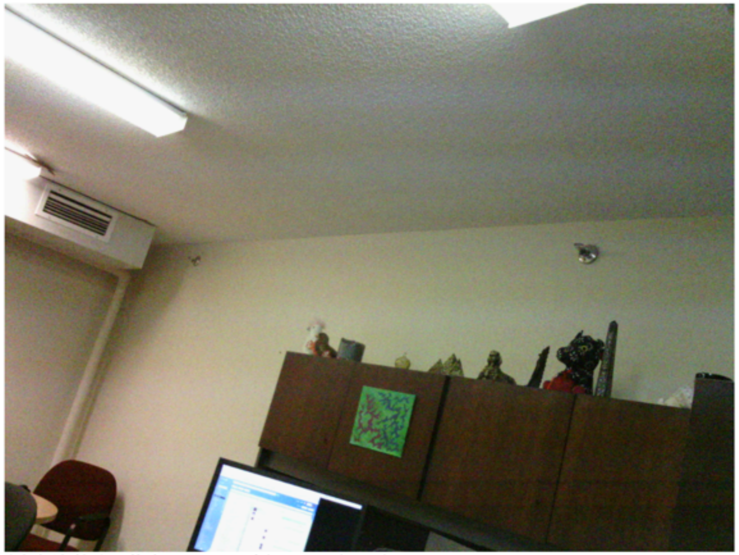
(-8.764608) a clock tower in the middle of a city



(-10.298248) a plate of food with a sandwich and a salad

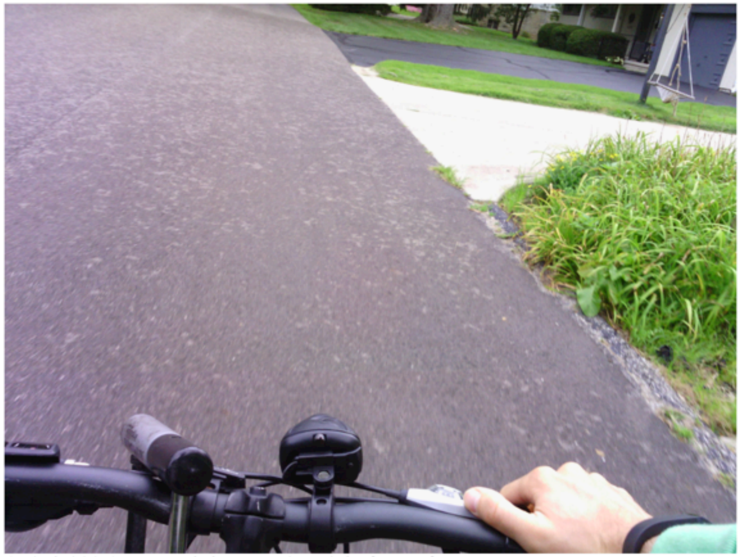


(-8.783713) a clock on a wall in a room

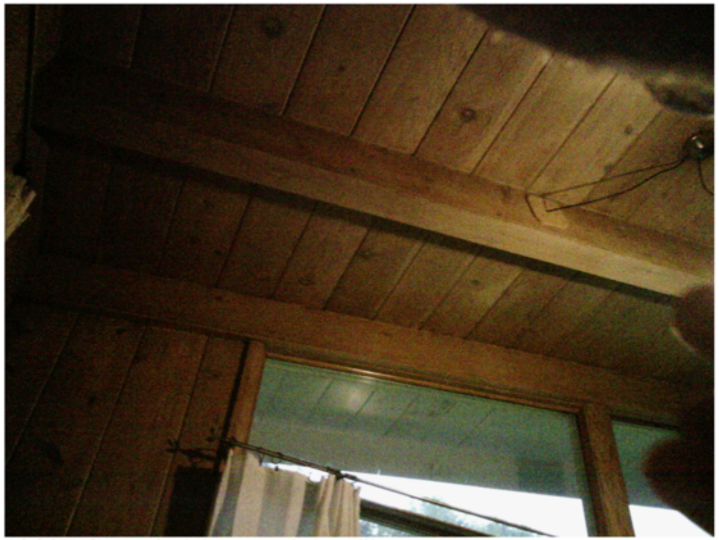


(-8.048537) a living room with a couch and a tv

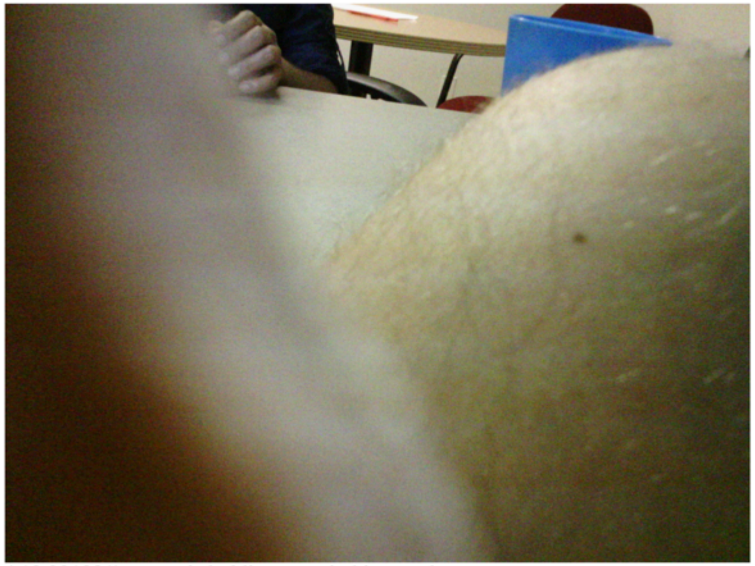
# Automatic image captioning by deep nets (failure)



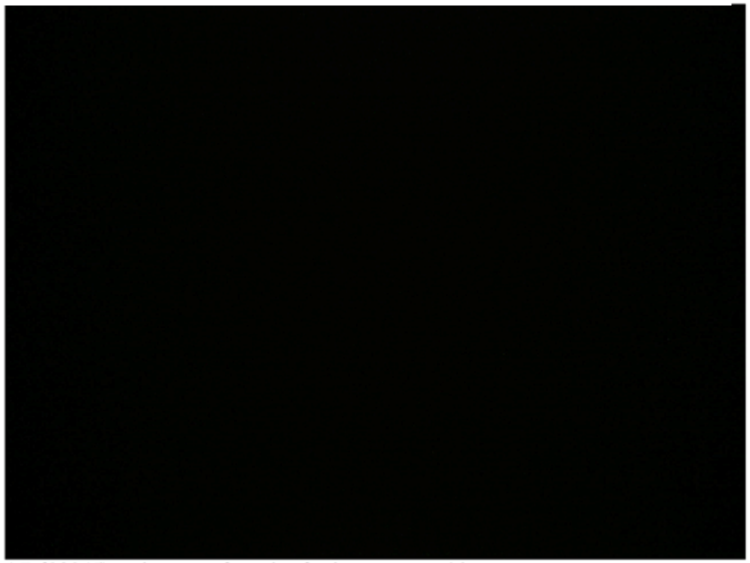
(-10.510004) a man sitting on a bench in front of a tree



(-8.265713) a cat is sitting on a window sill

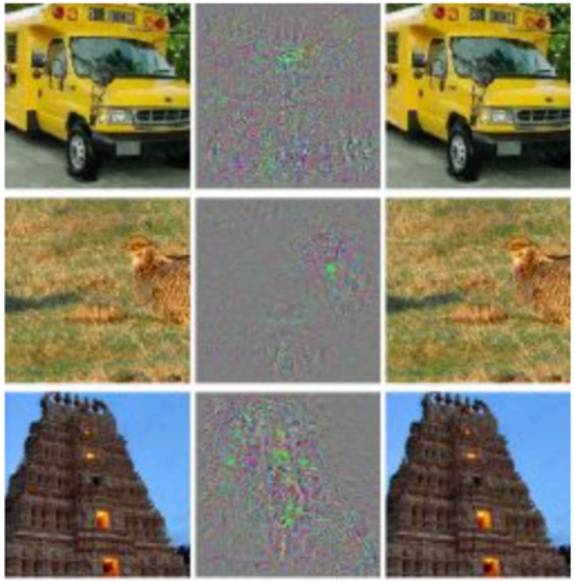


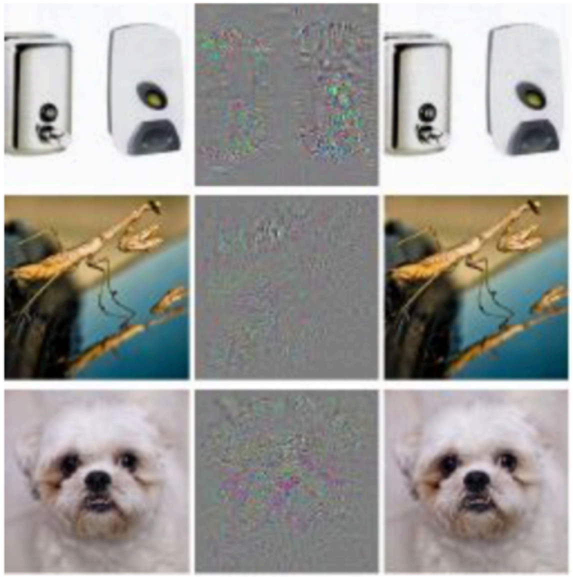
(-12.001291) a man is holding a cat in his mouth

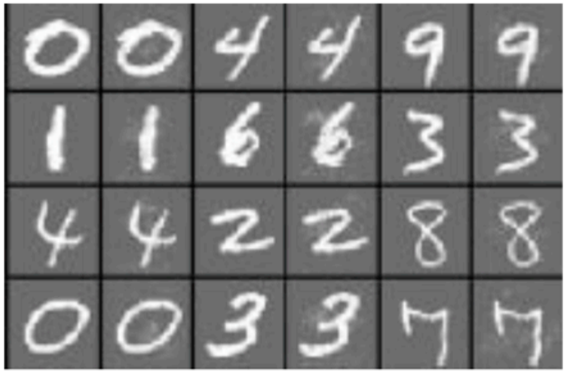


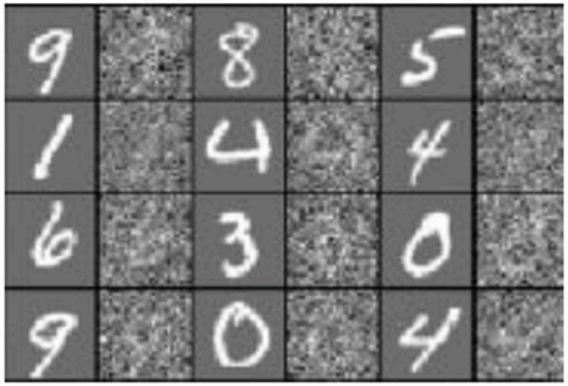
(-7.629245) a close up of a pair of scissors on a table

# Adversarial learning – Szegedy 2013



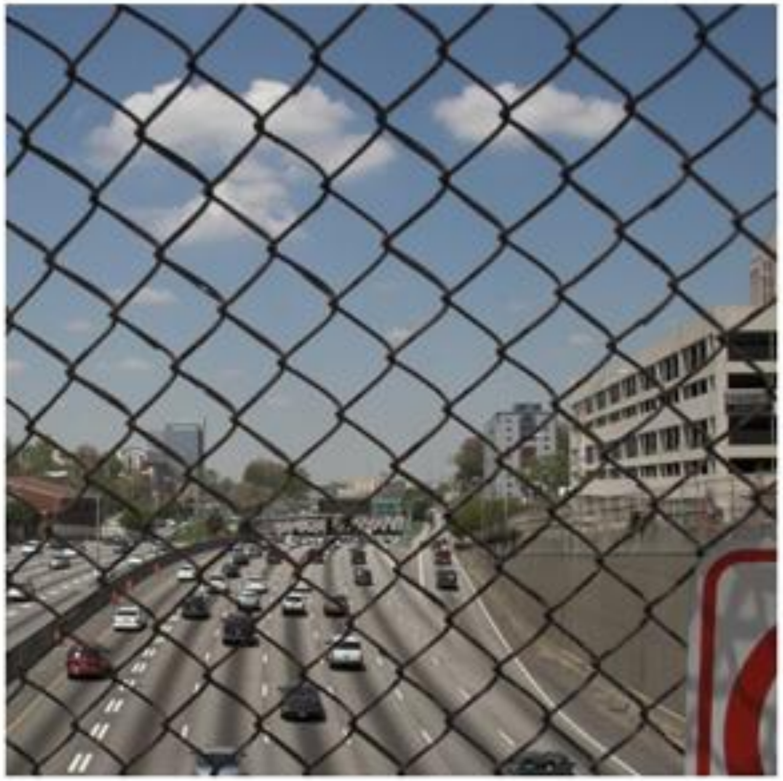


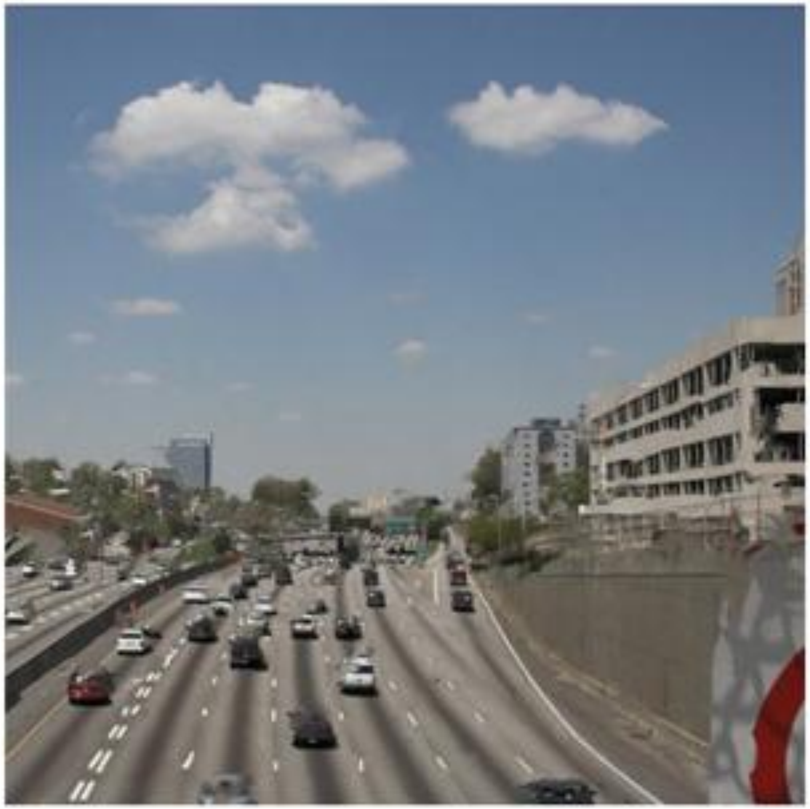




# Image restoration

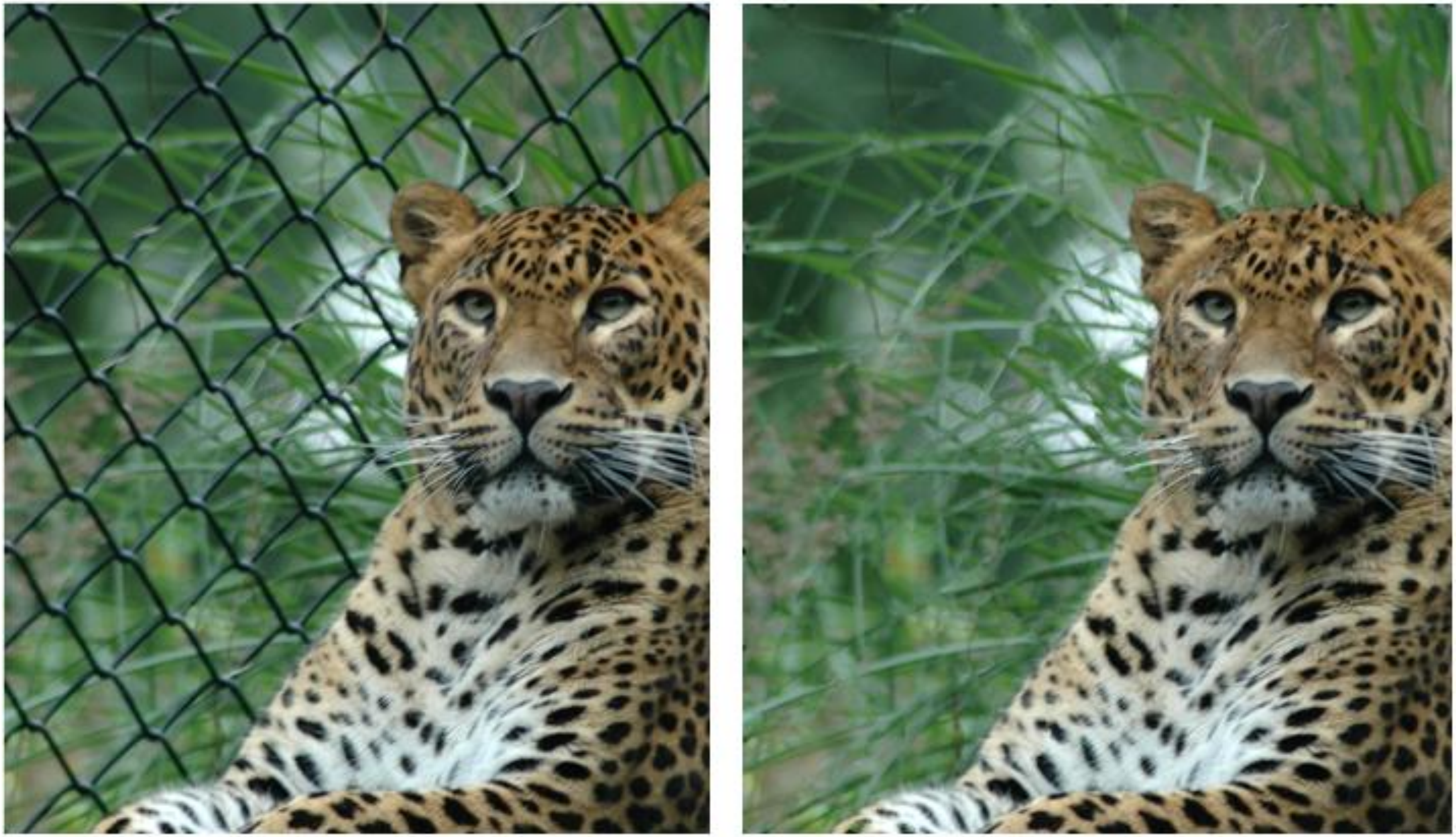
- Image ‘de-fencing’ [Liu08]





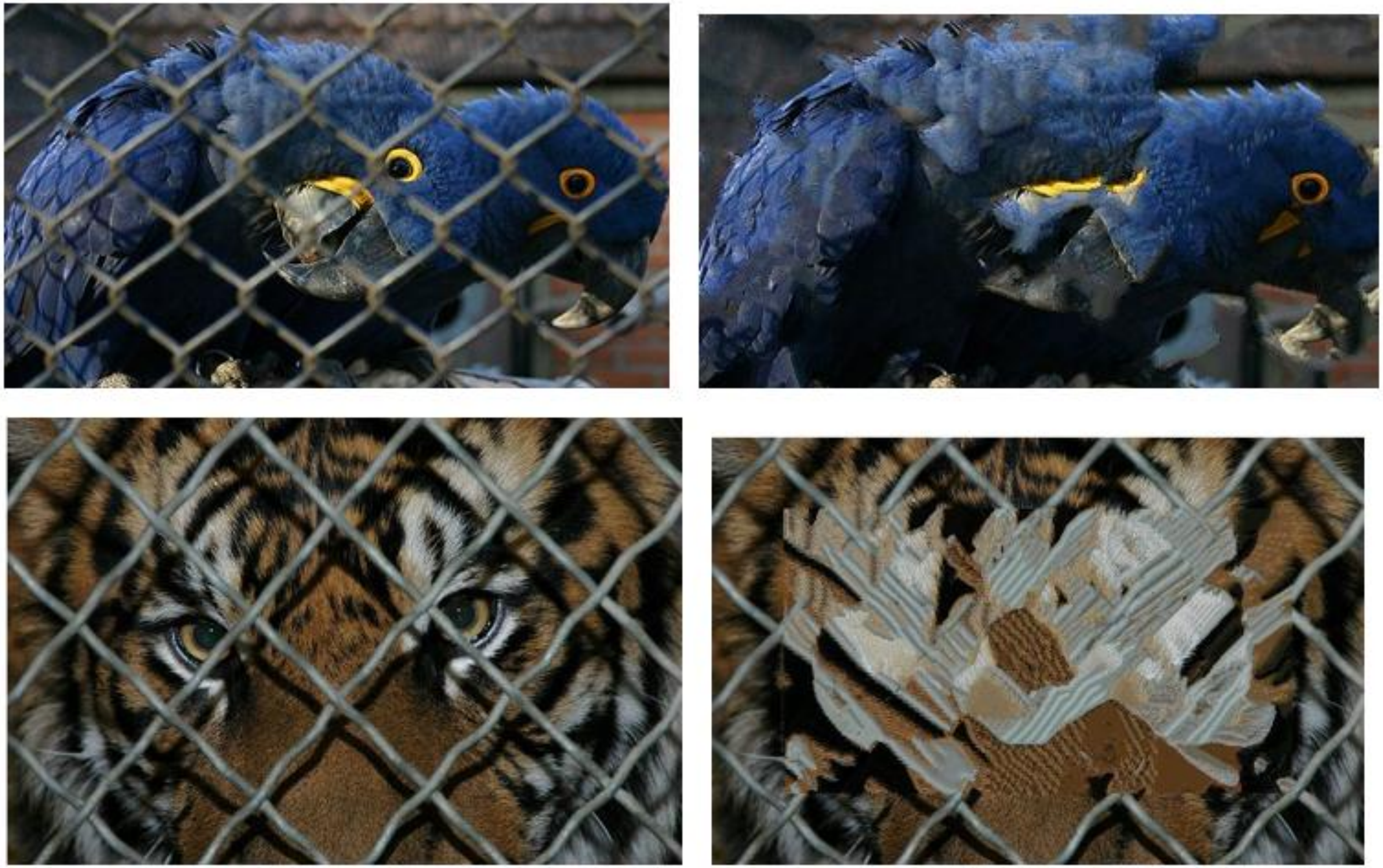
[Liu08]

# Image restoration



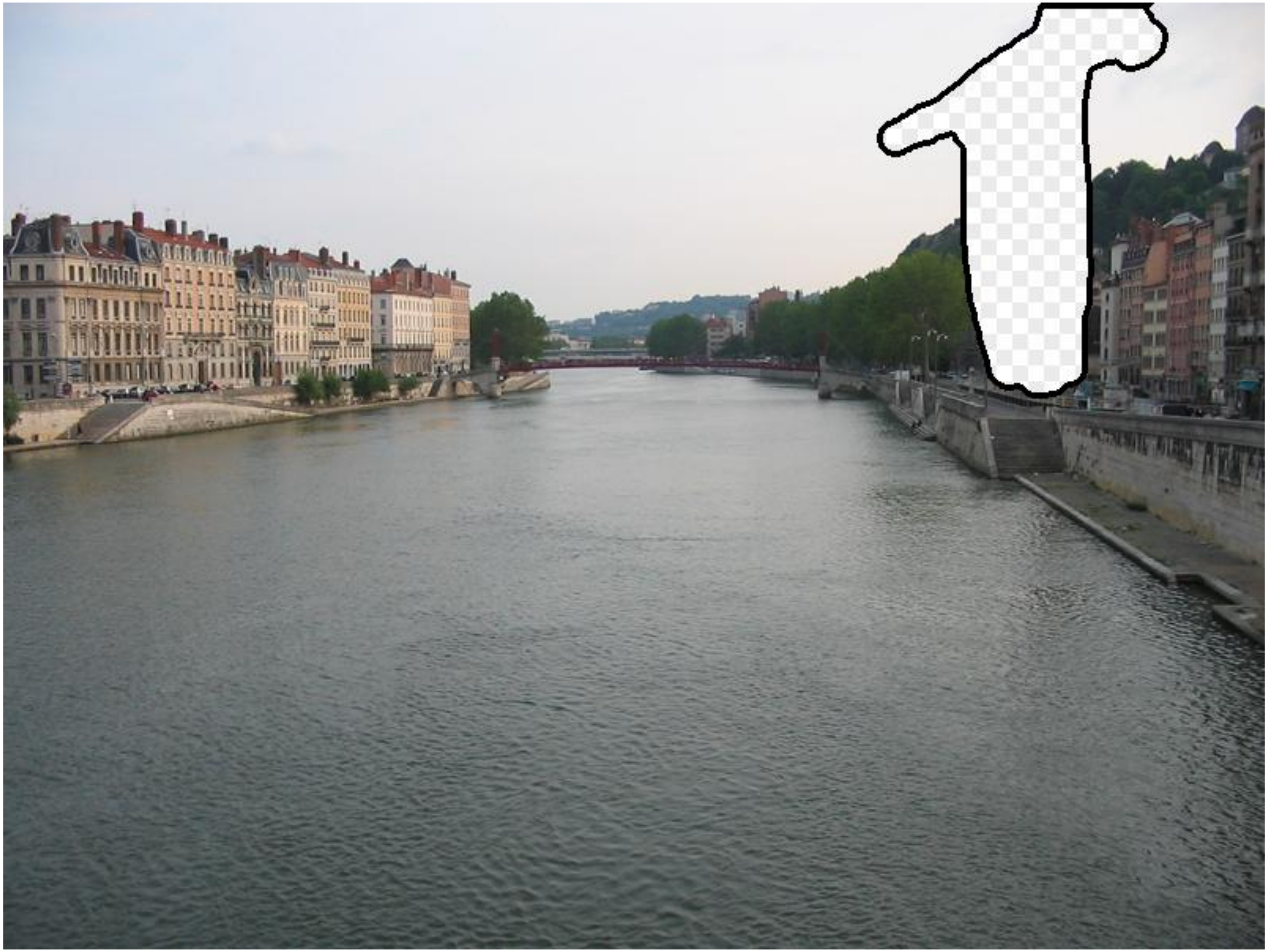
[Liu08]

# Image restoration



[Liu08]

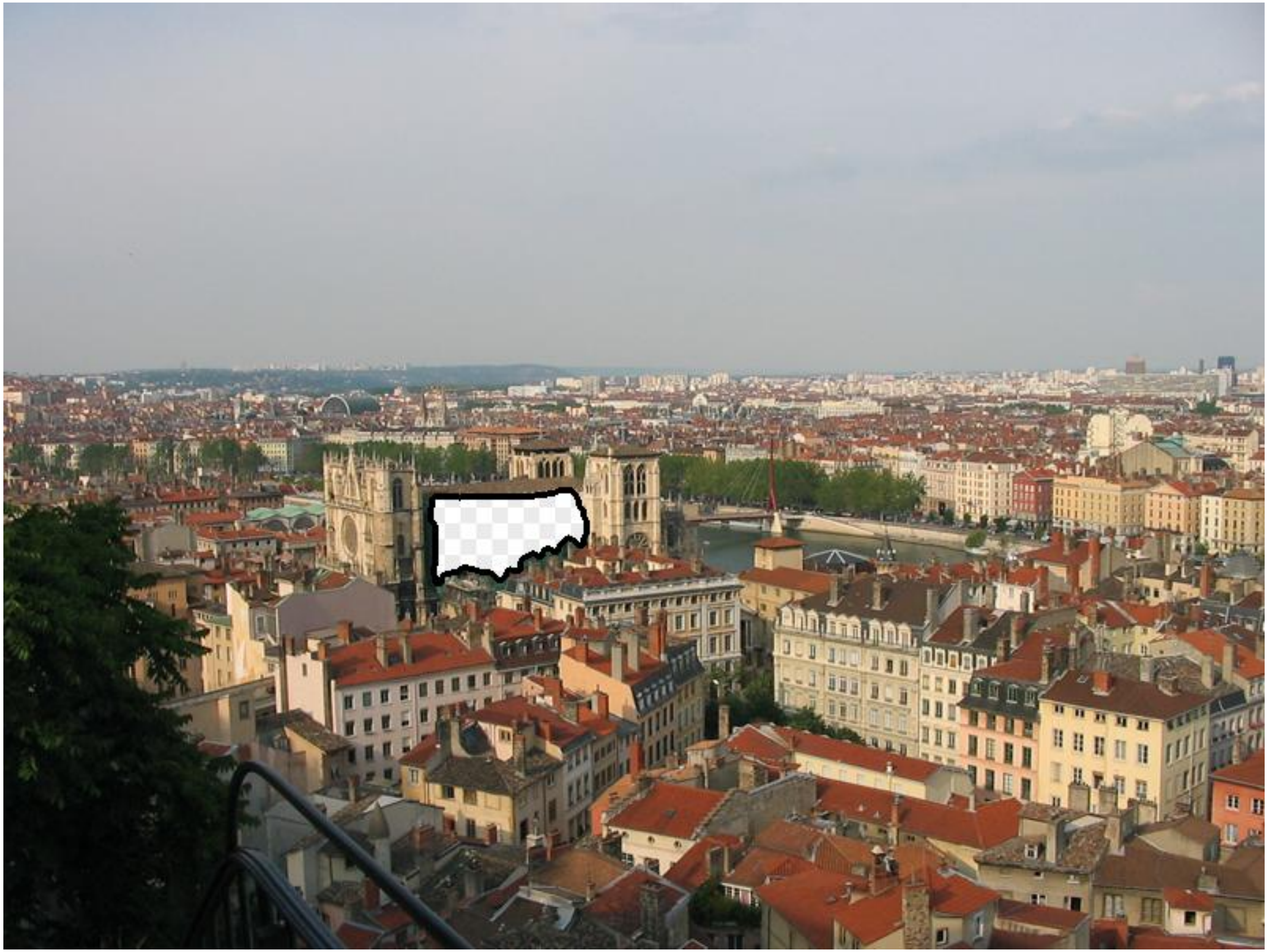
Hays and Efros, SIGGRAPH 2007



Hays and Efros, SIGGRAPH 2007

Hays and Efros, SIGGRAPH 2007

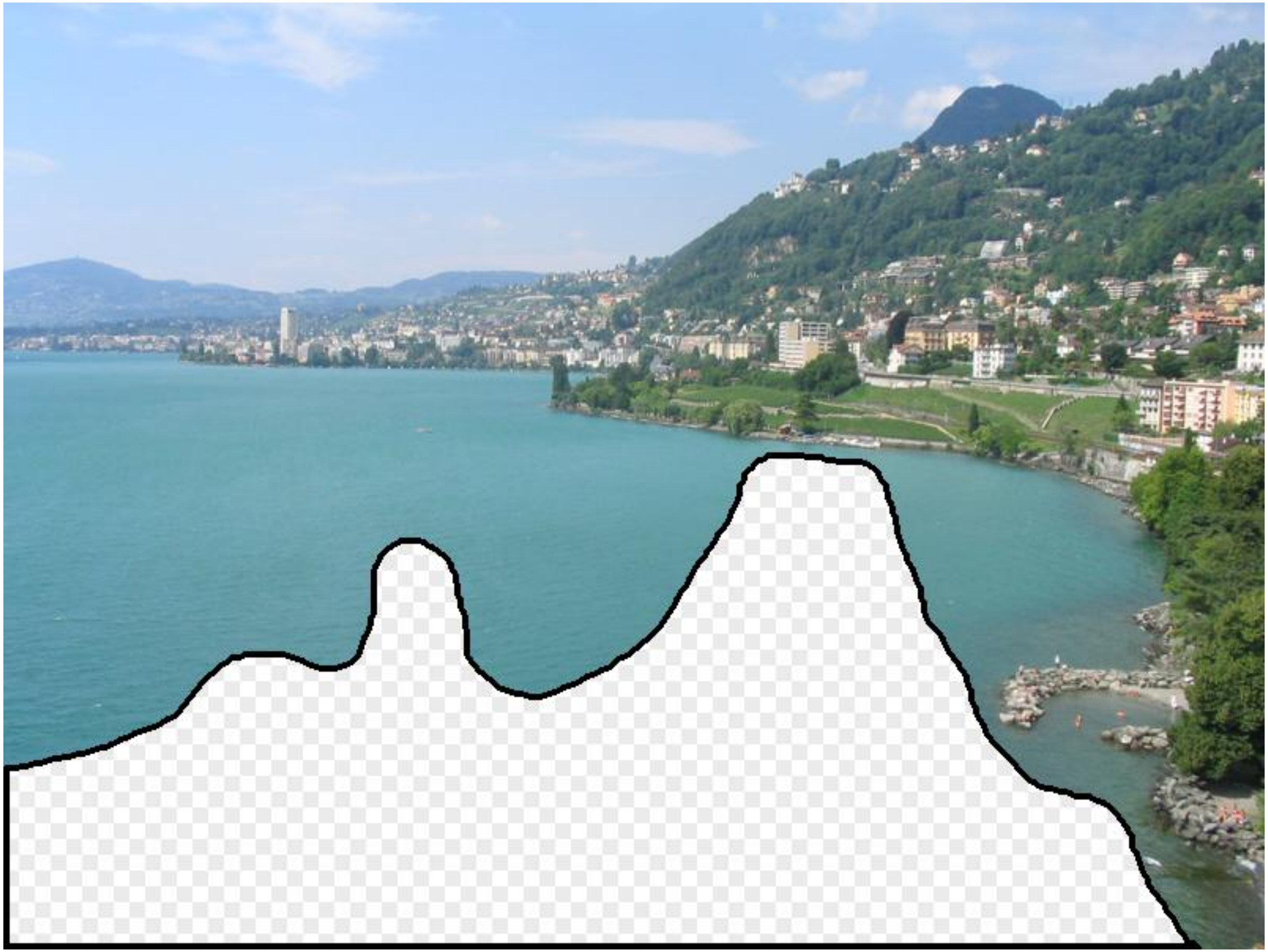
Hays and Efros, SIGGRAPH 2007



Hays and Efros, SIGGRAPH 2007

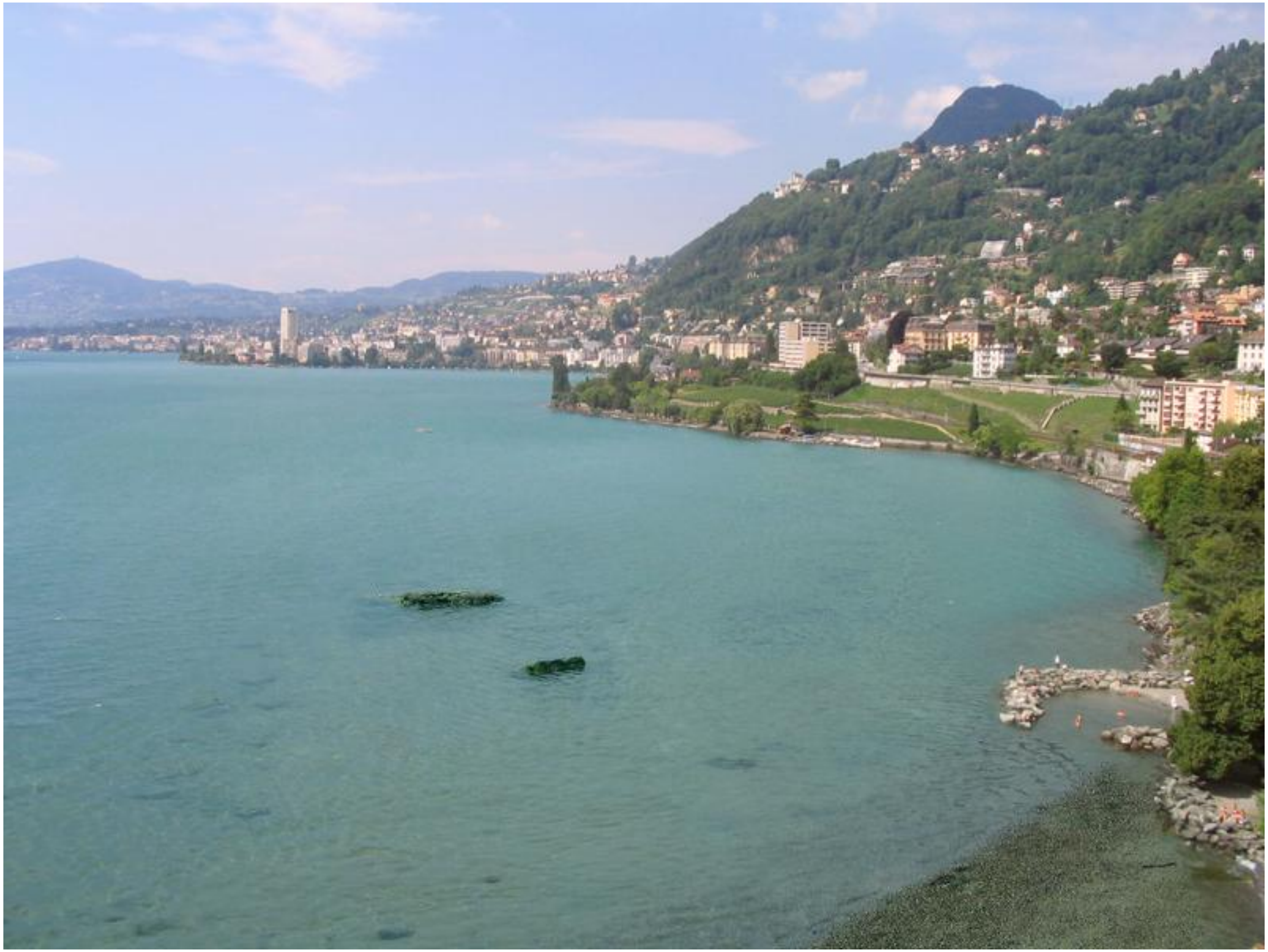
Hays and Efros, SIGGRAPH 2007

[Hays07]

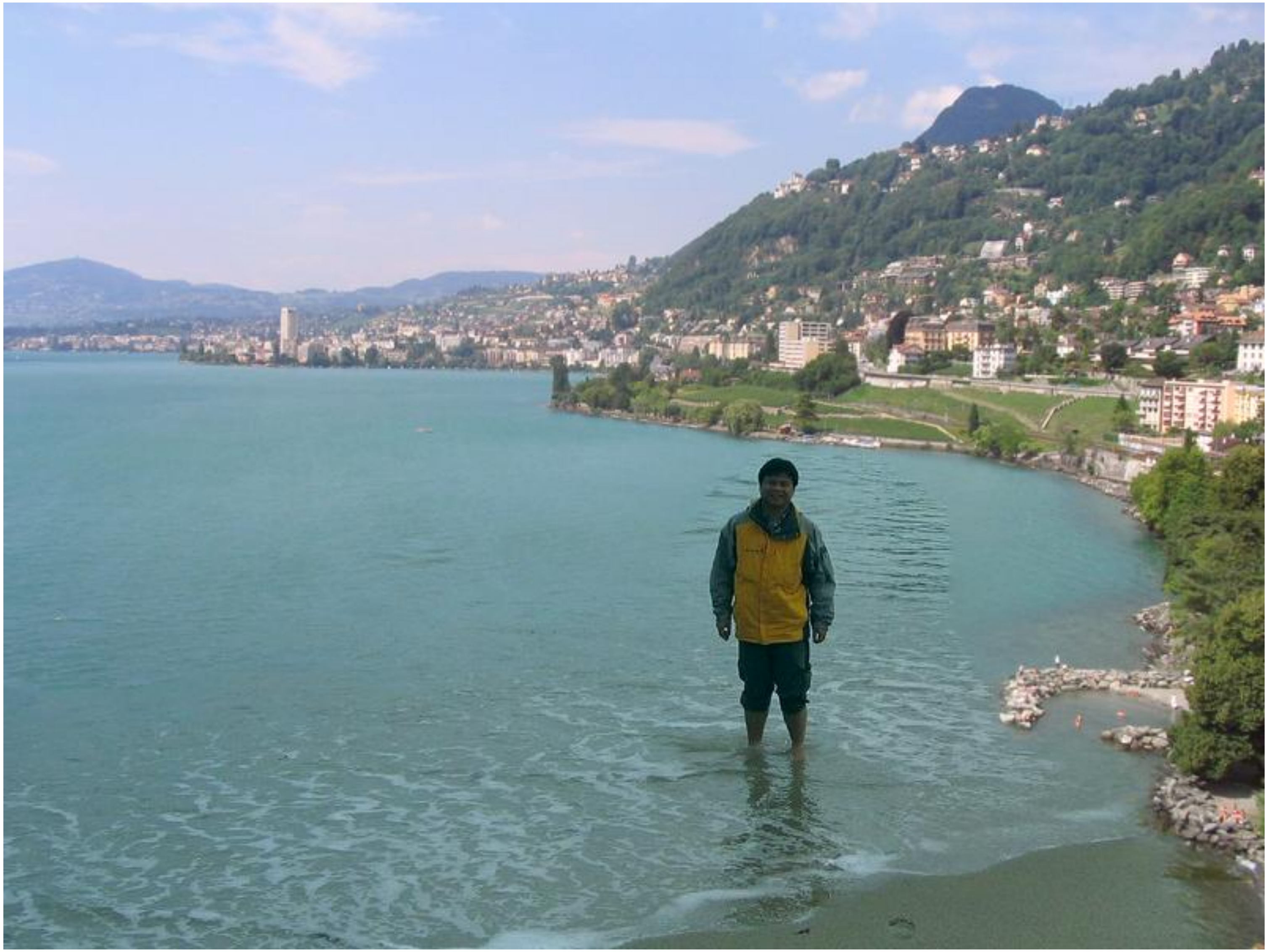


Hays and Efros, SIGGRAPH 2007

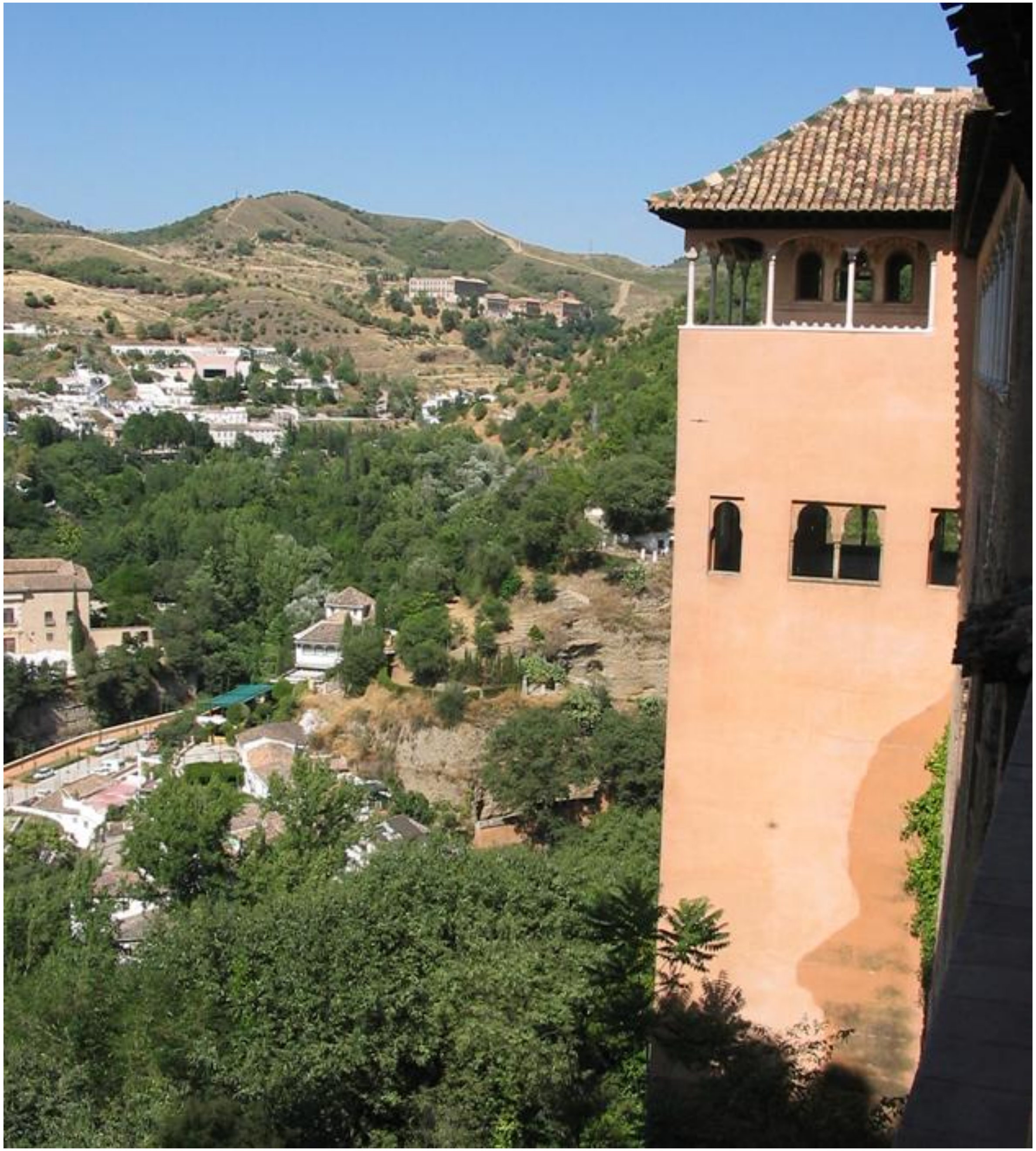
Hays and Efros, SIGGRAPH 2007



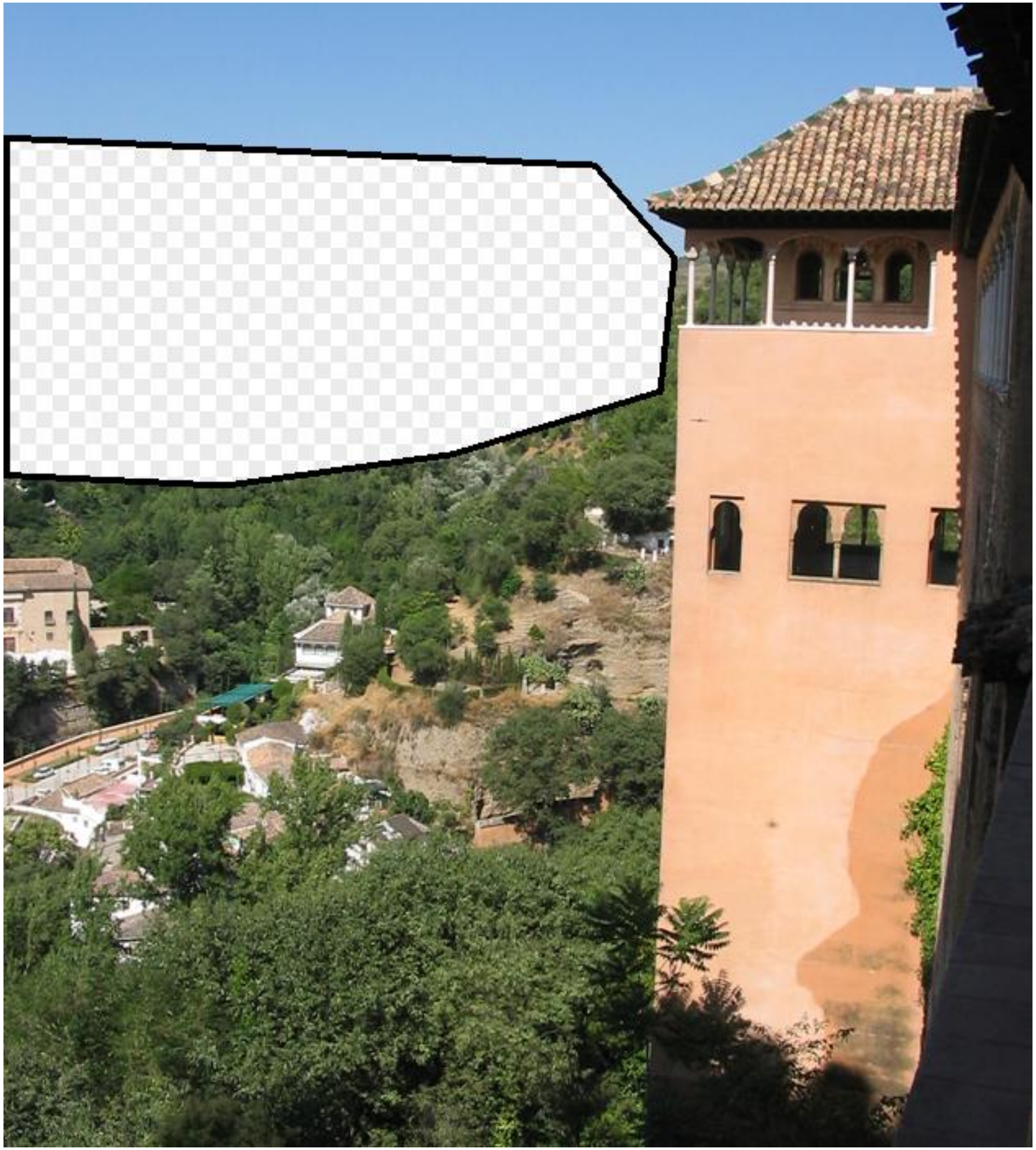
Hays and Efros, SIGGRAPH 2007



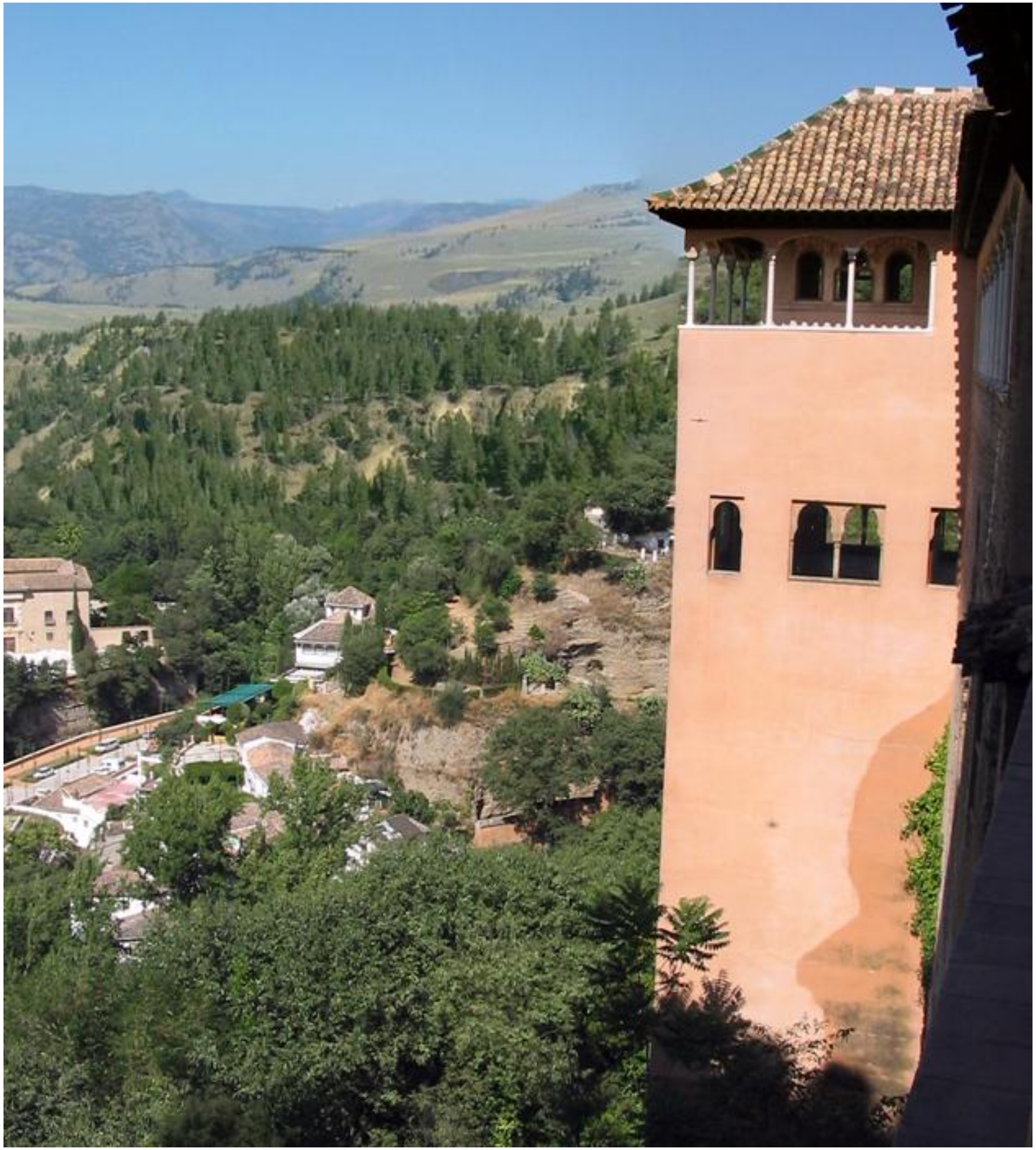
Hays and Efros, SIGGRAPH 2007



Hays and Efros, SIGGRAPH 2007

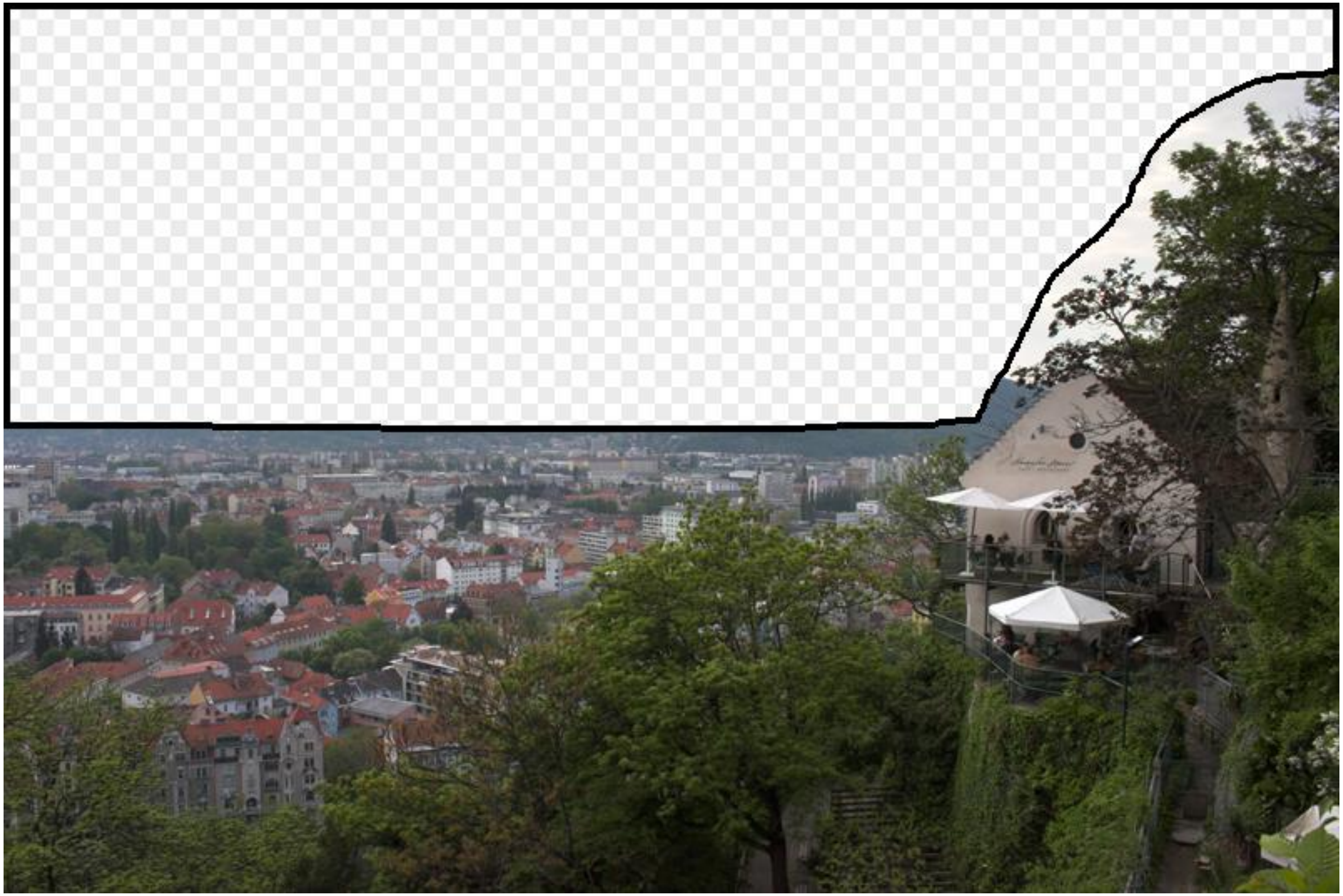


Hays and Efros, SIGGRAPH 2007



Hays and Efros, SIGGRAPH 2007

Hays and Efros, SIGGRAPH 2007



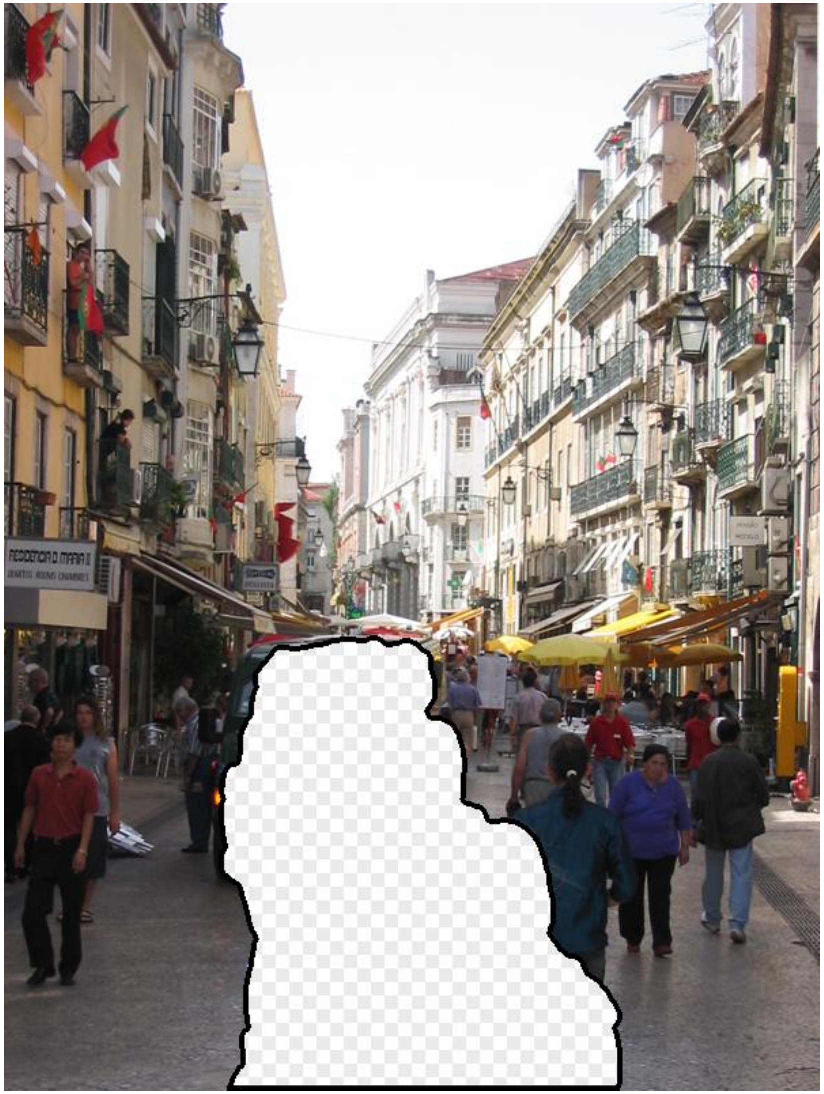
Hays and Efros, SIGGRAPH 2007

Hays and Efros, SIGGRAPH 2007

# Failures

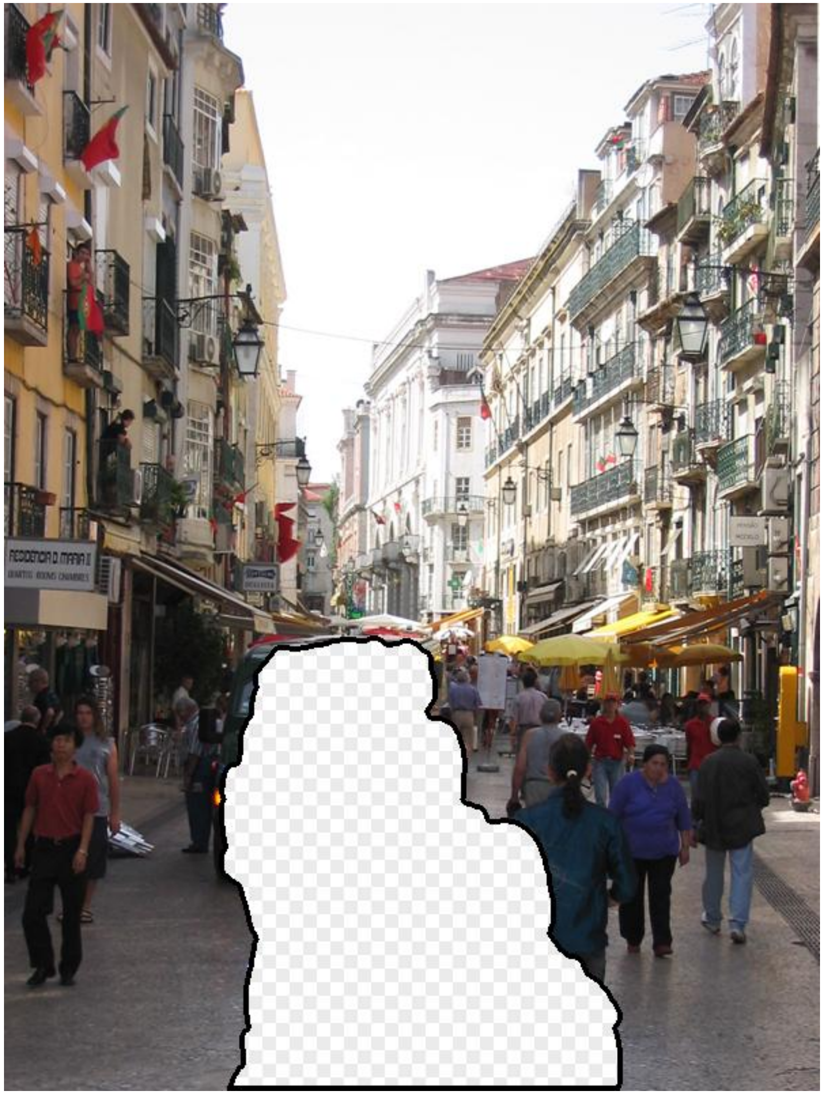
Hays and Efros, SIGGRAPH 2007

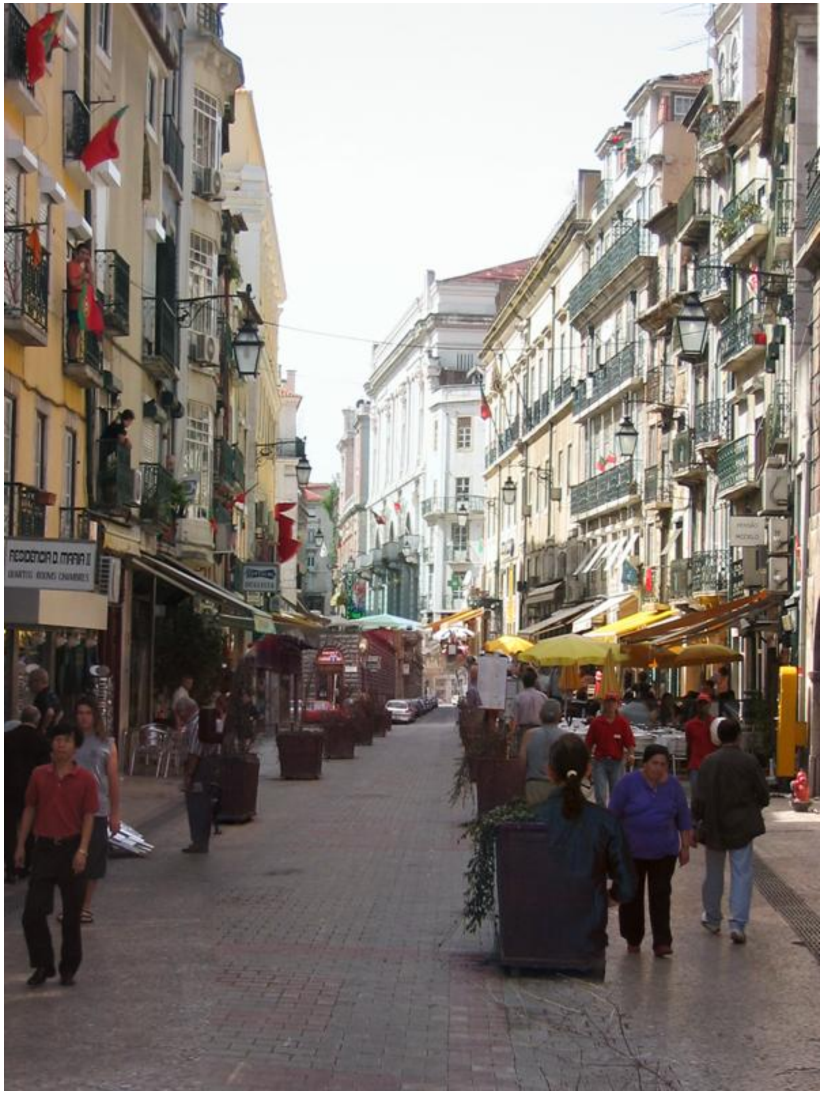
# Failures



Hays and Efros, SIGGRAPH 2007

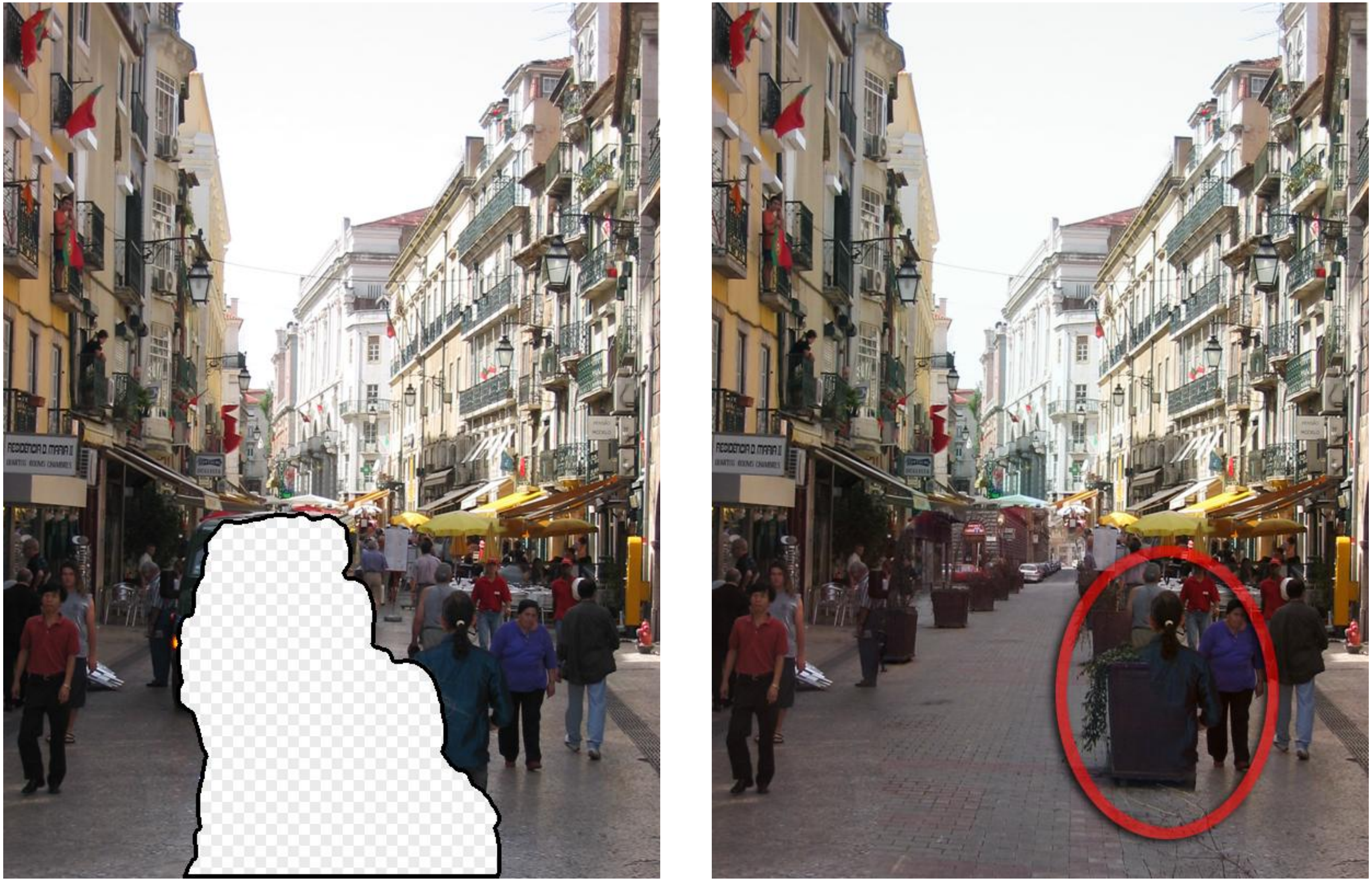
# Failures





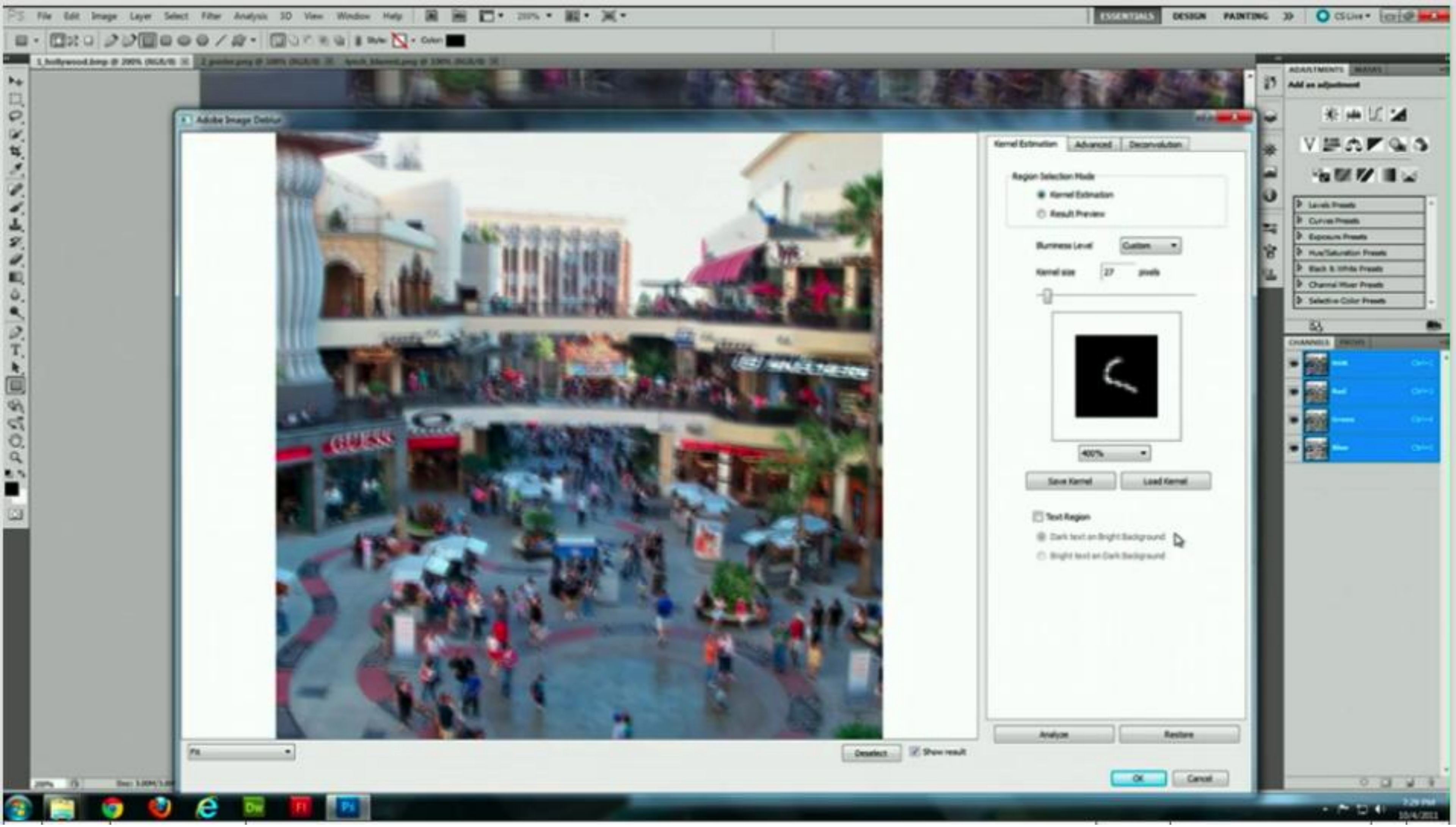
Hays and Efros, SIGGRAPH 2007

# Failures



Hays and Efros, SIGGRAPH 2007

# De-blurring



Adobe Image Deblur

Kernel Estimation | Advanced | Deconvolution

Region Selection Mode
● Kernel Estimation
○ Result Preview

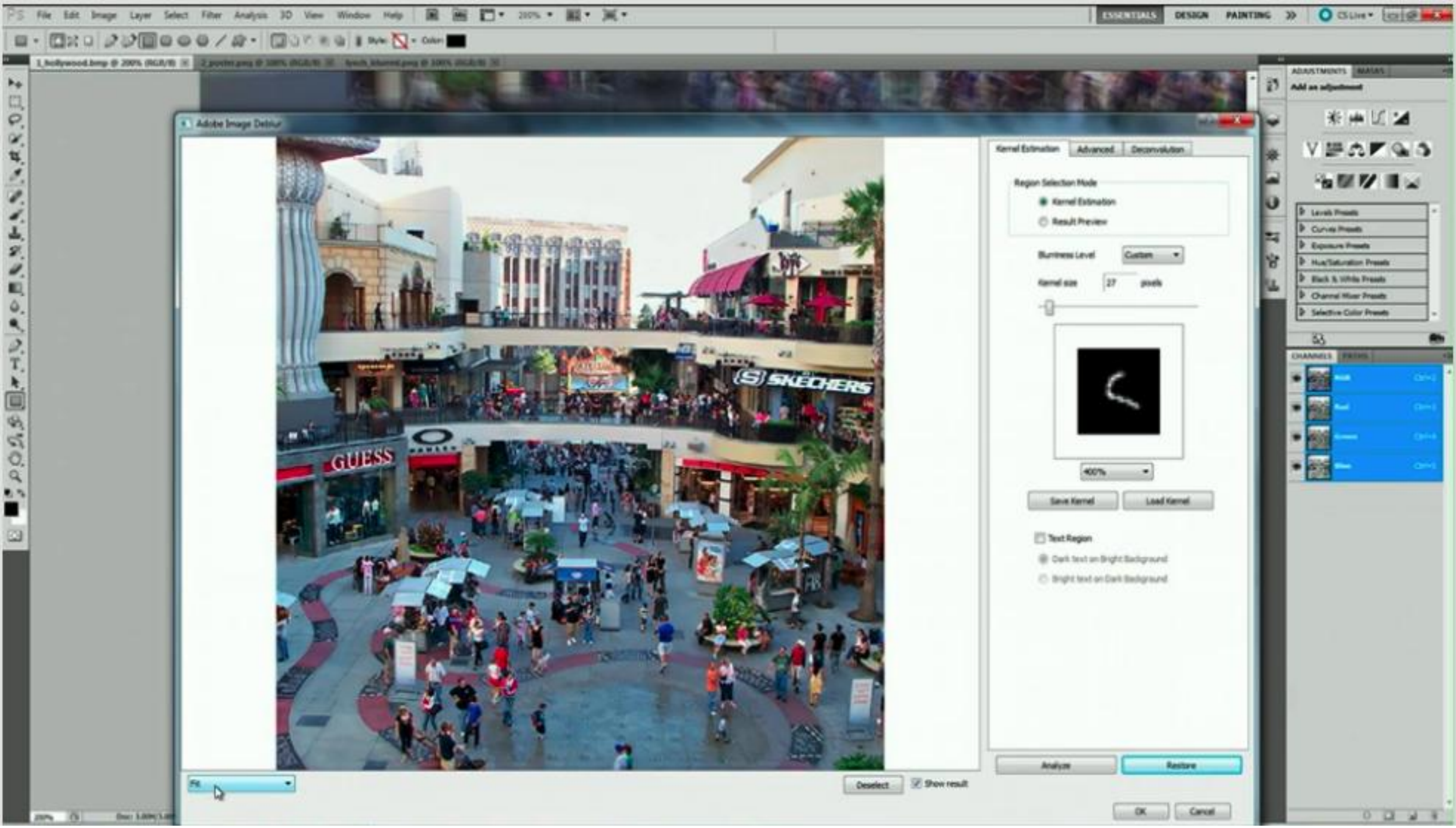
Blurriness Level: Custom

Kernel size: 27 pixels

☐ Text Region
● Dark text on Bright Background
○ Bright text on Dark Background

http://tv.adobe.com/watch/max-2011-sneak-peeks/max-2011-sneak-peek-image-deblurring/

# De-blurring



Region Selection Mode
● Kernel Estimation
○ Result Preview

Blurriness Level: Custom

Kernel size: 27 pixels

400%

□ Text Region
● Dark text on Bright Background
○ Bright text on Dark Background

Fit

☑ Show result

http://tv.adobe.com/watch/max-2011-sneak-peeks/max-2011-sneak-peek-image-deblurring/

# Automatic 3-D reconstruction

- From Internet photo collections [Snavely06]

“Statue of Liberty” “Half Dome, Yosemite” “Colosseum, Rome”

Flickr photos

3D model