

Words and pictures: the problem

D.A. Forsyth, UIUC

with: Kobus Barnard, U.Arizona; Pinar Duygulu, Bilkent U.;
Nando de Freitas, UBC; Tamara Berg, UIUC; Derek Hoiem,
UIUC; Ian Endres, UIUC; Ali Farhadi, UIUC; Gang Wang,
UIUC;

Part 1: General points

- Image organization is important, with many applications
- Matching to image queries isn't good enough
- Browsing is important
- Iconographic effects force us to rerank results
- Words and pictures are curiously complementary

! Ten year old slide !

LOTS of BIG collections of images

Corel Image Data	40,000 images
Fine Arts Museum of San Francisco	83,000 images online
Cal-flora	20,000 images, species information
News photos with captions (yahoo.com)	1,500 images per day available from yahoo.com
Hulton Archive	40,000,000 images (only 230,000 online)
internet.archive.org	1,000 movies with no copyright
TV news archives (televisionarchive.org, informedia.cs.cmu.edu)	Several terabytes already available
Google Image Crawl	>330,000,000 images (with nearby text)
Satellite images (terrarserver.com, nasa.gov, usgs.gov)	(And associated demographic information)
Medial images	(And associated with clinical information)

1e9 or so

* and the BBC is releasing its video archive, too;
and we collected 500,000 captioned news images;
and it's easy to get scanned mediaeval manuscripts;
etc., etc.,

More BIG collections

- Object recognition datasets
 - 80 million images (Russell 08)
 - Caltech 5, 101, 256 (Fei-Fei 04; Griffin 07)
 - PASCAL (Everingham 06)
- Flickr
 - vast quantities of pix
- YouTube
 - vast quantities of video

How many images on the web?

- Google image search
 - 1.4×10^9 ? (search google on .jpg)
 - 9.14×10^8 ? (search google images on .jpg, no filtering)
 - 9.57×10^8 ? (ditto, strong filtering)
 - 7.3×10^8 (the, no filtering)
 - 6.9×10^8 (the, filtering)
- Bing image search
 - 2×10^9 (.jpg, no filtering)
 - 1.9×10^9 (.jpg, strict filtering)
 - 2.3×10^6 (the, no filtering)
 - 2.2×10^6 (the, strict filtering)
- Conclusion:
 - about 1×10^9 (?), with about 10% rejected by filters

Imposing order

- Iconic matching
 - child abuse prosecution
 - managing copyright (BayTSP)
- Tagging and sharing
 - Flickr
 - Youtube

Current, profitable applications

Imposing order

- Browsing for:
 - web presence for museums (Barnard et al, 01)
 - home picture, video collections
 - selling pictures
- Searching
 - home picture, video collections
 - selling pictures
 - scanned writing (Manmatha, 02)
 - collections of insects, flowers, etc.
 - movie search; celebrity search
 - product search
- Building world knowledge
 - a face gazetteer (Miller et al, 04)



Maybe applications



Maybe applications

Consumer Products



Marc by Marc Jacobs
Adorable peep-toe pumps, great for any occasion. Available in an array of uppers. Metallic fabric trim and bow detail. Metallic leather lined footbed. Lined printed design. Leather sole. 3 3/4" heel.

Zappos.com



soft and glassy patent calfskin trimmed with natural vachetta cowhide, open top satchel for daytime and weekends, interior double slide pockets and zip pocket, seersucker stripe cotton twill lining, kate spade leather license plate logo, imported
2.8" drop length
14"h x 14.2"w x 6.9"d

Katespade.com



It's the perfect party dress. With distinctly feminine details such as a wide sash bow around an empire waist and a deep scoopneck, this linen dress will keep you comfortable and feeling elegant all evening long.

* Measures 38" from center back, hits at the knee.

* Scoopneck, full skirt.

* Hidden side zip, fully lined.

* 100% Linen. Dry clean.

bananarepublic.com

E-commerce transactions in 2004, 2005, 2006 of \$145 billion, \$168 billion, and \$198 billion (Forrester Research).

Part 1: General points

- Image organization is important, with many applications
- Matching to image queries isn't good enough
- Browsing is important
- Iconographic effects force us to rerank results
- Words and pictures are curiously complementary

Searching

- Search involves
 - Specifying a need (e.g. example image, query word)
 - Get it met (e.g. by matching, ranking)
- Image example queries
 - global features
 - (e.g. Flickner et al. 95, Carson et al. 99, Wang 00, various entire conferences)
 - local features
 - (e.g. Photobook - Pentland et al 96; Blobworld - Carson et al, 98)
 - relevance feedback
 - (e.g. Cox et al 00; Santini 00; Schettini 02; etc.)
- Special case
 - naughty pictures
 - (eg Forsyth et al. 96, 99; Wang et al. 98; Chan et al 99)

Picture Queries

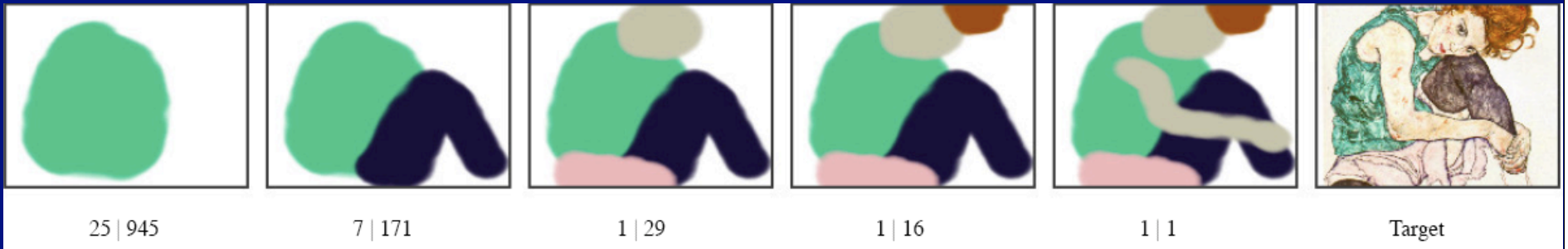


Jacobs et al, 1995

Procedure

- For each image in collection
 - compute a signature
- For query
 - compute match score to each image
- Signatures
 - Many possibilities
 - Color histograms (not enough spatial info)
 - Wavelet coefficients

Query



Jacobs et al, 1995

Image retrieval

- Whole image queries don't work that well
 - users don't seem to like them
 - clumsy interface?
 - hard to find what you want?
 - errors confusing?

What will users pay for?

- Work by Peter Enser and colleagues on the use of photo movie collections
(Enser McGregor 92; Ornager 96; Armitage Enser 97; Markkula Sormunen 00; Frost et al 00; Enser 00)
- Typical queries:

What is this about?

“... smoking of kippers...”

“The depiction of vanity in painting, the depiction of the female figure looking in the mirror, etc.”

“Cheetahs running on a greyhound course in Haringey in 1932”

What do we need to get right?

- Recall
 - percentage of relevant items recovered
- Precision
 - percentage of items recovered that are relevant
- Sometimes high recall is important
 - eg search for prior art
- Sometimes high precision is important
 - eg picture of the president for a magazine article

Part 1: General points

- Image organization is important, with many applications
- Matching to image queries isn't good enough
- **Browsing is important**
- Iconographic effects force us to rerank results
- Words and pictures are curiously complementary

Browsing

- Searching big, unknown collections is hard for naive user
 - skilled users don't benefit from vision-based tools
 - image-based search is of overrated significance
- Browsing
 - seems to be preferred by naive users (Frost et al, '00)
 - but browsing requires organization too
 - generally underrated problem

*Notable exceptions ---Sclaroff, Taycher, and La Cascia, 98;
Rubner, Tomasi, and Guibas, 00; Smith Kanade, 97.

Google “rooms”



... virtual tour > room photos
644 x 446 - 39k - jpg
www.mandalaybay.com



Bed Room Sets
599 x 402 - 33k - jpg
www.chiphi-pi.org



16 Creative and Sexy Art Hotel Rooms ...
468 x 354 - 111k - jpg
weburbanist.com
[[More from weburbanist.com](http://weburbanist.com)]



Rooms >
450 x 300 - 25k - jpg
www.radisson.com
[[More from www.radisson.com](http://www.radisson.com)]



Bookcase Secret Room Door
468 x 391 - 98k - jpg
weburbanist.com



The large room known today as the ...
350 x 353 - 48k - jpg
www.royalacademy.org.uk



To reserve a room call 212-596-1200 ...
640 x 480 - 93k - jpg
www.columbiaclub.org



Now let's see some amazing rooms.
450 x 300 - 19k - jpg
freshhome.com



Room for physically-challenged
600 x 395 - 244k - jpg
www.hotelnikkohanoi.com.vn



basement family room
450 x 325 - 48k - jpg
www.thisoldhouse.com



Handicap Room
300 x 301 - 22k - jpg
intl-house.howard-hotels.com



Spacious Guest Room
450 x 300 - 29k - jpg
www.radisson.com



Rooms may also include twin beds and ...
370 x 486 - 40k - jpg
www.inisrael.com



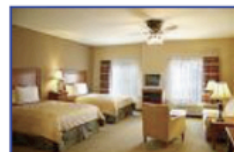
This bright room on the 2nd floor of ...
1728 x 1152 - 283k - jpg
biosphere.ec.gc.ca



These twenty rooms ...
468 x 352 - 97k - jpg
weburbanist.com



Texas' enormous locker room facility ...
530 x 343 - 34k - ipa



Two Queen Room
450 x 300 - 26k - jpg
www.countrvinn.com



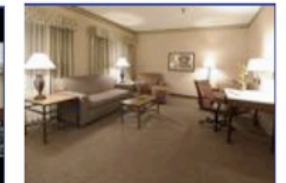
trent room The Trent Room was first ...
346 x 450 - 54k - ipa



Image of changing room
450 x 388 - 75k - ipa



Tour the USC Marshall Capture Room
637 x 481 - 160k - ipa



large drawing room in two room suite
737 x 551 - 70k - ipa

Bing “room”



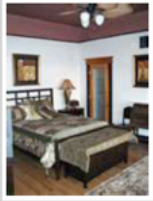

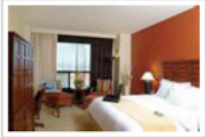

room - Bing Images 6/30/09 10:2

Web **Images** Videos Shopping News Maps More | MSN | Windows Live Sign in | United States | Extras ▼

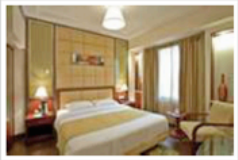
ALL RESULTS IMAGES 1-18 of 125,000,000 results **safe search off** ([change](#))

Images

- SIZE
- LAYOUT
- COLOR
- STYLE
- PEOPLE


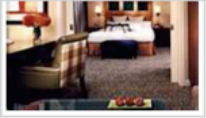






CLUB-ROOM_450.jpg



450 x 300 · 21 kB
www.countryinns.com

[show similar images](#)
[feedback on this image](#)





© 2009 Microsoft | [Privacy](#) | [Legal](#) | [Help](#) | [Feedback](#) | Images are scaled down and may be

Part 1: General points

- Image organization is important, with many applications
- Matching to image queries isn't good enough
- Browsing is important
- **Iconographic effects force us to rerank results**
- Words and pictures are curiously complementary

Iconography

- Images on the web are “selected”
 - visual “modes” of representation
 - if we rank by match quality, we might not see all modes
- Iconographic biases are widespread
 - cf Google image search with Flickr



Loeff et al, 06

Google “rooms”



... virtual tour > room photos
644 x 446 - 39k - jpg
www.mandalaybay.com



Bed Room Sets
599 x 402 - 33k - jpg
www.chiphi-pi.org



16 Creative and Sexy Art Hotel Rooms ...
468 x 354 - 111k - jpg
weburbanist.com
[[More from weburbanist.com](http://weburbanist.com)]



Rooms >
450 x 300 - 25k - jpg
www.radisson.com
[[More from www.radisson.com](http://www.radisson.com)]



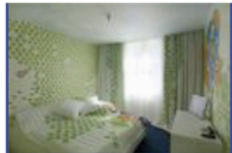
Bookcase Secret Room Door
468 x 391 - 98k - jpg
weburbanist.com



The large room known today as the ...
350 x 353 - 48k - jpg
www.royalacademy.org.uk



To reserve a room call 212-596-1200 ...
640 x 480 - 93k - jpg
www.columbiaclub.org



Now let's see some amazing rooms.
450 x 300 - 19k - jpg
freshhome.com



Room for physically-challenged
600 x 395 - 244k - jpg
www.hotelnikkohanoi.com.vn



basement family room
450 x 325 - 48k - jpg
www.thisoldhouse.com



Handicap Room
300 x 301 - 22k - jpg
intl-house.howard-hotels.com



Spacious Guest Room
450 x 300 - 29k - jpg
www.radisson.com



Rooms may also include twin beds and ...
370 x 486 - 40k - jpg
www.inisrael.com



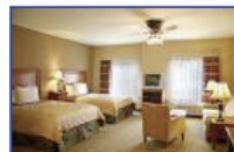
This bright room on the 2nd floor of ...
1728 x 1152 - 283k - jpg
biosphere.ec.gc.ca



These twenty rooms ...
468 x 352 - 97k - jpg
weburbanist.com



Texas' enormous locker room facility ...
530 x 343 - 34k - ipa



Two Queen Room
450 x 300 - 26k - jpg
www.countrvinn.com



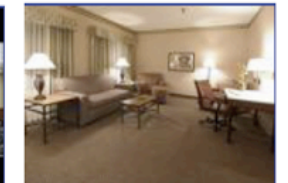
trent room The Trent Room was first ...
346 x 450 - 54k - ipa



Image of changing room
450 x 388 - 75k - ipa



Tour the USC Marshall Capture Room
637 x 481 - 160k - ipa



large drawing room in two room suite
737 x 551 - 70k - ipa

Flickr "rooms"

View: Most relevant • Most recent • Most interest

he love
by fsumar

New living room
by flowers & machinery
29 comments 13 notes
Tagged with art, home, vintage
Taken on December 3, 2007
December 4, 2007
See more of flowers & machinery's photos, or visit her profile.

South Side of Madison
(Fisheye) by Joeysplan
6 comments 9 favorite notes
Tagged with pets, cat, close-up
Taken on January 16, 2008
January 16, 2008
See more of Joeysplan's photos, or visit his profile.

[the waiting Room]
by bass_nroll
55 comments 84 notes
Tagged with contrast, canopy, sleepingbags ...
Taken on April 16, 2007, updated on April 16, 2007
Taken in Madonna di Campiglio, Italy (map)
See more of bass_nroll's photos, or visit his profile.

Live in rooms for light. by *Peanut (Lambert)
147 comments 13 notes
Tagged with nikkor50mmf1.8, argbacktoworktoday, sameer
Taken on January 3, 2009, updated on January 5, 2009

Reranking

- Idea:
 - Don't repeat visual "modes"
 - Display some of each, ideally "important" ones
- Motivation:
 - in image search, precision matters, recall doesn't
 - (usually?)

Reranking

- Build graph linking search results
 - links are local appearance similarity measures
 - (eg matching interest points)
 - identify important images
- graph is weighted
 - by the number of interest points that match
 -

The largest number of neighbours is not a good “importance” test because it tends to find large clusters of very similar images and ignore large scale structure



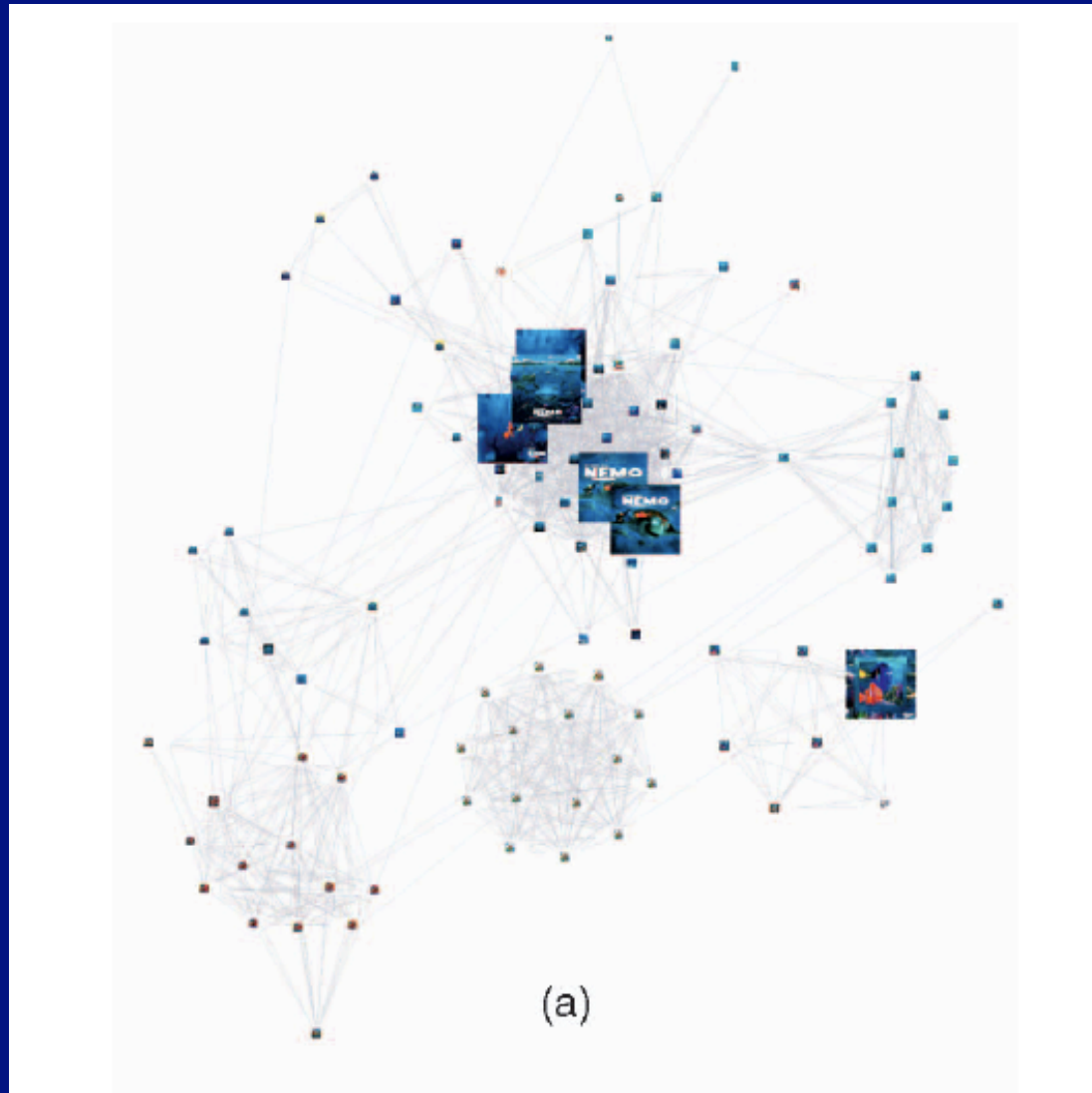
Ying Baluja 08

Reranking

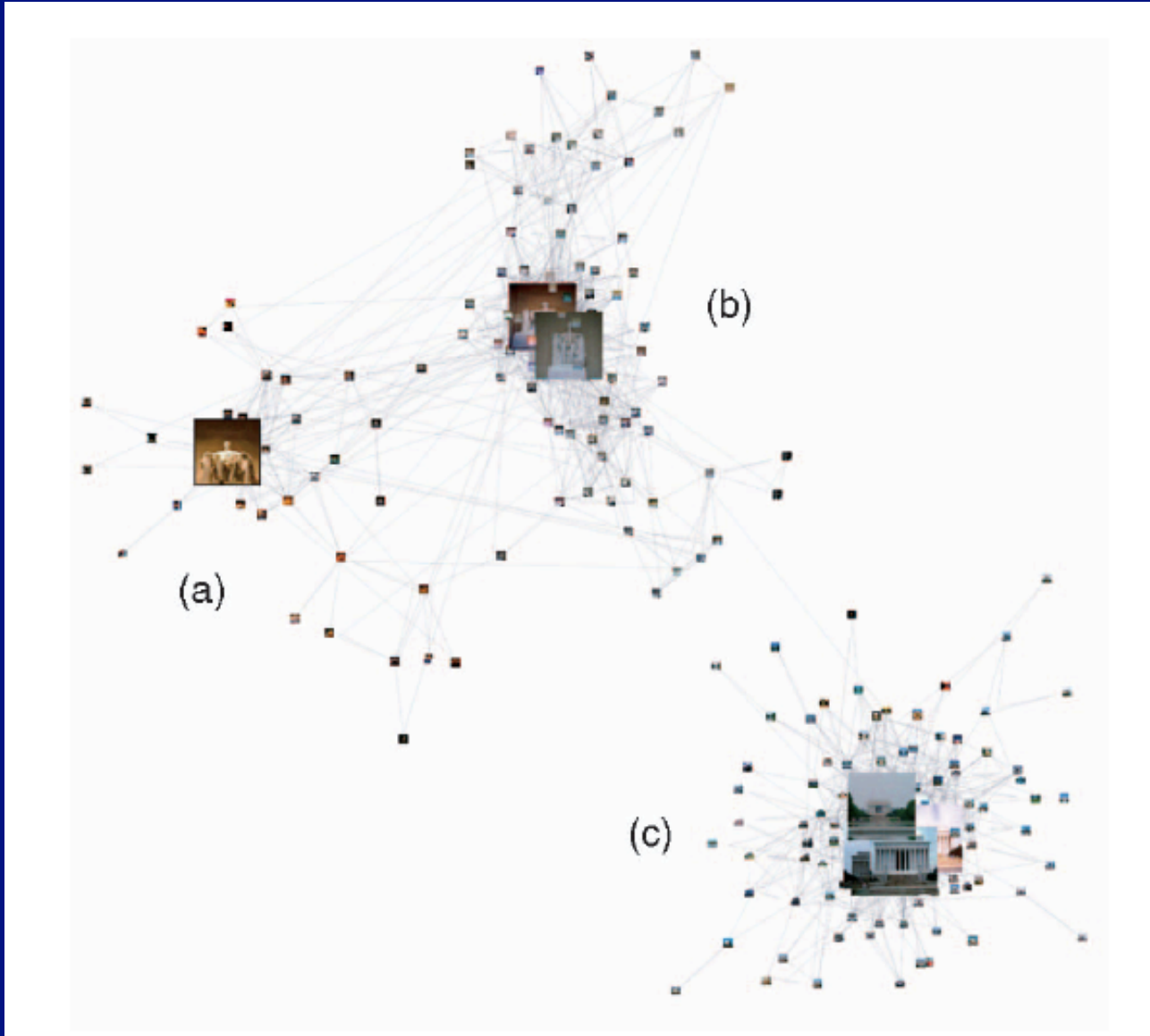
- Model:
 - random walk on weighted graph
 - high probability of arriving at an image implies it's important
 - i.e. degree counts, but so do weights on links and degree of neighbours
- importance is frequency with which one visits points

Reranking

- Big graphs
 - (can't build the graph)
- Approach
 - use a hash table to build the links
 - simulate the random walk
 - trick: at any node, with small probability transition to any other
 - otherwise, follow weights
 - trick: exploit the hash table
 - for transition, choose feature UAR, then choose collisions in hash bucket



Ying Baluja 08



Ying Baluja 08

Part 1: General points

- Image organization is important, with many applications
- Matching to image queries isn't good enough
- Browsing is important
- Iconographic effects force us to rerank results
- **Words and pictures are curiously complementary**

Image retrieval - query on segments

Query on



Example from Berkeley
Blobworld system



Carson et al 02

Query on
“Rose”

Example from Berkeley
Blobworld system



Annotation results in complementary words and pictures

Query on
“Rose”
and



Example from Berkeley
Blobworld system



Carson et al 02

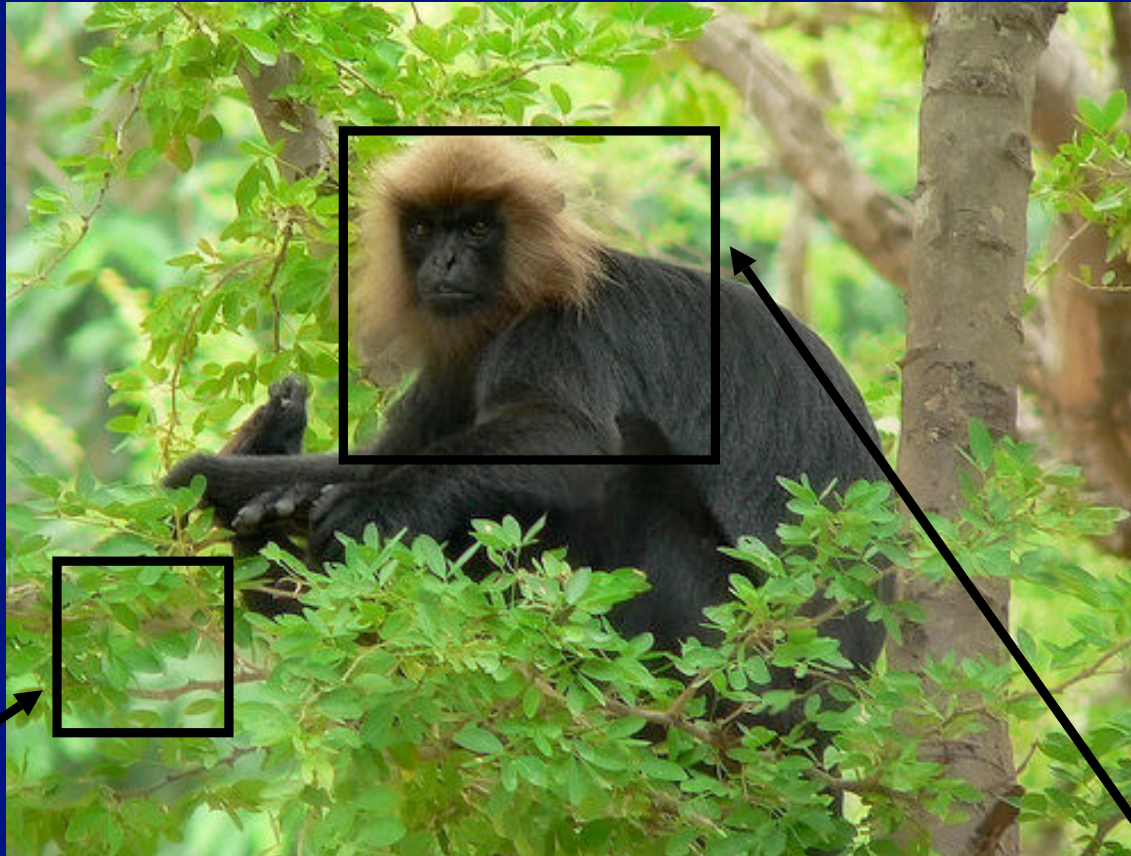
Complementary relations



Tags:

leaves, endangered, green, i love nature, chennai, nilgiri langur, monkey, forest, wildlife, perch, black, wallpaper, ARK OF WILDLIFE, topv111, WeeklySurvivor, top20HallFame, topv333, 100v10f, captive, simian

What do words+picture tell us?



Tags:

leaves, endangered, green, i love nature, chennai, nilgiri langur, monkey, forest, wildlife, perch, black, wallpaper, ARK OF WILDLIFE, topv111, WeeklySurvivor, top20HallFame, topv333, 100v10f, captive, simian

Searching with words

- Most pictures don't have words "attached"
- Attach in simple ways
 - in image name
 - in caption
 - in nearby text
- All really useful, but dangerous
 - 12739.jpg?
 - common to have pictures on web pages without easily identified captions
 - nearby text might, might not, be relevant
- Idea: Predict annotations from picture