Advanced Multimedia

Image Retrieval
Tamara Berg
Announcements

Start thinking about projects
   Visit office hours this week or Tues after Spring break to discuss project ideas.
Project Proposals – in class Thurs April 8 (5-10 minutes)
   – Description of problem
   – Challenges you need to solve
   – Timeline for when you will accomplish each part of your project
Previous Projects

• What do images sound like? Translation between images and sound.
• Maps + Music – interactive map for exploring music by artist home town.
• Virtual Dress up room (morphing + shopping images).
• Topic classification of news articles for suggestive browsing.
More Sources for Project Ideas

- Flickr mashups - [http://www.programmableweb.com/api/flickr/mashups](http://www.programmableweb.com/api/flickr/mashups)
- Music + images – visual browser for music collections
- Lots of API’s available for accessing data – flickr, google, google maps, yahoo! news, etc…
- Merging location info with pictures/music
- Image retrieval using text + image + metadata information
- Create a morph sequence between images of all the students in your class
- Text classification
Image Retrieval by text query

Related searches: apple fruit  apple logo  red apple  apple clipart
Retrieval using text info

- Idea – most images have associated text.
- Analyze words around picture & associated with picture (title, words, links, etc).
- For a query word return pictures based on standard web search on text associated with image.
Retrieval using human info

Peekaboom – you and a random partner take turns “peeking” & “booming”
Luis von Ahn, Ruoran Liu and Manuel Blum

Just leave the content analysis/labeling to people.
Retrieval using image info

Content based image retrieval:

• Analyze visual content of images
  – Extract features
  – Build visual descriptor of each image (query and database images).

• For a query image, match descriptors between query and database images.

• Return closest matches in ranked order by similarity.
  – What similarity measures have we talked about?
Retrieval using text+image info

Tags: banana, monkey banana, monkeys, primate, homeless, cardboard, funny photo, orangutan, hairy, brown, zoo, animal, brown fur ...

Web – billions of web pages almost always containing text and images.

Flickr – over 2 billion user uploaded pictures, 2+ million uploads per day. About half come with some sort of associated text.
It's the perfect party dress. With distinctly feminine details such as a wide sash bow around an empire waist and a deep scoop neck, this linen dress will keep you comfortable and feeling elegant all evening long.

bananarepublic.com

Sources of complimentary information!

Combine information for improved performance at recognition, search, or organization
Retrieval by image query

Query Image
Retrieval by image query
Retrieval by image query

Query Image

Ranked Results – database images ranked by similarity to query
Demo

Photo by: marielito
Demo

Represent the image as a spatial grid of average pixel colors
Convert data base of images to this representation
Represent query image in this representation.
Find images from data base that are similar to query.
Commercial systems

- http://www.like.com/
- http://tineye.com/login
- http://labs.systemone.at/retrievr/
- http://www.polarrose.com/
- http://images.google.com
- http://www.picsearch.com/
State of the art retrieval research

• “PageRank for Product Image Search”, Shumeet Baluja, Yushi Jing.
• “80 Million Tiny Images”, Antonio Torralba, Rob Fergus, William T Freeman.
• “Scene Completion Using Millions of Photographs”, James Hays, Alexei A Efros.
PageRank for Product Image Search

Research Paper By: Shumeet Baluja, Yushi Jing
Motivation

Image search – A Graph Theory problem?

A potential search result will have features which are common in majority of results

✓ Identifying “authority” nodes on an inferred visual-similarity graph
✓ Analyzing the visual link structure
PageRank – A link analysis algorithm

- Numeric value that represents how important a page is on the web.
- The more votes that are cast for a page, the more important the page must be.
- The importance of the page that is casting the vote determines how important the vote itself is.

A page's PageRank = 0.15 + 0.85 * (a "share" of the PageRank of every page that links to it)
PageRank → ImageRank: Images are linked through visual links based on similarity!
PageRank

• For web pages – use links between two pages as a measure of their similarity.

• For images – use number of matching features between two images as a measure of their similarity.
  – Features – SIFT features (based on histograms of edges in different directions).
  – Two features are considered matching if their SSD distance is below a threshold.
Similarity defined!

**Similarity between two images**

“The number of interest points shared between two images divided by their average number of interest points”

(a) A v.s. B  
(b) A v.s. C  
(c) A v.s. D  

(d) B v.s. C  
(e) B v.s. D  
(f) C v.s. D

Since all the variations (B, C, D) are based on the original painting (A), A contains more matched local features than others.
ImageRank
Putting it together

- Image u has a visual-hyperlink to image v, then there is some probability that the user will jump from u to v.
- A relevant query will have many other images pointing to it, and will therefore be visited often.
- Images visited often are deemed important.
- If image, v, is important and it links to image w, it is casting its vote for w's importance.
- Because v is itself important, the vote should count more than a "non-important" vote.
Full Retrieval System – Where do we win?

• Queries with homogeneous visual concepts
  • System produces good results by identifying the vertices that are located at the “center” of weighted similarity graph.
  • Example: Monalisa.

• Queries with heterogeneous visual concepts
  • Approach is able to identify a relevant and diverse set of clusters there is no bias – how? Example: Jaguar, Apple
  • Eigen vector centrality measure pays attention to ‘global’ structure of network/graph and ignores local patterns – that's how!

• Query Dependent Ranking
  • Shall we generate the similarity graph S for the billions of images on web? NO!
  • OR Precluster web images based using metadata such as text, anchor text.
  • OR Use existing search engines to get initial result set.
Application & Failures

- Unlike ranking, the goal is not to reorder the full set of images, but to select only the “best” ones to show.
- Examples (Precise and small set (1-3) of images needed)
  - Google Product Search (Single Image).
  - Mixed-Result-Type Search (Text+Image)
- Failures
  - Inflated logo score.
- Screenshot Images (Logos of operating Systems/browser panels).
Questions?
80 Million Tiny Images
Antonio Torralba, Rob Fergus, William T Freeman

53,464 nouns, 79 million images
Very low resolution (32x32)

Many thanks to Rob Fergus for slides!
How many images is enough to solve retrieval?

- Number of possible images is huge
- Natural images make up only a small percentage of all possible
- If we have enough natural images does the problem become easy?
  - If so, how many do we need?
Overview

- Use simple algorithms: nearest neighbors
Thumbnail Collection Project

- Collect images for ALL objects
  - List obtained from WordNet
  - 75,378 non-abstract nouns in English

- Example first 20:

  a-bomb
  a-horizon
  a._conan_doyle
  a._e._burnside
  a._e._housman
  a._e._kennelly
  a.e.
  a_battery
  a_cappella_singing
  a_horizon

  a_kempis
  aalborg
  aalii
  aalborg
  aalost
  aalto
  aar
  aardvark
  aardwolf
  aare
  aare_river
Thumbnail Collection

- 7 different search engines
Dataset Statistics

• Overall stats
  – 79,302,017 images
  – 75,062 different words

• Details
  – Two formats: square & rectangular
  – Gathered at 4.5 images/second
  – Downloaded 97,245,098 images
  – 18% duplicate rate
  – Disk usage: ~ 700Gb
  – Collection time: ~ 9 months
Histogram Images/Word

Total, unique, non-uniform images: 79,302,017

Total number of words: 75,062

Mean # images per word: 1,056
Labeling Noise

- Manual labeling of 78 classes
- Best: Google & Altavista
- Worst: Cydral & Webshots
Suitable Image Representation

• Want minimal representation for task:
  – Classifying scene and dominant objects

• Compact representation has low storage requirements

• We blur & subsample to give low-res image (32x32 color)
Why such tiny images?

- Small
  - Easy to store in a reasonable amount of space
  - Can process lots of them in a short amount of time

- Humans can do recognition well at small scale
Human Labeling of Tiny Scenes

32x32
Human Labeling of Tiny Scenes

32x32

office

waiting area

dining room

dining room

Context!

256x256
Human Performance at Scene Recognition

The role of context in object recognition
A. Oliva, A. Torralba
Trends in Cognitive Sciences, in press.
December 2007.
Non-parametric Classifier

- Nearest-neighbors

- For each query, obtain sibling set (neighbors)

- 3 different types of distance metric

- Hand-designed, use whole image
Metric 1 - $D_{ssd}$

- Sum of squared differences (SSD)

$$D_{ssd}^2 = \sum_{x,y,c}$$

To give invariance to illumination:
Each image normalized to be zero mean, unit variance
Metric 2 - $D_{warp}$

- SSD but allow small transformations

$$D^2_{warp} = \min_{\theta} \sum_{x,y,c}$$

Find min using gradient descent

Target  SSD  Warping
Metric 3 - $D_{shift}$

- As per Warping but also allow sub-window shifts

$$D_{shift}^2 = \min_{\text{Local sub-window}} \sum_{x,y,c}$$

- Quick since images are so small
Sibling Sets with Different Metrics

- Sibling set is 50 images
Quality of Sibling Set using $D_{shift}$

Size of dataset

- Target
- 7,900
- 722,000
- 79,600,000
- $10^5$
- $10^6$
- $10^8$
Label Assignment

- Distance metrics give set of nearby images
- How to compute label?

Issues:
- Labeling noise
- Keywords can be very specific
  - e.g. yellowfin tuna
Wordnet – a Lexical Dictionary

http://wordnet.princeton.edu/

Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun aardvark

Sense 1
aardvark, ant bear, anteater, Orycteropus afer
  => placental, placental mammal, eutherian, eutherian mammal
  => mammal
  => vertebrate, craniate
    => chordate
    => animal, animate being, beast, brute, creature
      => organism, being
      => living thing, animate thing
        => object, physical object
        => entity

Each noun corresponds to a path in Wordnet tree
Wordnet Voting Scheme

1. a) Input image
2. b) Neighbors
3. Ground truth
4. d) Wordnet voted branches

One image - one vote
Vote for branches in Wordnet tree
Wordnet Voting Scheme

a) Input image
b) Neighbors
c) Ground truth
d) Wordnet voted branches
Wordnet Voting

- Overcomes differences in level of semantic labeling:
  - e.g. “person” & “sir arthur conan doyle”

- Totally incorrect labels form hopefully uniform background noise

- Assumes semantic and visual consistency are closely related
Recognition Experiments
Person Recognition

- 23% of all images in dataset contain people

- Wide range of poses: not just frontal faces
Person Recognition – Test Set

- 1016 images from Altavista using “person” query
- High res and 32x32 available
- Disjoint from 79 million tiny images
Re-ranked Altavista Images

Original

Re-ranked
Person Recognition

- **Task**: person in image or not?

![Graph showing Precision vs. Recall for different image ranking methods: Tiny images ranking, Viola-Jones (high-res), Viola-Jones (32x32), Altavista ranking.]

**Important concepts:**

As you walk down the ranked images,

Precision – of the things in the ranking so far what percentage are correct.

Recall – of all the correct items what percentage of them did you return so far.
Count the votes at the corresponding node of Wordnet tree for classification.
Extrapolation of how well it would do at Google scale dataset
Object Classification

Performance drops as classes become more specific
Automatic Colorization

<table>
<thead>
<tr>
<th>Grayscale input High resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grayscale 32x32 siblings</td>
</tr>
<tr>
<td>Color siblings high resolution</td>
</tr>
<tr>
<td>Average of color siblings</td>
</tr>
<tr>
<td>Colorization of input using average</td>
</tr>
<tr>
<td>Colorization of input using specific siblings</td>
</tr>
</tbody>
</table>
Automatic Colorization Result

Grayscale input High resolution

Colorization of input using average
Automatic Orientation

• Look at mean distance to neighbors

Subspace of natural images

Images at wrong orientation have neighbors further away
Automatic Orientation Examples

Average correlation to 50 closest neighbors
Automatic Orientation

- Many images have ambiguous orientation
- Look at top 25% by confidence:
- Examples of high and low confidence images:
Conclusions

• Can get good results simple algorithms & lots of data

• Issues with Internet images: labeling noise & image baises.

• Bring in learning: Distance metrics, text & images

Webpage: http://people.csail.mit.edu/torralba/tinyimages

Thanks to Rob Fergus for slides!