Image Retrieval

Tamara Berg
Advanced Multimedia
Reminders

• Last HW out, due March 22
  – Translating between images and sound!
Projects

Start thinking about project ideas

Anything with at least 1 form of multimedia is fair game
  Must involve some coding (needn’t be in matlab)
  Visit office hours to discuss

Project Proposals – March 27
Previous Projects

• Maps + Music – interactive map for exploring music by artist home town.
• Virtual Dress up room (morphing shopping images onto virtual models).
• Topic classification of news articles for user suggested browsing.
• Bunny hop video game – driven by web cam.
• Kinect virtual drum kit
More Sources for Project Ideas

• Google maps mashups - [http://googlemapsmania.blogspot.com/](http://googlemapsmania.blogspot.com/)
• 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
• Kinect game

If you tell me what you want to do I can suggest specific resources
Image Retrieval by text query
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 + user data

Watch what people click on!
Retrieval using text+image info

Tags: banana, monkey, bananas, 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.
Retrieval using text+image info

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

Query Image

Database Images
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.
Quick overview of other common kinds of Features
Reminder: Edges

source: Svetlana Lazebnik
Edge filters

Approximations of derivative filters:

Prewitt: \( M_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \); \( M_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \)

Sobel: \( M_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \); \( M_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \)

Roberts: \( M_x = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \); \( M_y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \)

Convolve filter with image to get edge map

Source: K. Grauman
Important concept: Histograms

Graphical display of tabulated frequencies, shown as bars. It shows what proportion of cases fall into each of several categories. The categories are usually specified as non-overlapping intervals of some variable.
Color Histograms

Representation of the distribution of colors in an image, derived by counting the number of pixels of each of given set of color ranges in a typically (3D) color space (RGB, HSV etc).

What are the bins in this histogram?
Shape Descriptors: shape context

Representation of the local shape around a feature location (star) – as histogram of edge points in an image relative to that location. Computed by counting the edge points in a log polar space.

So what are the bins of this histogram?
Shape descriptors: SIFT

• Descriptor computation:
  • Divide patch into 4x4 sub-patches
  • Compute histogram of gradient orientations (convolve with filters that respond to edges in different directions) inside each subpatch
  • Resulting descriptor: 4x4x8 = 128 dimensions


source: Svetlana Lazebnik
Texture Features

Convolve with various filters – spot, oriented bar.

Compute histogram of responses.

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.

• “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

1. Image $u$ has a visual-hyperlink to image $v$, then there is some probability that the user will jump from $u$ to $v$.

2. A relevant query will have many other images pointing to it, and will therefore be visited often.

3. Images visited often are deemed important.

4. If image, $v$, is important and it links to image $w$, it is casting its vote for $w$'s importance.

5. 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?
Scene Completion Using Millions of Photographs

James Hays and Alexei A. Efros
Carnegie Mellon University

Thanks to James & Alyosha for slides!

Fill in unknown region from source image parts.
Efros and Leung result – no notion of semantics, also assumes necessary data is present elsewhere in the image.
Scene Matching for Image Completion
Challenges:

- Computational costs of searching lots of images
- Should fill in missing regions with semantically valid fragments

Scene Completion Result
The Algorithm

Input image

Scene Descriptor

Image Collection

20 completions

Context matching + blending

200 matches

Hays and Efros, SIGGRAPH 2007
Data

We downloaded **2.3 Million** unique images from Flickr groups and keyword searches.

Groups: lonelyplanet, urban-fragments, ruraldecay ...
Keywords: outdoors, vacation, river...
Discard duplicates and small images
Scene Matching
Compute oriented edge response at multiple scales (5 spatial scales, 6 orientations)
Scene Descriptor

Gist scene descriptor (Oliva and Torralba 2001)

“semantic” descriptor of image composition
aggregated edge responses over 4x4 windows
scenes tend to be semantically similar under this descriptor if very close

Hays and Efros, SIGGRAPH 2007
Scene Descriptor

Gist scene descriptor - with missing regions masked (weighted based on percentage of valid pixels)

Hays and Efros, SIGGRAPH 2007
Scene Descriptor

Color descriptor – color of the query image downsampled to 4x4

Distances calculated by SSD between query image descriptors & imgs in database

Total Dist = color dist + 2*gist dist

Gist scene descriptor
(Oliva and Torralba 2001)
Hays and Efros, SIGGRAPH 2007
200 total
Context Matching

Need to more precisely align matching scenes to local img context around missing region
local context = all pixels within 80 pixel radius of hole’s boundary
Compute pixel-wise error of 200 best scene matches across all valid translations and 3 scales
Compute texture similarity of proposed fill-in to removed region

Hays and Efros, SIGGRAPH 2007
Final result – blended between the two images along the cut to merge seamlessly
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
Pro – allows insertion of novel objects
The Algorithm

Input image
The Algorithm

Compute a global description of the whole image

Hays and Efros, SIGGRAPH 2007
The Algorithm

Input image  →  Scene Descriptor  →  Image Collection

Compare to LOTS of images

Hays and Efros, SIGGRAPH 2007
The Algorithm

Input image → Scene Descriptor → Image Collection

Get top matches

200 matches

Hays and Efros, SIGGRAPH 2007
The Algorithm

Input image

Scene Descriptor

Image Collection

Compare more locally and merge pieces of matching images

Context matching + blending

200 matches

Hays and Efros, SIGGRAPH 2007
The Algorithm

Input image → Scene Descriptor → Image Collection

20 completions (final results) → Context matching + blending → 200 matches

Hays and Efros, SIGGRAPH 2007
... 200 scene matches

Hays and Efros, SIGGRAPH 2007
... 200 scene matches

Hays and Efros, SIGGRAPH 2007
... 200 scene matches

Hays and Efros, SIGGRAPH 2007
... 200 scene matches
... 200 scene matches
... 200 scene matches
... 200 scene matches
Failures
Failures
Failures

Cause of failure – atypical scene caused lack of good matches

Hays and Efros, SIGGRAPH 2007
Failures
Failures
Failures
Failures
Failures

Cause of failure – fine scale texture mismatch

Hays and Efros, SIGGRAPH 2007
Failures
Failures
Failures
Failures

Cause of failure – no notion of “objects”
Evaluation
Real Image. This image has not been manipulated

or

Fake Image. This image has been manipulated
User Study Results - 20 Participants

Criminisi et al.

Our algorithm

Real Photographs

Percentage of images marked fake

Maximum response time (seconds)
Why does it work?
10 nearest neighbors from a collection of 20,000 images

Hays and Efros, SIGGRAPH 2007
10 nearest neighbors from a collection of 2 million images

Hays and Efros, SIGGRAPH 2007