What makes Big Visual Data hard?

Jun-Yan Zhu               Alexei A. Efros

UC Berkeley

© Quint Buchholz
Our Goals

1. To make you fall in love with Big Visual Data
   • Very difficult to handle.
   • but holds the key to achieving real visual understanding

2. To discuss the challenges and ask for help in tackling this Big Data Problem
Driven by Visual Data

- Texture Synthesis
- Dating Historical Images
- Seeing Through Water
- Unsupervised Object Discovery
- Action Recognition
- Illumination Estimation
- Inferring 3D from 2D
- Geo-location
Two Kinds of Things in the World

Navier-Stokes Equation

\[ \frac{\partial u}{\partial t} = -(u \cdot \nabla) u + v \nabla^2 u - \frac{1}{d} \nabla p + f \]

+ weather
+ location
+ …
Lots of data available
“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, genetics, economics,… visual understanding?

• Enter: The **Magic of Data**
  – Great advances in several fields:
    • e.g. speech recognition, machine translation, Google
The A.I. for the postmodern world
The Good News

Really stupid algorithms + Lots of Data
= “Unreasonable Effectiveness”
Big Visual Data

- Almost 90% of web traffic is visual!
- flickr: 6 billion images
- YouTube: 100 hours uploaded per minute
- Imgur: 1 billion images served daily
- Facebook: 70 billion images
- 3.5 trillion photographs

Cisco: 1 billion images served daily
The Bad News

Visual Data is difficult to handle

- text:
  - clean, segmented, compact, 1D, indexable

- Visual data:
  - Noisy, unsegmented, high entropy, 2D/3D
What makes Big Visual Data hard?

for Computers

1. Finding Correspondences
2. Mining Visual Data
3. Connecting Visual Data

for Human Beings

1. Visualizing Visual Data
2. Visual Communication
Computing distances is hard

\[ CLIME - CRIME \]

= hamming distance of 1 letter

= Euclidian distance of 5 units

= Grayvalue distance of 50 values

= ?
How similar are two pictures?
“It irritated him that the “dog” of 3:14 in the afternoon, seen in profile, should be indicated by the same noun as the dog of 3:15, seen frontally...”

“My memory, sir, is like a garbage heap.”

-- from *Funes the Memorious*

Organizing the “Garbage Heap”:

• Finding visual correspondences across data
• Mining Visual Data
• Connecting visual data to enable understanding (Visual Memex)
Improving Visual Correspondence
Lots of Tiny Images

• 80 million tiny images: a large dataset for non-parametric object and scene recognition
A. Torralba, R. Fergus, W.T. Freeman. PAMI 2008
Automatic Colorization
*SIGGRAPH* 2007
[Hays & Efros, SIGGRAPH’07]
Scene Descriptor

[Oliva & Torralba 01']
2 Million Flickr Images
10 nearest neighbors from a collection of 20,000 images
10 nearest neighbors from a collection of 2 million images
... 200 scene matches
Improving Visual Correspondence
Improving Visual Correspondence
Visual Data has a Long Tail

The rare is common!
One of these is from Paris

...this is Paris
We showed 20 subjects:
- 100 Random Street View Images
- 50 from Paris
- They classified Paris non-Paris
- Accuracy: 79%

How do they know?
We showed 20 subjects:
- 100 Random Street View Images
- 50 from Paris
- They classified Paris non-Paris
- Accuracy: 79%

How do they know?
Our Goal:

Given a large geo-tagged image dataset, we automatically discover visual elements that characterize a geographic location.
Our Hypothesis

• The visual elements that capture Paris:
  – Frequent: Occur often in Paris
  – Discriminative: Are not found outside Paris

Note: same idea as TF-IDF if we knew the elements.
Need Both Conditions

• Discriminative only:
Need Both Conditions

• Frequently occurring only:
The Data: Google Street View
K-means Clustering
not geo-informative!
visually incoherent!
Our Approach
Our Approach

I. Use geo-supervision
Our Approach

I. Use geo-supervision

II. Don’t partition the space top-down; build clusters bottom-up
Step 1: Nearest Neighbors for Every Patch

Using normalized correlation of HOG features as a distance metric
Step 2: Find the Parisian Clusters by Sorting

Sort by # Paris Neighbors
Rank: 1146
Good Patches may have Bad Neighbors!

- The naïve distance metric gives equal weight to the vertical bar and the sign.
Step 3: Updating the Similarity Function

- Learn a similarity function that separates Paris from not-Paris
  - I.e. reweight the dimensions of the feature space
  - Recast problem as classification & use SVMs [Cortes & Vapnik 1995]
  - [Shrivastava et al. 2011] uses a similar technique for image retrieval
Resulting Matches

patch weight
matches

Learn Weights

patch weight matches

Paris Not Paris

[Legend: Green Paris, Red Not Paris]
Step 4: Iterate using the new matches
Random Paris
Paris: A Few Top Elements
In the U.S.

Elements from San Francisco

Elements from Boston
Elements from Prague

Elements from London

Elements from Barcelona
Paris, France
What if we force the algorithm to match elements in Prague and London?
So, what makes Paris look like Paris?

• The proposed algorithm finds visual elements that appear frequently in Paris, and not elsewhere.

• What makes X look like X?
  – What makes a bathroom?
  – What makes a ‘50’s car?
  – What makes an Apple product?
Organizing the “Garbage Heap”

• Finding visual correspondences across data
• Mining Visual Data
• Connecting visual data to enable understanding (Visual Memex)
How to connect visual data to enable understanding (Visual Memex)

Nodes = instances
Edges = associations

types of edges:
• visual similarity
• spatial, temporal co-occurrence
• geometric structure
• language
• geography
• ...

[Malisiewicz and Efros 09]
How to build a Visual Memex with rich and dense relationships?

Image-Level Embedding
[van der Maaten and Hinton 2008]

Pixel-Level Graph
[Zhou et al 2014]

Object Graph
[Malisiewicz and Efros 2009]

2D Image to 3D shape
[Aubry et al 2014]
What makes Big Visual Data hard?

for Computers

1. Finding Correspondences
2. Mining Visual Data
3. Connecting Visual Data

for Human Beings

1. Visualizing Visual Data
2. Visual Communication
Data Visualization: the First Step

Data: Siggraph paper scores

4.5  4.0  3.5  2.5  3.5

Average score 3.6

Data: a collection of photos

Average image

...
Image Averaging

Sir Francis Galton
1822-1911

Average Images in Art

“60 passagers de 2e classe du metro, entre 9h et 11h” (1985)
Krzysztof Pruszkowski

“Dynamism of a cyclist” (2001)
James Campbell

Idris Khan
“100 Special Moments” (2004) by Jason Salavon

Newlyweds  Little Leaguer  Kids with Santa
Not so simple…

Jason Salavon
“Kids with Santa”

Google query result:
“kids with Santa”

Automatic Average
Why Difficult?

Google results

Visual Modes

Misaligned
“Object-Centric Averages” (2001) by Antonio Torralba

Manual Annotation and Alignment

Average Image
With Alignment

Google results

Visual Modes

Misaligned  Aligned
Our Goal:

An interactive system to rapidly explore and align a large image collection using image averaging.
Weighted Averages Overview

Image Collection \( \{I_1 \cdots I_N\} \) (e.g. “Kids with Santa” images)

Average \( I_{avg} \)

Image Weights \( \{s_1 \cdots s_N\} \)

\[
I_{avg} = \sum_{i=1}^{N} \sum_{i=1}^{N} I_i s_i
\]
Zappos “Shoes” (5, 703 Images)

Sketching Brush

ShadowDraw [Lee et al. 2011]
“Face” Dataset (13,233 Images)

Coloring Brush
Coloring Brush

\[ I_1 \]

\[ I_2 \]

Average

Weight \[ \Rightarrow \]

\[ S_i + \text{similarity}(\cdot, \cdot) \]
Flickr + Google Query: ‘Eiffel Tower’ (412 Images)

Sketching Brush + Coloring Brush
How to Start?

Blurry Average

Explorer Brush
Explorer Brush: Select a Local Mode

Local Visual Modes

Visual Mode Discovery

Average

\[ s_i = s_i + \text{similarity}(\text{Image}_1, \text{Image}_2) \]

Mid-level Discriminative Patch Discovery
[Doersch et al. 2012]
Google Query ‘Church’ (11,007 Images)
Weighted Averages + Alignment

Image Collection \(\{I_1 \cdots I_N\}\) (e.g. “Kids with Santa” images)

Average \(I_{\text{avg}}\)

Image Weights \(\{s_1 \cdots s_N\}\)
Image Alignment

User Edit

Image 1

Image 2

Average Image

Mean Position
Flickr + Google Query
‘Bridge of Sighs’
(829 Images)

Bridge of Sighs
Oxford
Image Warping

User Edits

Image 1

Image 2

Average Image

Moving Least Square

[Schaefer et al. 2006]
Creating Multiple Averages

Google Query ‘Kids with Santa’ (1,640 Images)
Automatic Clustering

- K-means, GMM
- Spectral Clustering
  - e.g. [Shi and Malik 2000]
- Discriminative Clustering
  - e.g. [Hoai and Zisserman 2013]
Google Query ‘Wedding Kiss’ (16, 868 Images)

Automatic Clustering

K-means

Spectral Clustering [Shi and Malik 2000]

Discriminative Clustering [Hoai and Zisserman 2013]
Automatic Alignment

[Learned-Miller 06]
[Huang et al. 07]

[Mattar et al. 12]
Interactive Clustering and Alignment

Average image
Our Contribution:

User-Guided Clustering

+ 

User-Guided Alignment
Face Keypoint Alignment

“Average Face by Country”
using FaceResearch.org

[Cootes et al. 1998]

[Blinz & Vetter, 1999]
Different Cat Breeds (Simple Average)
Different Cat Breeds (Our Result)

Abyssinian  Sphynx  Birman  Bombay  Egyptian Mau  Ragdoll
British Shorthair  Persian  Maine Coon  Russian Blue  Siamese  Bengal
Application: Keypoint Annotation

Car Parts Annotation

Average Image
Application: Online Shopping

Recommended Products

BUY NOW!
What makes Big Visual Data hard?

for Computers
1. Finding Correspondences
2. Mining Visual Data
3. Connecting Visual Data

for Human Beings
1. Visualizing Visual Data
2. Visual Communication
How to connect Humans’ Mental Picture to Big Visual Data?

The Language Bottleneck

words

Mental Picture

The Identi-Kit System

Forensic Sketch

Shoes

Image
THANK YOU!