ECS 289G: Visual Recognition
Fall 2015
Yong Jae Lee
Department of Computer Science
Plan for today

• Questions?
• Sign-up for paper
• Research overview
Standard supervised visual learning

- Number of training images required can be costly
- Assumes closed-world setting where all categories are known
Unsupervised visual discovery

Visual world

Discovered categories
Unsupervised visual discovery

Visual world

Object segmentations in images and video
Unsupervised visual discovery

- No human to explicitly guide visual recognition process
Why visual discovery?

Exploring new environments
Why visual discovery?

Summarization

MSR Sensecam
Why visual discovery?

- **flickr**: 6 billion images
- **facebook**: 70 billion images
- **imgur**: 1 billion images served daily
- **YouTube**: 10 billion images
- **photobucket**: 100 hours uploaded per minute

From **Cisco**: Almost 90% of web traffic is visual!

**Most of it is unlabeled!!**
Inputs today

Understand and organize and index all this data!!
Let’s first explore...

what we can do with big data!
Everyday use of big data: Predictive text
Big visual data

From Cisco:
Almost 90% of web traffic is visual!

flickr
6 billion images

facebook
70 billion images

imgur
1 billion images served daily

YouTube
10 billion images

100 hours uploaded per minute
Predictive drawing?
Drawing examples
Drawing examples

Subject 1

Subject 2

We need help..
Tracing
Drawing a face
Our Idea: ShadowDraw

User study drawings

[Lee, Zitnick, Cohen, ShadowDraw, SIGGRAPH 2011]
Creating shadows

Collection of images...

• Shadow gives gist of many images simultaneously
Creating shadows

- Requires partial matching and dynamic updates
Approach overview

Database (offline)

Database Image → Edges → For each sub-window → Min Hash

Query time (online)

User Drawing → For each sub-window → Histogram (ID x dx x dy) → Top 100

Verify → Fine Alignment → Spatial Scoring → Shadow
Blending weights

Shadows are the composite of many images
Rendering

Pen Strokes + Pen Position = Output

* Shadow
User studies

- 30,000 images, 20 categories
- 16 drawers (8 men, 8 women), 8 evaluators
- 5 objects (shoe, face, bicycle, butterfly, rabbit)
User studies

With ShadowDraw

Without ShadowDraw

Good drawers
User studies

With ShadowDraw

Without ShadowDraw

Bad drawers
User studies

With ShadowDraw

Without ShadowDraw

Average drawers
User studies

- Significant improvement for "Average" group
User studies

After training all users improved:

- Subject’s *personal style* is maintained!
AverageExplorer: Interactive Exploration and Alignment of Visual Data Collections

Yong Jae Lee
University of California, Davis

SIGGRAPH 2014 work with Jun-Yan Zhu and Alexei A. Efros
Digital Dark Matter

[Perona 2010]
Image Averaging

Sir Francis Galton
1822-1911

Multiple Individuals

Composite

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

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

Newlyweds
Little Leaguer
Kids with Santa
AverageExplorer

• video
Research goal: Visual discovery

- No human to explicitly guide visual recognition process
Intuition: context-aware discovery

Previous methods:
Cluster images/regions based on their appearance


Our idea:
Let familiar objects serve as context for unfamiliar objects
Intuition: context-aware discovery

[Lee & Grauman, Object-Graphs, CVPR 2010, TPAMI 2012]
Model the topology of category predictions relative to the unknown (unfamiliar) region.
Object-graphs

An unknown region within an image

Closest nodes in its object-graph

Consider spatially near regions above and below, record distributions for each known class.

$$g(s) = [ H_0(s), H_1(s), \ldots, H_R(s) ]$$
Example object-graphs

- Colors indicate the predicted known category (max posterior)

- unknown
- building
- sky
- road
Clusters from region-region affinities

$K(s_i, s_j) = K_{app}(s_i, s_j) + K_{obj-graph}(s_i, s_j)$

Object-level context provides more robust affinities
Example discoveries
Mining differences

• Recent methods mine *specific* visual patterns

**Visual world**

Paris

non-Paris

Mid-level visual elements

Problem

• Much in our visual world undergoes a *gradual change*

*Temporal:*

![Fashion Timeline - Short History of Women's Dress and Style](image)

<table>
<thead>
<tr>
<th>Decade</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>1910's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>1920's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>1930's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>Early 1940's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>Late 1940's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>1950's</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>1960's</td>
<td><img src="image" alt="Image" /></td>
</tr>
</tbody>
</table>

*Timeline of Coca-Cola Logos:*

- 1887-1900
- 1900-1941
- 1941-1969
- 1958-1969
- 1969-1987
• Much in our visual world undergoes a *gradual change*

**Spatial:**

![Map illustrating spatial distribution](image-url)
when? 1972
where? Krakow, Poland

“The View From Your Window” challenge
Key Idea

1) Establish connections

1926 1947 1975

“closed-world”

2) Model style-specific differences

[Lee, Efros, Hebert, Style-aware mid-level representation, ICCV 2013]
Mining style-sensitive elements

(a) Peaky (low-entropy) clusters
Mining style-sensitive elements

(b) Uniform (high-entropy) clusters
Making visual connections

• Take top-ranked clusters to build correspondences
Making visual connections

• Train a detector (HoG + linear SVM) [Singh et al. 2012]
Making visual connections

Top detection per decade

[Singh et al. 2012]
Making visual connections

- We expect style to change gradually...

1920s

1930s

1940s

Natural world “background” dataset
Making visual connections

1920s  1930s  1940s  1950s  1960s  1970s  1980s  1990s

Top detection per decade
Results: Learned styles

Average of top predictions per decade
Mining spatial extent of relative attributes

[Attribute: “high-at-the-heel”]

[strong, ..., weak]

[Xiao & Lee, Spatial Extent of Relative Attributes, ICCV 2015]
Mining spatial extent of relative attributes

[Xiao & Lee, Spatial Extent of Relative Attributes, ICCV 2015]
Mining spatial extent of relative attributes

[Xiao & Lee, Spatial Extent of Relative Attributes, ICCV 2015]
Collect-Cut

Discovered Ensemble from Unlabeled Multi-Object Images

Collect-Cut (ours)  Best Bottom-up (with multi-segs)

Unsupervised Segmentation Examples

[Lee & Grauman, Collect-Cut, CVPR 2010]
Problem: Video object segmentation

How to segment the foreground objects in video when
  • background is moving and changing
  • categories of foreground objects are unknown in advance

Input: Unannotated video

Desired output: Segmentation of high-ranking foreground object

• Existing methods group pixels using low-level features, which can result in an “over-segmentation.” [Brendel & Todorovic 2009, Vazquez-Reina et al. 2010, Grundmann et al. 2010, Brox & Malik 2010]
Mining first-person camera data

GoPro
Google Glass
Looxcie
Tobii
SMI
Pivothead
Mining first-person camera data

90’s
Problem: Summarizing egocentric videos

**Input:** Egocentric video of the camera wearer’s day

**Output:** Storyboard summary of discovered *important people and objects*

Results: Egocentric video summarization
Coming up

• Sign-up for paper

• Next class

• Read both papers

• Write a review for one of them