Let’s look at some lakefront property

*actually fences / walls
Opportunities of Scale

Computer Vision

James Hays

Many slides from James Hays, Alyosha Efros, and Derek Hoiem

Graphic from Antonio Torralba
Outline

Opportunities of Scale: Data-driven methods

- The Unreasonable Effectiveness of Data
- Scene Completion
- Im2gps
- Recognition via Tiny Images
- Project 5 Intro
Computer Vision so far

• The geometry of image formation
  – Ancient / Renaissance

• Signal processing / Convolution
  – 1800, but really the 50’s and 60’s

• Hand-designed Features for recognition, either instance-level or categorical
  – 1999 (SIFT), 2003 (Video Google), 2005 (Dalal-Triggs), 2006 (spatial pyramid)

• Learning from Data
  – 1991 (EigenFaces) but late 90’s to now especially
What has changed in the last decade?

- The Internet
- Crowdsourcing
- Learning representations from the data these sources provide (deep learning)
Google and massive data-driven algorithms

A.I. for the postmodern world:

– all questions have already been answered...many times, in many ways

– Google is dumb, the “intelligence” is in the data
The Unreasonable Effectiveness of Data

Peter Norvig
Google
If a machine can convincingly simulate an intelligent conversation, does it necessarily understand? In the experiment, Searle imagines himself in a room, acting as a computer by manually executing a program that convincingly simulates the behavior of a native Chinese speaker.

Most of the discussion consists of attempts to refute it. "The overwhelming majority," notes BBS editor Stevan Harnad, "still think that the Chinese Room Argument is dead wrong." The sheer volume of the literature that has grown up around it inspired Pat Hayes to quip that the field of cognitive science ought to be redefined as "the ongoing research program of showing Searle's Chinese Room Argument to be false."
Questions from the piece:

Q1. Does the Chinese Room argument prove the impossibility of machine consciousness?
A1: Hell no. ... See More

Can Machines Become Moral?
The question is heard more and more often, both from those who think that machines cannot become moral, and who think that to believe otherwise is a dangerous illusion, and from those who think that machines must become moral....

BIGQUESTIONSONLINE.COM | BY DON HOWARD
Big Idea

• Do we need computer vision systems to have strong AI-like reasoning about our world?
• What if invariance / generalization isn’t actually the core difficulty of computer vision?
• What if we can perform high level reasoning with brute-force, data-driven algorithms?
Scene Completion

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

http://graphics.cs.cmu.edu/projects/scene-completion/
How it works

• Find a similar image from a large dataset
• Blend a region from that image into the hole
Hopefully, if you have enough images, the dataset will contain very similar images that you can find with simple matching methods.
How many images is enough?
Nearest neighbors from a collection of 20 thousand images
Nearest neighbors from a collection of 2 million images
Image Data on the Internet

• Flickr (as of Sept. 19\textsuperscript{th}, 2010)
  – 5 billion photographs
  – 100+ million geotagged images
• Facebook (as of 2009)
  – 15 billion

Image Data on the Internet

• Flickr (as of Nov 2013)
  – 10 billion photographs
  – 100+ million geotagged images
  – 3.5 million a day

• Facebook (as of Sept 2013)
  – 250 billion+
  – 300 million a day

• Instagram
  – 55 million a day
Scene Completion: how it works

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

Scene Gist Descriptor
(Oliva and Torralba 2001)
Scene Descriptor

Scene Gist Descriptor
(Oliva and Torralba 2001)
2 Million Flickr Images
Context Matching
Graph cut + Poisson blending
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
... 200 scene matches
Which is the original?
Outline

Opportunities of Scale: Data-driven methods

– The Unreasonable Effectiveness of Data
– Scene Completion
– Im2gps
– Recognition via Tiny Images
– Project 5 Intro
Hopefully, if you have enough images, the dataset will contain very similar images that you can find with simple matching methods.
Graph cut + Poisson blending
im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

http://graphics.cs.cmu.edu/projects/im2gps/
How much can an image tell about its geographic location?
Nearest Neighbors according to gist + bag of SIFT + color histogram + a few others
Example Scene Matches
Voting Scheme
Effect of Dataset Size

![Graph showing the effect of dataset size on the percentage of geolocations within 200km. The green line represents the first nearest neighbor scene match, and the red dashed line represents chance - random scenes. The x-axis represents the database size in thousands of images (log scale), and the y-axis represents the percentage of geolocations within 200km.]
Population density ranking

High Predicted Density

Low Predicted Density
Where is This?

Where is This?
Where are These?

15:14, June 18th, 2006

16:31, June 18th, 2006
Where are These?

15:14, June 18th, 2006

16:31, June 18th, 2006

17:24, June 19th, 2006
Results

- **im2gps** – 10% (geo-loc within 400 km)
- **temporal im2gps** – 56%
Tiny Images

80 million tiny images: a large dataset for non-parametric object and scene recognition

http://groups.csail.mit.edu/vision/TinyImages/
c) Segmentation of 32x32 images
Human Scene Recognition

![Graph showing the correct recognition rate and true positive rate for different image resolutions. The graph compares color images and grayscale images. The x-axis represents image resolution, and the y-axis represents the correct recognition rate and true positive rate. The graph indicates that color images have a higher correct recognition rate compared to grayscale images.](image-url)
Humans vs. Computers: Car-Image Classification

Humans for 32 pixel tall images

Various computer vision algorithms for full resolution images
Powers of 10

Number of images on my hard drive: \(10^4\)

Number of images seen during my first 10 years: \(10^8\)
(3 images/second \(\times\) 60 \(\times\) 60 \(\times\) 16 \(\times\) 365 \(\times\) 10 = 630720000)

Number of images seen by all humanity: \(10^{20}\)
106,456,367,669 humans\(^1\) \(\times\) 60 years \(\times\) 3 images/second \(\times\) 60 \(\times\) 60 \(\times\) 16 \(\times\) 365 = 1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: \(10^{88}\)

Number of all 32x32 images: \(10^{7373}\)
256 \(32^{*32}*3 \sim 10^{7373}\)
Scenes are unique
But not all scenes are so original
Lots Of Images

7,900
Lots Of Images

A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008
Lots
Of
Images
Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match
Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match
Summary

• With billions of images on the web, it’s often possible to find a close nearest neighbor

• In such cases, we can shortcut hard problems by “looking up” the answer, stealing the labels from our nearest neighbor

• For example, simple (or learned) associations can be used to synthesize background regions, colorize, or recognize objects
Project 5

• http://www.cc.gatech.edu/~hays/compvision/proj5/