Visual Data on the Internet

http://www.boingboing.net/2009/07/30/bb-video-send-me-a-l.html (starts at 2:40 min)
Big issues

• What is out there on the Internet? How do we get it? What can we do with it?
• How do we compute distances between images?
Subject-specific Data

Photos of Coliseum

Portraits of Bill Clinton
Much of Captured World is “generic”
Generic Data

street scenes

Food plates

faces

pedestrians
The Internet as a Data Source

- Social Networking Sites (e.g. Facebook, MySpace)
- Image Search Engines (e.g. Google, Bing)
- Photo Sharing Sites (e.g. Flickr, Picasa, Panoramio, photo.net, dpchallenge.com)
- Computer Vision Databases (e.g. CalTech 256, PASCAL VOC, LabelMe, Tiny Images, imagenet.org, ESP game, Squigl, Matchin)
How Big is Flickr?

• As of June 19th, 2009

• Total content:
  – 3.6 billion photographs
  – 100+ million geotagged images

• Public content:
  – 1.3 billion photographs
  – 74 million geotagged images
How Annotated is Flickr? (tag search)

- Party – 7,355,998
- Paris – 4,139,927
- Chair – 232,885
- Violin – 55,015
- Trashcan – 9,818
Trashcan Results

- http://www.flickr.com/search/?q=trashcan+NOT+party&m=tags&z=t&page=5
Is Generic Data useful?

A motivating example…
[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]
Diffusion Result
Efros and Leung result
Scene Matching for Image Completion
Scene Completion Result
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
Why does it work?
Nearest neighbors from a collection of 20 thousand images
Nearest neighbors from a collection of 2 million images
“Unreasonable Effectiveness of Data”  
[Halevy, Norvig, Pereira 2009]

• Parts of our world can be explained by elegant mathematics
  – physics, chemistry, astronomy, etc.

• But much cannot
  – psychology, economics, genetics, etc.

• Enter The Data!
  – Great advances in several fields:
    • e.g. speech recognition, machine translation
    • Case study: Google
• 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
How about visual data?

• text is simple:
  – clean, segmented, compact, 1D

• Visual data is much harder:
  – Noisy, unsegmented, high entropy, 2D/3D

Quick Overview
  Comparing Images
  Uses of Visual Data

  The Dangers of Data
Distance Metrics

- $\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} = \text{Euclidian distance of 5 units}$

- $50 - 50 = \text{Grayvalue distance of 50 values}$

- ? = ?
SSD says these are not similar
Tiny Images

- 80 million tiny images: a large dataset for non-parametric object and scene recognition
c) Segmentation of 32x32 images
Human Scene Recognition

![Graph showing the correct recognition rate and true positive rate with varying image resolution for color and grayscale images.]

- **Correct recognition rate**
- **True positive rate**

**Legend:**
- Color image
- Grayscale

**Axes:**
- Image resolution
- a) Scene recognition
Tiny Images Project Page

http://groups.csail.mit.edu/vision/TinyImages/
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 * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: \(10^{20}\)
106,456,367,669 humans \(^1\) * 60 years * 3 images/second * 60 * 60 * 16 * 365 =

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

Number of all 32x32 images:
\(256^{32^*3^2} \sim 10^{7373}\)
Scenes are unique
But not all scenes are so original
But not all scenes are so original
Lots Of Images

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

Of

Images

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

Grayscale input High resolution

Colorization of input using average

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

• Many images have ambiguous orientation
• Look at top 25% by confidence:
• Examples of high and low confidence images:
Automatic Orientation Examples
Tiny Images Discussion

• Why SSD?
• Can we build a better image descriptor?
Image Representations: Histograms

global histogram

- Represent distribution of features
  - Color, texture, depth, …
Image Representations: Histograms

Joint histogram
- Requires lots of data
- Loss of resolution to avoid empty bins

Marginal histogram
- Requires independent features
- More data/bin than joint histogram

Images from Dave Kauchak
Image Representations: Histograms

Adaptive binning

- Better data/bin distribution, fewer empty bins
- Can adapt available resolution to relative feature importance

Images from Dave Kauchak
Clusters / Signatures

- “super-adaptive” binning
- Does not require discretization along any fixed axis

Images from Dave Kauchak
Issue: How to Compare Histograms?

**Bin-by-bin comparison**
- Sensitive to bin size.
- Could use wider bins ...
- ... but at a loss of resolution

**Cross-bin comparison**
- How much cross-bin influence is necessary/sufficient?
Red Car Retrievals (Color histograms)

\[ \chi^2(h_i, h_j) = \frac{1}{2} \sum_{m=1}^{K} \frac{[h_i(m) - h_j(m)]^2}{h_i(m) + h_j(m)} \]

Histogram matching distance
Capturing the “essence” of texture

...for real images

We don’t want an actual texture realization, we want a texture invariant

What are the tools for capturing statistical properties of some signal?
Multi-scale filter decomposition

Filter bank

Input image
Filter response histograms
Heeger & Bergen ‘95

Start with a noise image as output

Main loop:

• Match pixel histogram of output image to input
• Decompose input and output images using multi-scale filter bank (Steerable Pyramid)
• Match subband histograms of input and output pyramids
• Reconstruct input and output images (collapse the pyramids)
Image Descriptors

- Blur + SSD
- Color / Texture histograms
- Gradients + Histogram (GIST, SIFT, HOG, etc)
- “Bag of Visual Words”
Gist Scene Descriptor
Gist Scene Descriptor

Gist scene descriptor
(Oliva and Torralba 2001)
Gist Scene Descriptor

Gist scene descriptor
( Oliva and Torralba 2001 )

Hays and Efros, SIGGRAPH 2007
Gist Scene Descriptor

Gist scene descriptor
( Oliva and Torralba 2001 )

Hays and Efros, SIGGRAPH 2007
im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images
How much can an image tell about its geographic location?
Im2gps
Example Scene Matches
Voting Scheme
Data-driven categories
Elevation gradient = 112 m / km
Elevation gradient magnitude ranking
Figure 2. Global population density map.
Population density ranking
Figure 4. Global land cover classification map.
Barren or sparsely populated
Urban and built up
Snow and Ice
Water
Where is This?

Where is This?
Where are These?

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

16:31,
June 18\textsuperscript{th}, 2006
Where are These?

15:14, June 18th, 2006
16:31, June 18th, 2006
17:24, June 19th, 2006
Problem Statement

\[
\begin{align*}
\Delta T_1 & \quad \Delta T_2 & \quad \Delta T_3 & \quad \Delta T_4 & \quad \Delta T_5 \\
T_1 & \quad T_2 & \quad T_3 & \quad T_4 & \quad T_5
\end{align*}
\]
Time-Series Model

Hidden Markov Model
Spatially Varying Human Mobility Model

Derived directly from Flickr photographer movements
Locations and timesteps are quantized
1 Beijing
2 Beijing
Results

• im2gps – 10% (geo-loc within 400 km)
• temporal im2gps – 56%
Scene matching with camera transformations
Image representation

Original image

GIST
[Oliva and Torralba’01]

Color layout
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
3. Find a match to fill-in the missing pixels
4. Stitched rotation
5. Display on a cylinder
Scene matching with camera view transformations:
Forward motion

1. Move camera
2. View from the virtual camera
3. Find a match to replace pixels
Tour from a single image

Navigate the virtual space using intuitive motion controls
Video

http://www.youtube.com/watch?v=E0rboU10rPo
Using Data for Graphics...
Semantic Photo Synthesis [EG’06]

Photo Clip Art [SG’07]

Inserting a single object -- still very hard!

- object size, orientation
- scene illumination

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

Use database to find well-fitting object

Lalonde et al, SIGGRAPH 2007
SkyFinder [SG’09]

(a) blue + normal-sky + horizon + R3
(b) cloudy + normal-sky + horizon + R5
(c) sunset + landscape + horizon + sun
(e) sunset + landscape + horizon + R5
(f) blue + object-in-sky + R1
(g) cloudy + full-sky + R5

Tao, Yuan, Sun, SIGGRAPH 2009
Webcam Clip Art [SG Asia’09]

Object transfer

illuminant transfer

Lalonde et al, SIGGRAPH Asia 2009
Image Restoration using Online Photo Collections [ICCV’09]

Dale, Johnson, Sunkavalli, Matusik, Pfister, ICCV’09
The Dangers of Data
Bias

• Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
• But it’s not random samples of visual world
• Many sources of bias:
  – Sampling bias
  – Photographer bias
  – Social bias
Real Paris
Real Notre Dame
Sampling Bias

- People like to take pictures on vacation
Photographer Bias

- People want their pictures to be recognizable and/or interesting

vs.

[Images of Notre Dame and a cloudy sky]
Photographer Bias

- People follow photographic conventions

VS.
Social Bias

“100 Special Moments” by Jason Salavon
Social Bias

Mildred and Lisa

Source: U.S. Social Security Administration

Gallagher et al CVPR 2008
Social Bias

Gallagher et al CVPR 2008

Gallagher et al, CVPR 2009
Reducing / Changing Bias

• Autonomous capture methods can reduce / change bias
  – But it won’t go away completely
• Sometimes you can just pick your data to suit your problem, but not always…