Opportunities of Scale

Computational Photography
Derek Hoiem, University of Illinois

Most slides from Alyosha Efros
Graphic from Antonio Torralba
Today’s class

• Opportunities of Scale: Data-driven methods
  – Scene completion
  – Im2gps
  – 3D reconstruction
  – Colorizing
  – Infinite zoom/panorama
  – and much more...
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
My dog once ate three oranges, but then it died.

Mi perro se comió una vez tres naranjas, pero luego murió.
Chinese Room

• John Searle (1980)
Image Completion Example

[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]
What should the missing region contain?
Which is the original?

(a)

(b)

(c)
How it works

• Find a similar image from a large dataset
• Blend a region from that image into the hole
Trick: 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 (now)
  – 6 billion images per month
  – More than 100 petabytes of images/video

• Flickr (as of Sept. 19th, 2010)
  – 5 billion photographs
  – 100+ million geotagged images

• Imageshack (as of 2009)
  – 20 billion

• Facebook (as of 2009)
  – 15 billion

Image 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
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?
Diffusion Result
Efros and Leung result
im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images
How much can an image tell about its geographic location?
Example Scene Matches
Voting Scheme
Population density ranking
Where is This?

Where is This?
Where are These?

15:14, June 18th, 2006

16:31, June 18th, 2006
<table>
<thead>
<tr>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:14</td>
<td>Graffiti underpass</td>
</tr>
<tr>
<td>16:31</td>
<td>Big Ben</td>
</tr>
<tr>
<td>17:24</td>
<td>Athens cityscape</td>
</tr>
</tbody>
</table>

Where are These?

15:14, June 18\textsuperscript{th}, 2006

16:31, June 18\textsuperscript{th}, 2006

17:24, June 19\textsuperscript{th}, 2006
Results

• im2gps – 10% (geo-loc within 400 km)
• temporal im2gps – 56%
3D Reconstruction from Flickr

- Create detailed 3D scenes from thousands of consumer photographs
- Challenges include variations in season, lighting, occluding objects, etc.
3D Reconstruction from Flickr: How it works

1. Download ~10,000 images, convert to grayscale, compute SIFT++ keypoints

2. Match images
   1. Get similar images with vocabulary tree (like in recognition from last class)
   2. Match keypoints across similar images and perform geometric verification with RANSAC (similar to photo stitching)

3. Form a graph of matched images

4. 3D Reconstruction by triangulating points, bundle adjustment
Large-scale 3D Reconstruction

Useful references

• Dense reconstruction: “Towards Internet-scale Multi-view Stereo”, Furukawa et al., CVPR 2010
  http://grail.cs.washington.edu/software/cmvs/

• Sparse reconstruction: “Building Rome in a Day”, Goesler et al., ICCV 2009
  http://grail.cs.washington.edu/projects/rome/

• Code: Bundler Software
Photo Clip Art [SG’07]

Inserting a single object -- still very hard!

Lalonde et al, SIGGRAPH 2007
Photo Clip Art [SG’07]

Use database to find well-fitting object

Lalonde et al, SIGGRAPH 2007
Tour from a single image
Scene matching with camera transformations
Tour from a single image

Navigate the virtual space using intuitive motion controls
Video

http://www.youtube.com/watch?v=E0rboU10rPo
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

Correct recognition rate

True positive rate

Image resolution

a) Scene recognition
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 \(\times 60 \times 60 \times 16 \times 365 =\)


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

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

\(256^{32 \times 32} \sim 10^{7373}\)
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
Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match
Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match
Summary

- Many questions have been asked before, photos have been taken before.

- Sometimes, we can shortcut hard problems by looking up the answer.