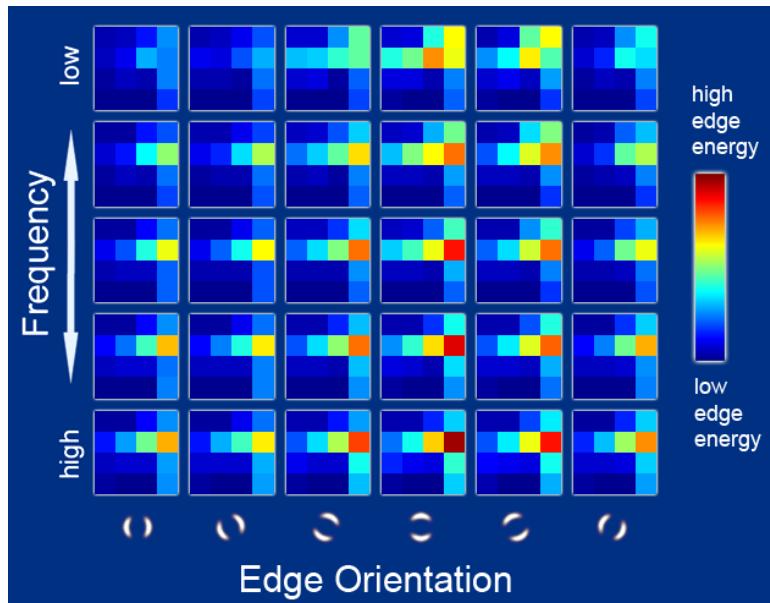


# History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

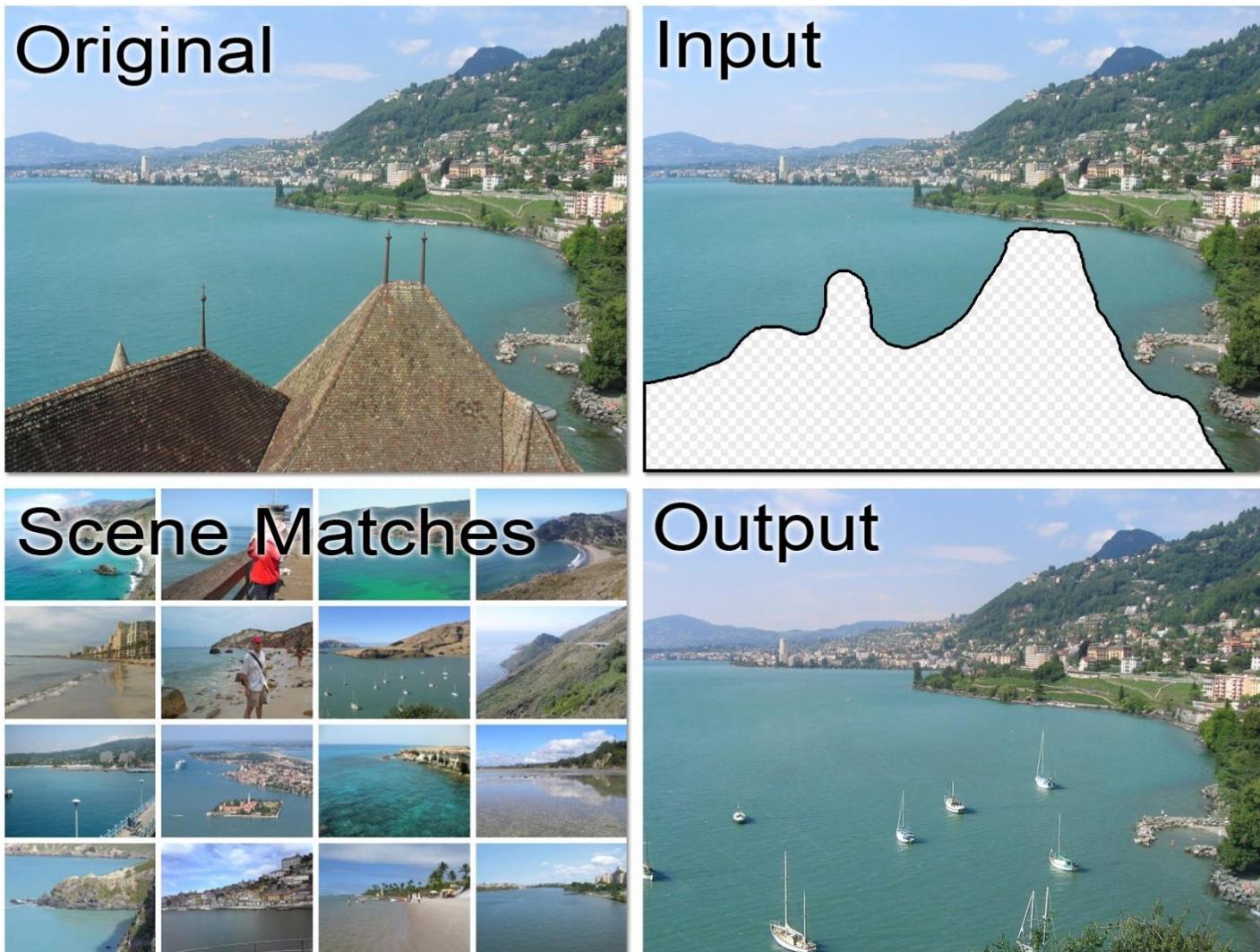
# Global scene descriptors

- The “gist” of a scene: Oliva & Torralba (2001)

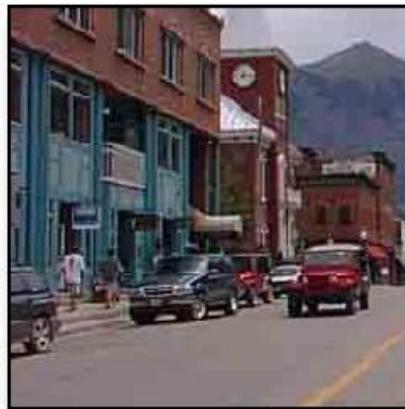


<http://people.csail.mit.edu/torralba/code/spatialevelope/>

# Data-driven methods



# Data-driven methods



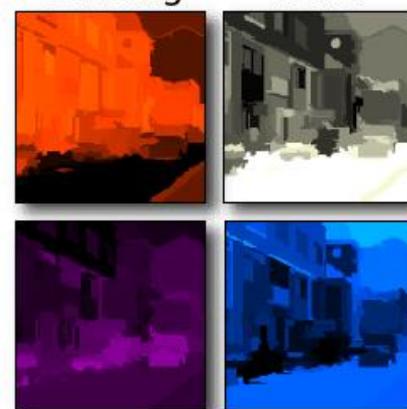
(a) Query Image



(b) Retrieval Set



(c) Superpixels

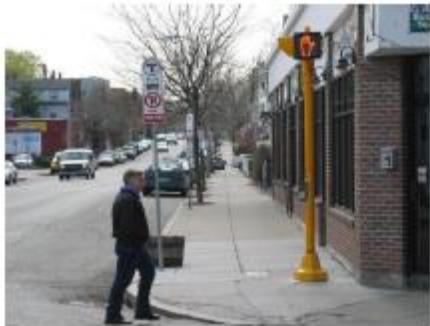


(d) Per-class Likelihoods



(e) Final Labeling

# Geometric context



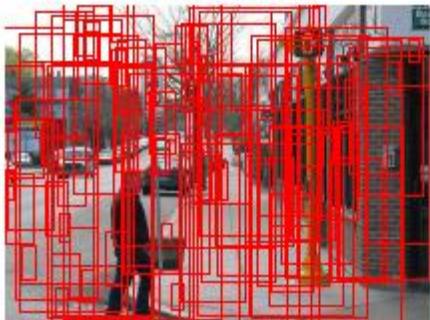
(a) Input image



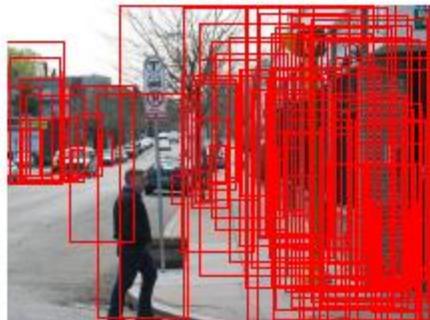
(c) Surface estimate



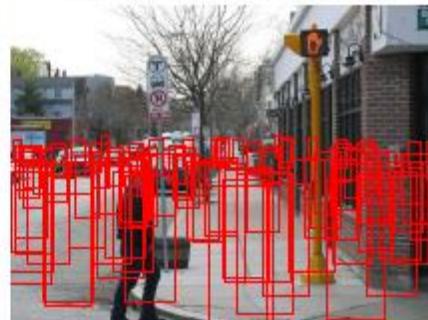
(e)  $P(\text{viewpoint} \mid \text{objects})$



(b)  $P(\text{person}) = \text{uniform}$



(d)  $P(\text{person} \mid \text{geometry})$



(f)  $P(\text{person} \mid \text{viewpoint})$



(g)  $P(\text{person} \mid \text{viewpoint, geometry})$

D. Hoiem, A. Efros, and M. Herbert. [Putting Objects in Perspective](#). CVPR 2006.

# What Matters in Recognition?

- Learning Techniques
  - E.g. choice of classifier or inference method
- Representation
  - Low level: SIFT, HoG, gist, edges
  - Mid level: Bag of words, sliding window, deformable model
  - High level: Contextual dependence
- Data
  - More is always better
  - Annotation is the hard part

# What Matters in Scene Recognition?

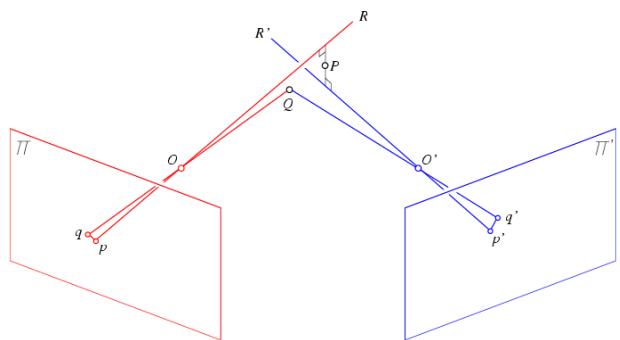
- Learning Techniques
  - ?
- Representation
  - ?
- Data
  - ?

# Large-scale Instance Retrieval

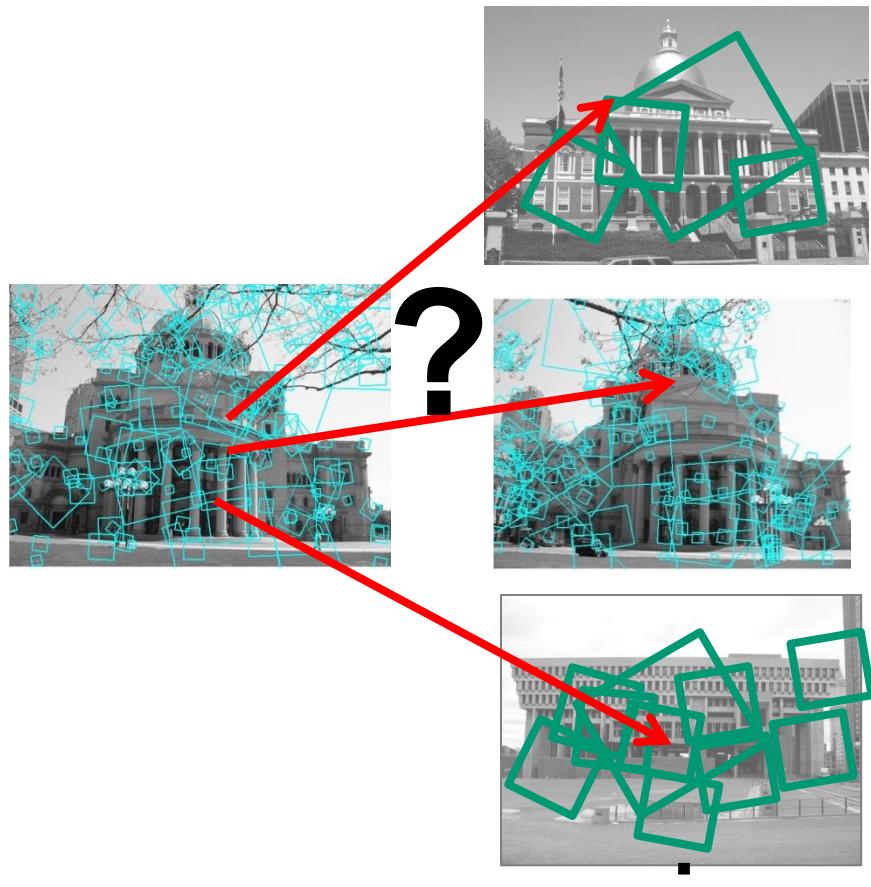
Computer Vision  
CS 143, Brown

James Hays

# Multi-view matching



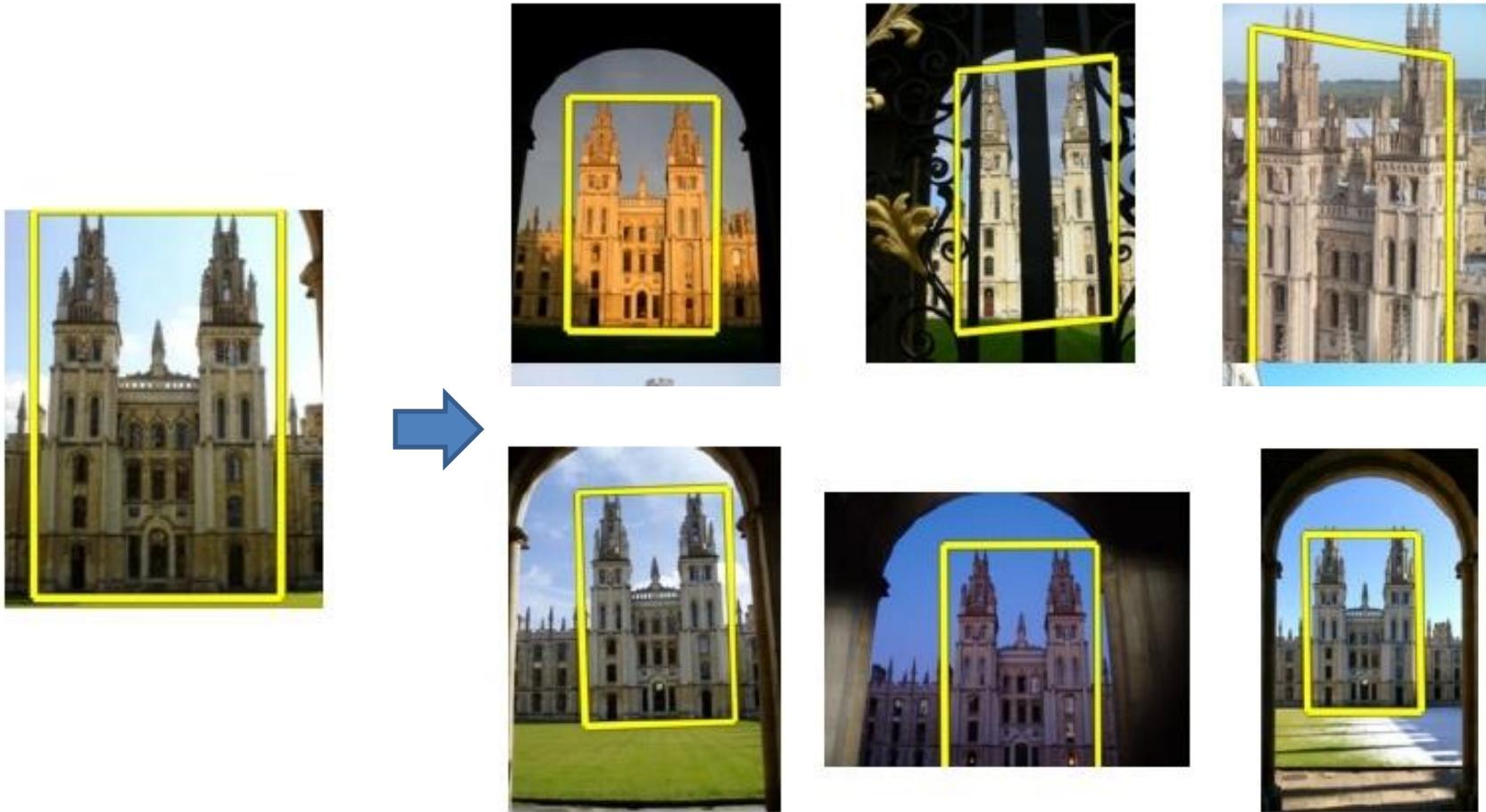
vs



Matching two given  
views for depth

Search for a matching  
view for recognition

# How to quickly find images in a large database that match a given image region?



# Video Google System

1. Collect all words within query region
2. Inverted file index to find relevant frames
3. Compare word counts
4. Spatial verification

Sivic & Zisserman, ICCV 2003

- Demo online at :  
<http://www.robots.ox.ac.uk/~vgg/research/vgoogle/index.html>



Query region



Kristen Grauman

Retrieved frames

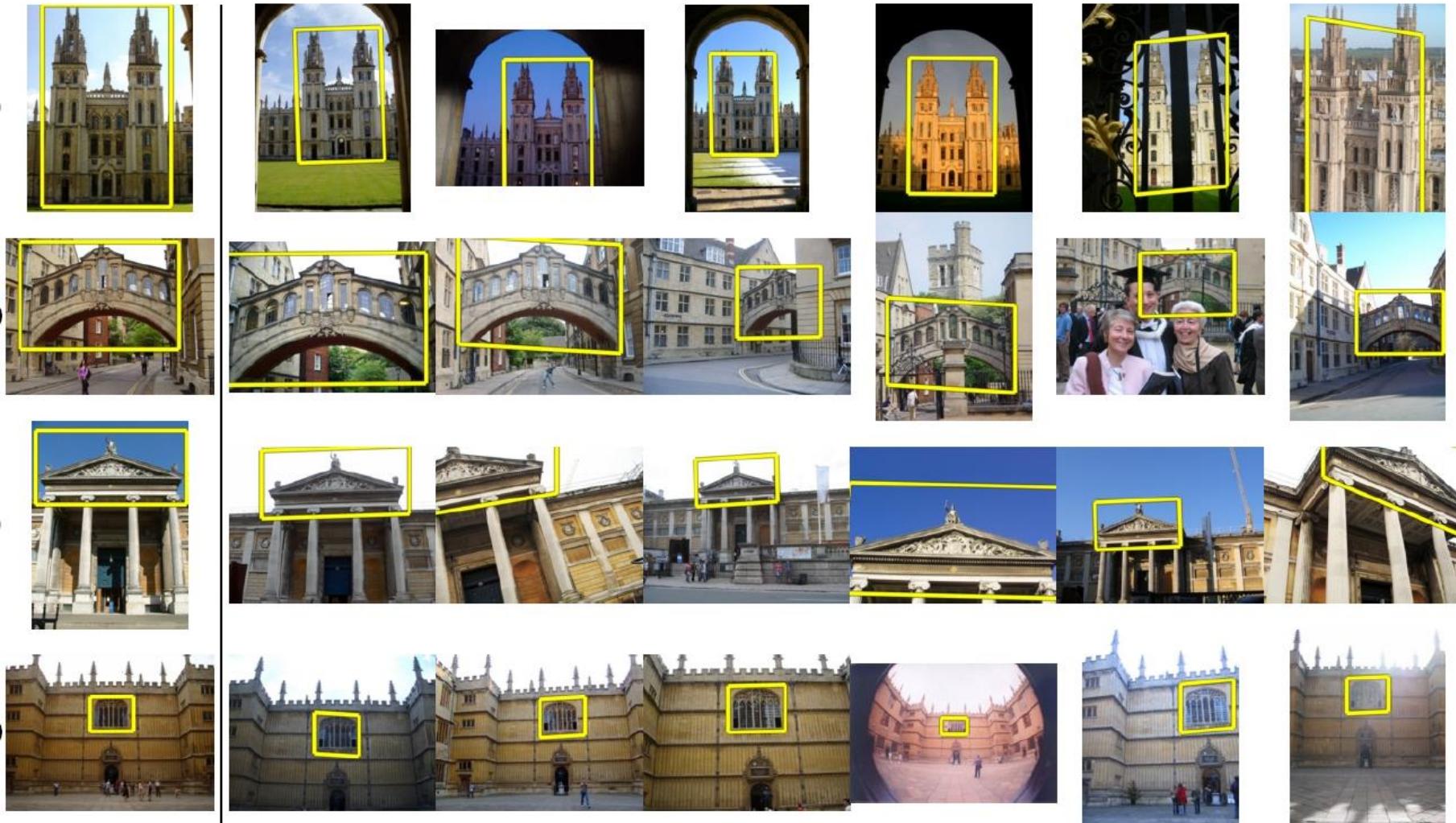
# Example Applications



## Mobile tourist guide

- Self-localization
- Object/building recognition
- Photo/video augmentation

# Application: Large-Scale Retrieval



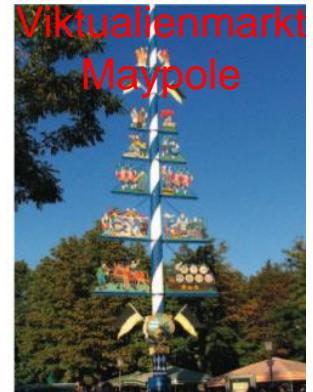
Query

Results from 5k Flickr images (demo available for 100k set)

# Application: Image Auto-Annotation



Left: Wikipedia image  
Right: closest match from Flickr





# Google Goggles

Use pictures to search the web.

Watch a video



## Get Google Goggles

**Android (1.6+ required)**

Download from [Android Market](#).

[Send Goggles to Android phone](#)

**New! iPhone (iOS 4.0 required)**

Download [from the App Store](#).

[Send Goggles to iPhone](#)

New!



[Text](#)



[Landmarks](#)



[Books](#)



[Contact Info](#)



[Artwork](#)



[Wine](#)



[Logos](#)



Ihr A

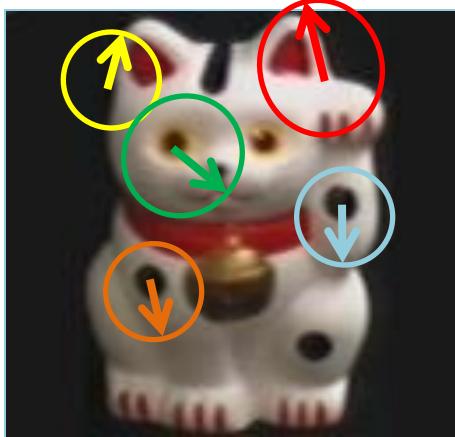
Fruh

auf Topin

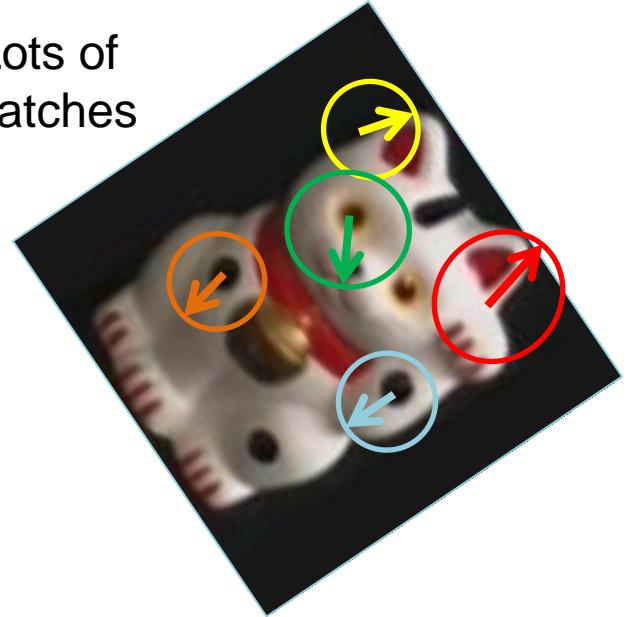


# Simple idea

See how many keypoints  
are close to keypoints in  
each other image



Lots of  
Matches



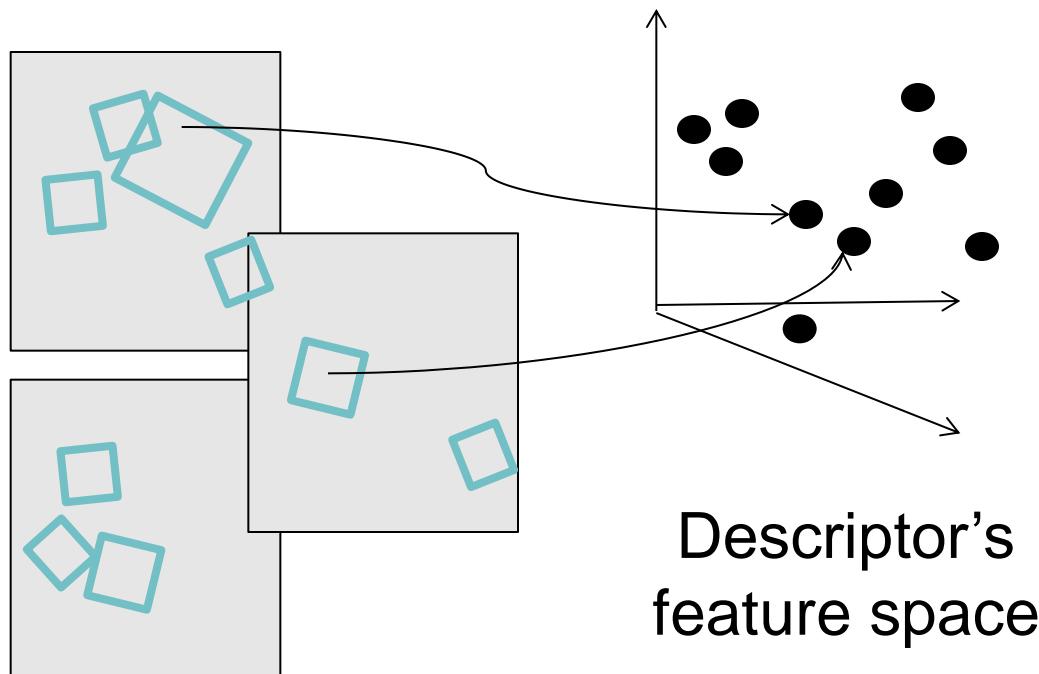
Few or No  
Matches



But this will be really, really slow!

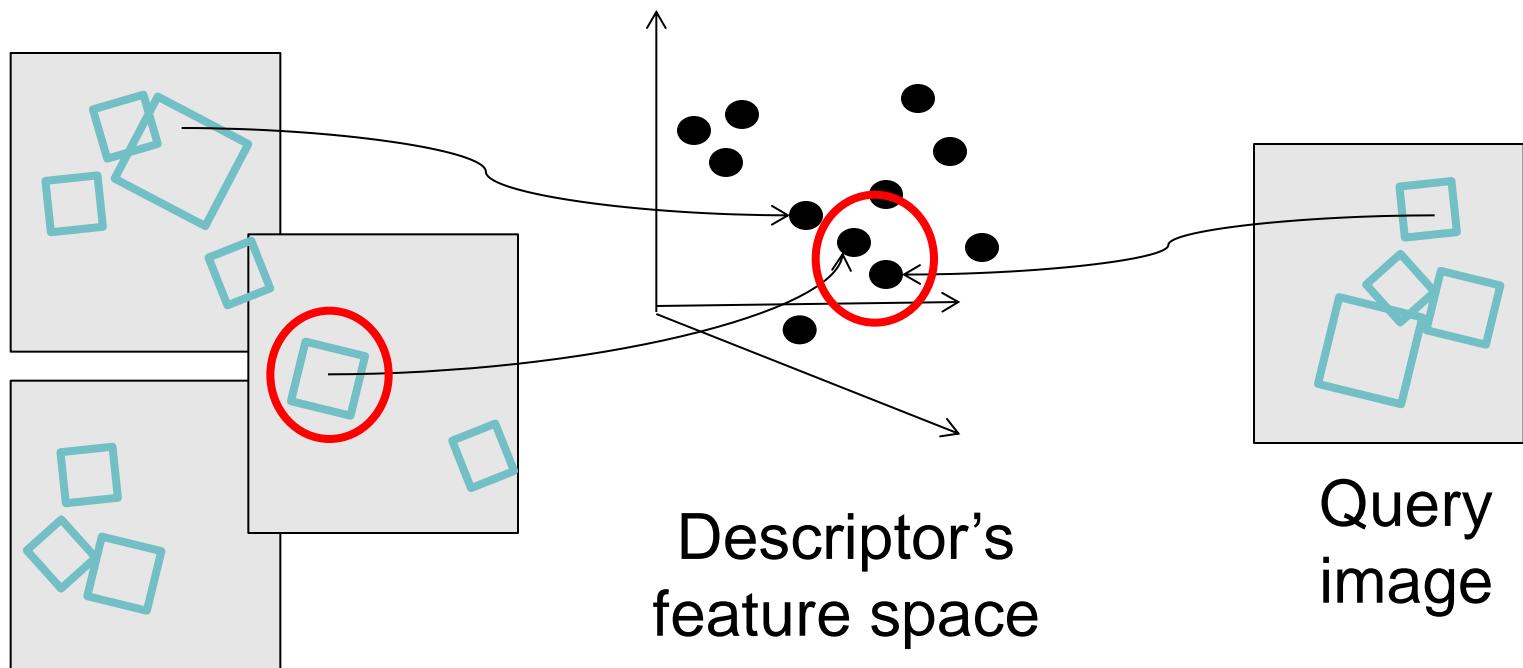
# Indexing local features

- Each patch / region has a descriptor, which is a point in some high-dimensional feature space (e.g., SIFT)



# Indexing local features

- When we see close points in feature space, we have similar descriptors, which indicates similar local content.



Database  
images

Descriptor's  
feature space

*Easily can have millions of  
features to search!*

# Indexing local features: inverted file index

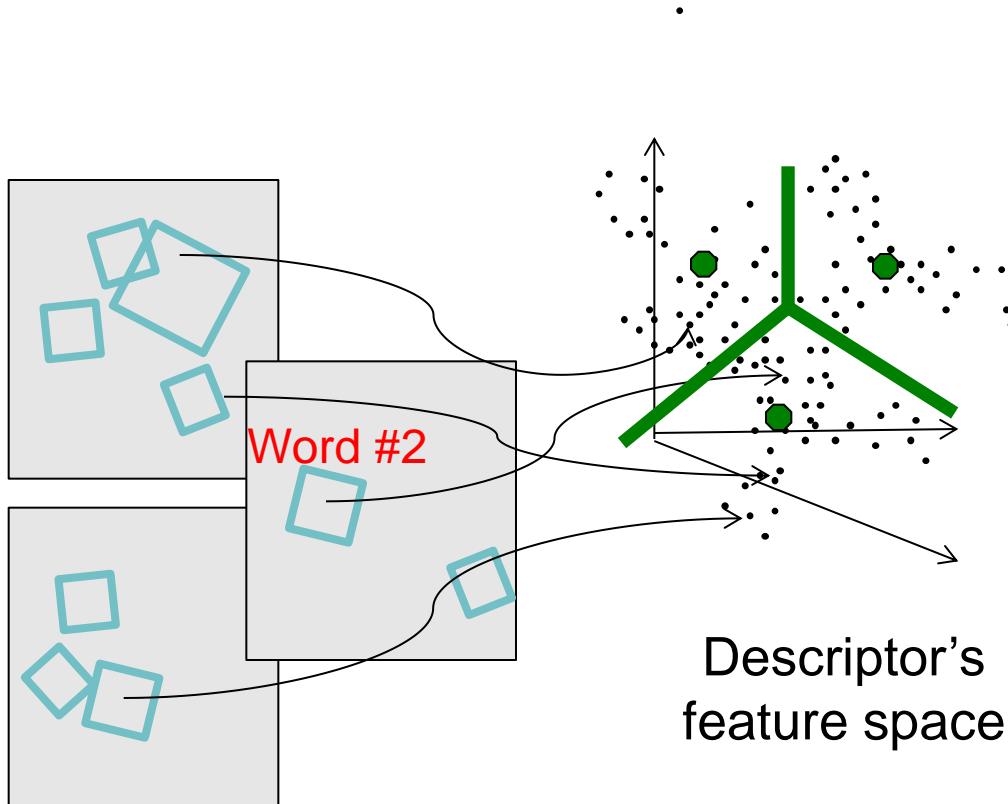
## Index

"Along I-75," From Detroit to Florida; <i>inside back cover</i>	Butterfly Center, McGuire; 134
"Drive I-95," From Boston to Florida; <i>inside back cover</i>	CAA (see AAA)
1929 Spanish Trail Roadway; 101-102,104	CCC, The; 111,113,115,135,142
511 Traffic Information; 83	Ca d'Zan; 147
A1A (Barrier Isl) - I-95 Access; 86	Caloosahatchee River; 152
AAA (and CAA); 83	Name; 150
AAA National Office; 88	Canaveral Natnl Seashore; 173
Abbreviations,	Cannon Creek Airpark; 130
Colored 25 mile Maps; cover	Canopy Road; 106,160
Exit Services; 196	Cape Canaveral; 174
Travelogue; 85	Castillo San Marcos; 169
Africa; 177	Cave Diving; 131
Agricultural Inspection Stns; 126	Cayo Costa, Name; 150
Ah-Tah-Thi-Ki Museum; 160	Celebration; 93
Air Conditioning, First; 112	Charlotte County; 149
Alabama; 124	Charlotte Harbor; 150
Alachua; 132	Chautauqua; 116
County; 131	Chipley; 114
Alafia River; 143	Name; 115
Alapaha, Name; 126	Choctawatchee, Name; 115
Alfred B Macay Gardens; 106	Circus Museum, Ringling; 147
Alligator Alley; 154-155	Citrus; 88,97,130,136,140,180
Alligator Farm, St Augustine; 169	CityPlace, W Palm Beach; 180
Alligator Hole (definition); 157	City Maps,
Alligator, Buddy; 155	Ft Lauderdale Expwys; 194-195
Alligators; 100,135,138,147,156	Jacksonville; 163
Anastasia Island; 170	Kissimmee Expwys; 192-193
Anhaica; 108-109,146	Miami Expressways; 194-195
Apalachicola River; 112	Orlando Expressways; 192-193
Appleton Mus of Art; 136	Pensacola; 26
Aquifer; 102	Tallahassee; 191
Arabian Nights; 94	Tampa-St. Petersburg; 63
Art Museum, Ringling; 147	St. Augustine; 191
Aruba Beach Cafe; 183	Civil War; 100,108,127,138,141
Aucilla River Project; 106	Clearwater Marine Aquarium; 187
Babcock-Web WMA; 151	Collier County; 154
Bahia Mar Marina; 184	Collier, Barron; 152
Baker County; 99	Colonial Spanish Quarters; 168
Barefoot Mallmen; 182	Columbia County; 101,128
Barge Canal; 137	Coquina Building Material; 165
Bee Line Expy; 80	Corkscrew Swamp, Name; 154
Belz Outlet Mall; 89	Cowboys; 95
Bernard Castro; 136	Crab Trap II; 144
Big "I"; 165	Cracker, Florida; 88,95,132
Big Cypress; 155,158	Crostown Expy; 11,35,98,143
Big Foot Monster; 105	Cuban Bread; 184
Billie Swamp Safari; 160	Dade Battlefield; 140
Blackwater River SP; 117	Dade, Maj. Francis; 139-140,161
Blue Angels	Danie Beach Hurricane; 184
	Daniel Boone, Florida Walk; 117
	Daytona Beach; 172-173
	De Land; 87
	Driving Lanes; 85
	Duval County; 163
	Eau Gallie; 175
	Edison, Thomas; 152
	Eglin AFB; 116-118
	Eight Reale; 176
	Ellenton; 144-145
	Emanuel Point Wreck; 120
	Emergency Callboxes; 83
	Epiphytes; 142,148,157,159
	Escambia Bay; 119
	Bridge (I-10); 119
	County; 120
	Estero; 153
	Everglade,90,95,139-140,154-160
	Draining of; 156,181
	Wildlife MA; 160
	Wonder Gardens; 154
	Falling Waters SP; 115
	Fantasy of Flight; 95
	Fayer Dykes SP; 171
	Fires, Forest; 166
	Fires, Prescribed; 148
	Fisherman's Village; 151
	Flagler County; 171
	Flagler, Henry; 97,165,167,171
	Florida Aquarium; 186
	Florida,
	12,000 years ago; 187
	Cavern SP; 114
	Map of all Expressways; 2-3
	Mus of Natural History; 134
	National Cemetery ; 141
	Part of Africa; 177
	Platform; 187
	Sheriff's Boys Camp; 126
	Sports Hall of Fame; 130
	Sun 'n Fun Museum; 97
	Supreme Court; 107
	Florida's Turnpike (FTP); 178,189
	25 mile Strip Maps; 66
	Administration; 189
	Coin System; 190
	Exit Services; 189
	HEFT; 76,161,190
	History; 189
	Names; 189
	Service Plazas; 190
	Spur SR91; 76
	Ticket System; 190
	Toll Plazas; 190
	Ford, Henry; 152

- For text documents, an efficient way to find all *pages* on which a *word* occurs is to use an index...
- We want to find all *images* in which a *feature* occurs.
- To use this idea, we'll need to map our features to “visual words”.

# Visual words

- Map high-dimensional descriptors to tokens/words by quantizing the feature space



- Quantize via clustering, let cluster centers be the prototype “words”
- Determine which word to assign to each new image region by finding the closest cluster center.

# Visual words

- Example: each group of patches belongs to the same visual word

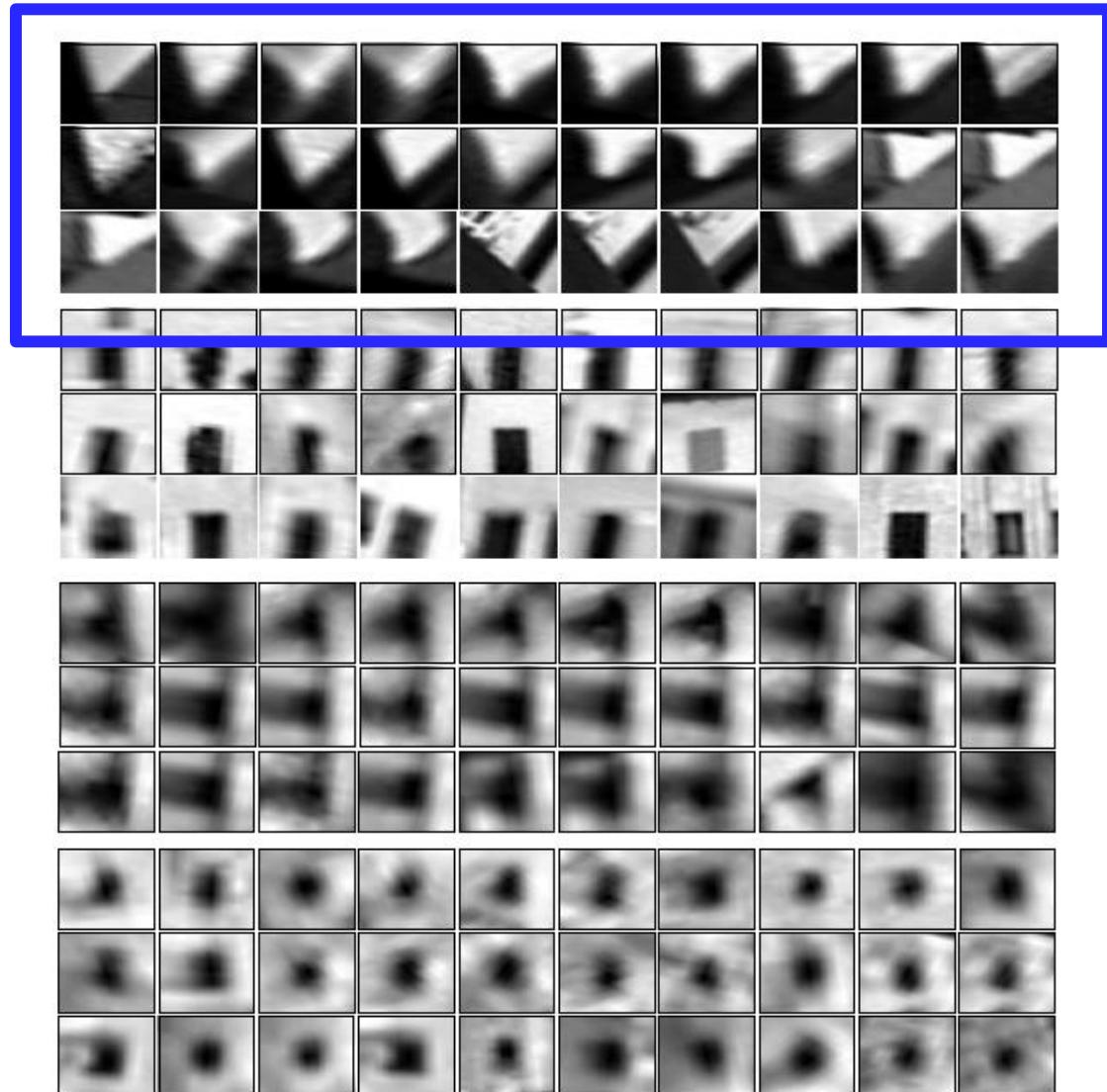
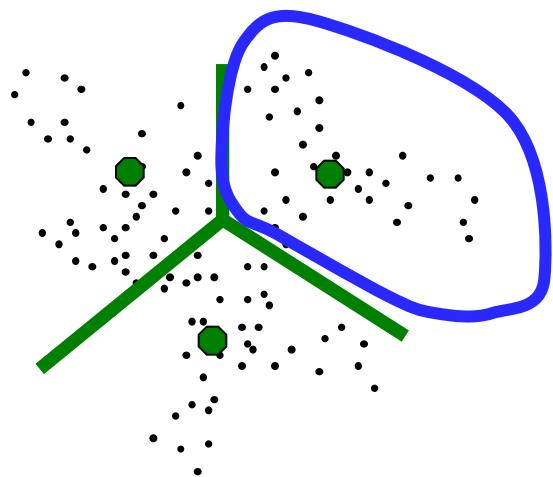


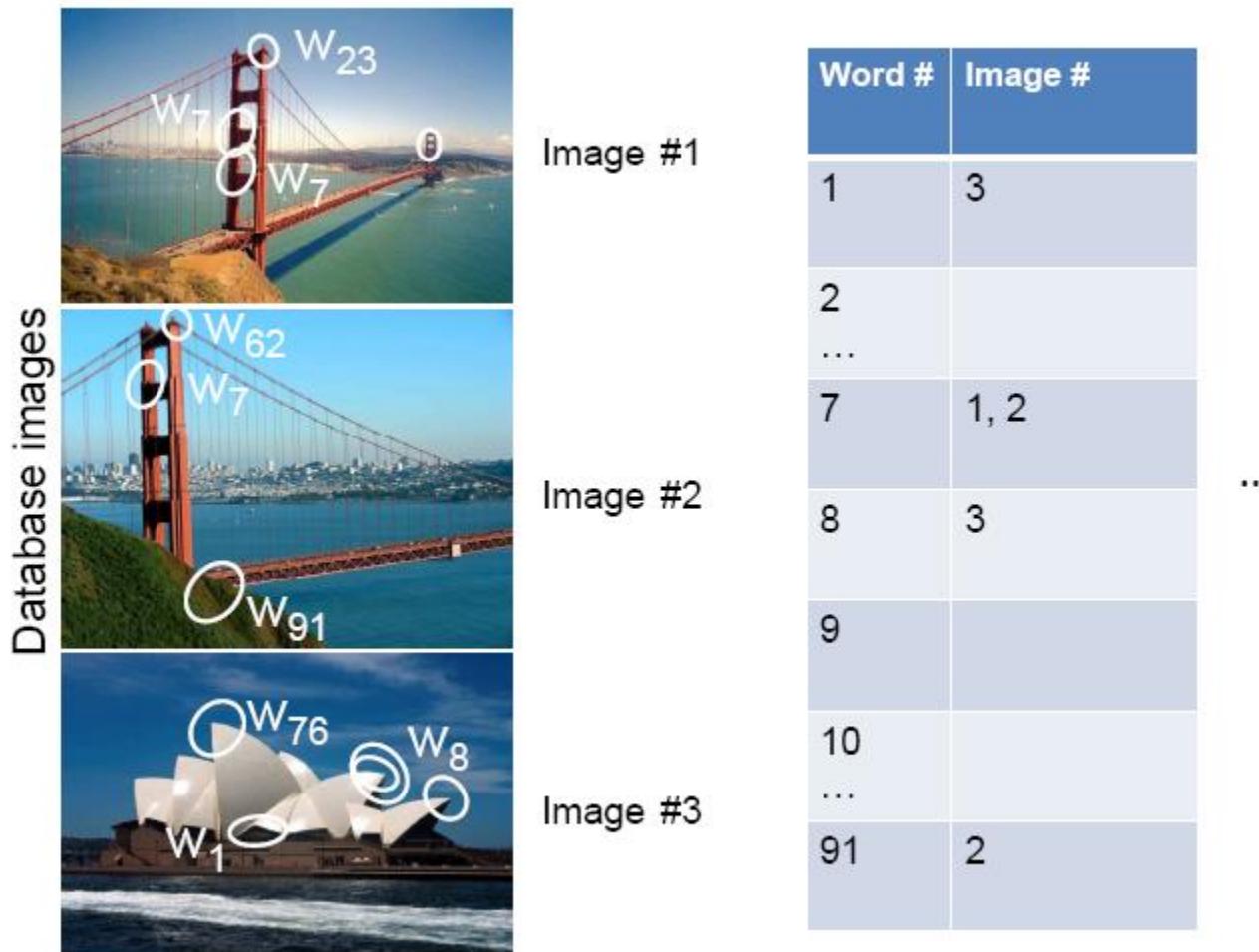
Figure from Sivic & Zisserman, ICCV 2003 Kristen Grauman

# Visual vocabulary formation

Issues:

- Vocabulary size, number of words
- Sampling strategy: where to extract features?
- Clustering / quantization algorithm
- Unsupervised vs. supervised
- What corpus provides features (universal vocabulary?)

# Inverted file index



- Database images are loaded into the index mapping words to image numbers

# Inverted file index



New query image

Word #	Image #
1	3
2	
7	1, 2
8	3
9	
10	
...	
91	2



- New query image is mapped to indices of database images that share a word.

# Instance recognition: remaining issues

- How to summarize the content of an entire image? And gauge overall similarity?
- How large should the vocabulary be? How to perform quantization efficiently?
- Is having the same set of visual words enough to identify the object/scene? How to verify spatial agreement?
- How to score the retrieval results?

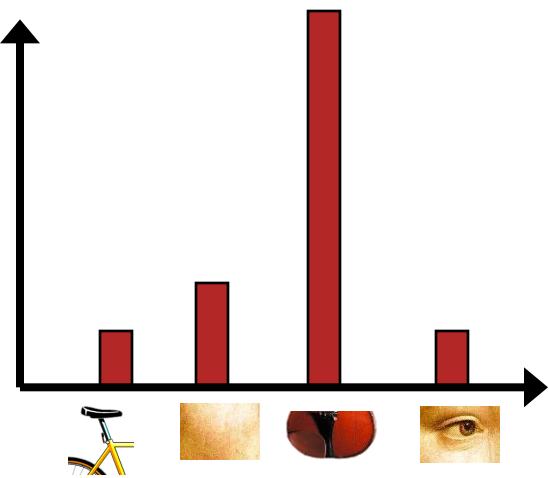
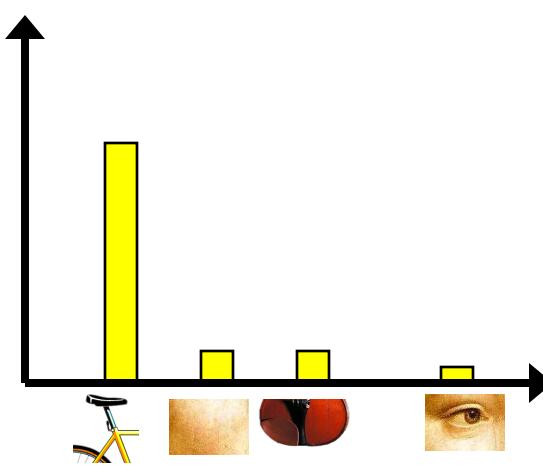
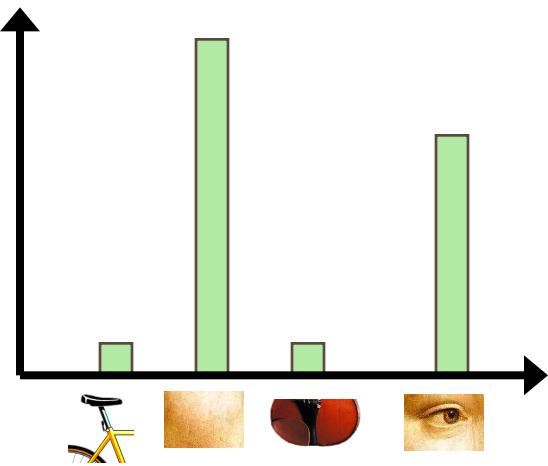
# Analogy to documents

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach us through our eyes. For a long time it was believed that the visual image was formed in the retina, and that visual centers in the cerebral cortex merely "read" the image. In 1959, however, two American scientists, David Hubel and Torsten Wiesel, discovered that the visual system is much more complex than had been thought. Following the path of the optic nerve fibers from the eye to the various centers of the cerebral cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a top-down analysis in a system of nerve cells stored in columns. In this system each column has its specific function and is responsible for a specific detail in the pattern of the retinal image.

**sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel**

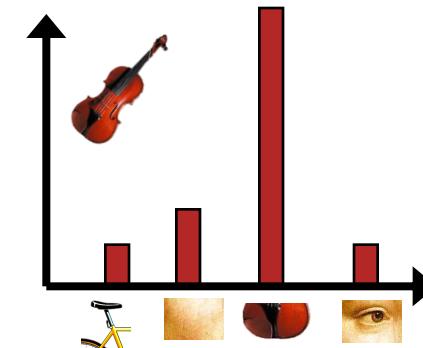
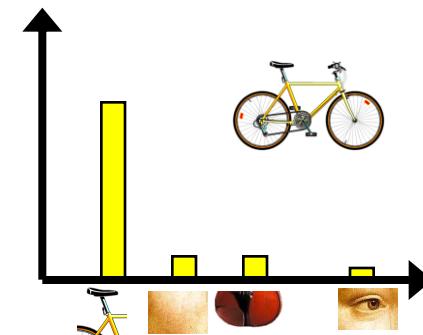
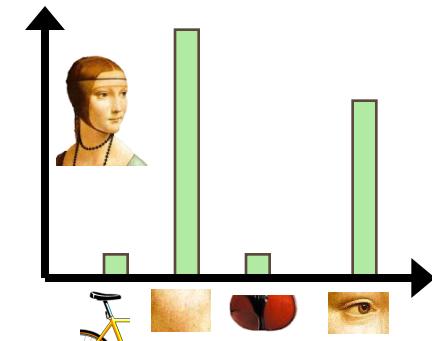
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$660bn. That would put China's exports to annoy the US. China's central bank, the People's Bank, deliberately agreed to let the Chinese yuan rise against the dollar. The Chinese government also needs to encourage domestic demand so that the country can buy more from the country. China has been allowed to let the Chinese yuan against the dollar rise, and permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

**China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value**



# Bags of visual words

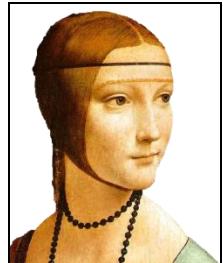
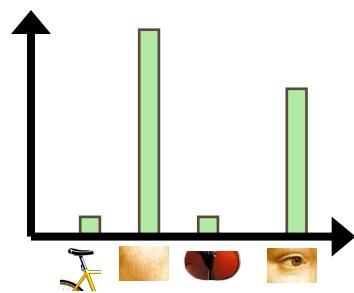
- Summarize entire image based on its distribution (histogram) of word occurrences.
- Analogous to bag of words representation commonly used for documents.



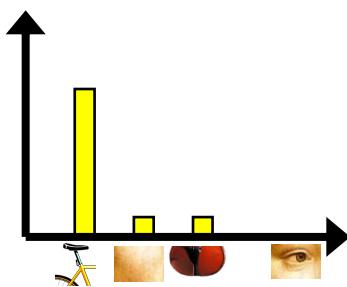
# Comparing bags of words

- Rank frames by normalized scalar product between their (possibly weighted) occurrence counts---*nearest neighbor* search for similar images.

[1 8 1 4]



[5 1 1 0]



$$sim(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|}$$

$$= \frac{\sum_{i=1}^V d_j(i) * q(i)}{\sqrt{\sum_{i=1}^V d_j(i)^2} * \sqrt{\sum_{i=1}^V q(i)}}$$

for vocabulary of  $V$  words

$$\vec{d}_j \quad \vec{q}$$

# Inverted file index and bags of words similarity



New query image

Word #	Image #
1	3
2	
7	1, 2
8	3
9	
10	
...	
91	2
...	

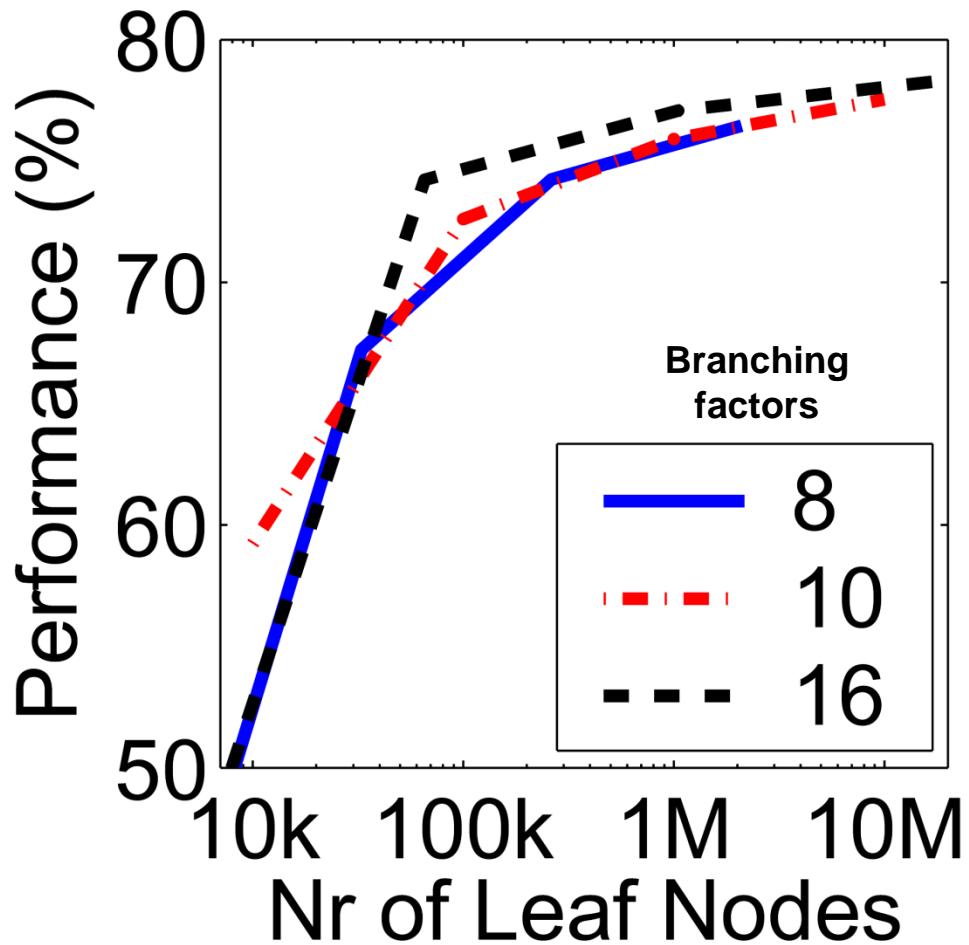


1. Extract words in query
2. Inverted file index to find relevant frames
3. Compare word counts

# Instance recognition: remaining issues

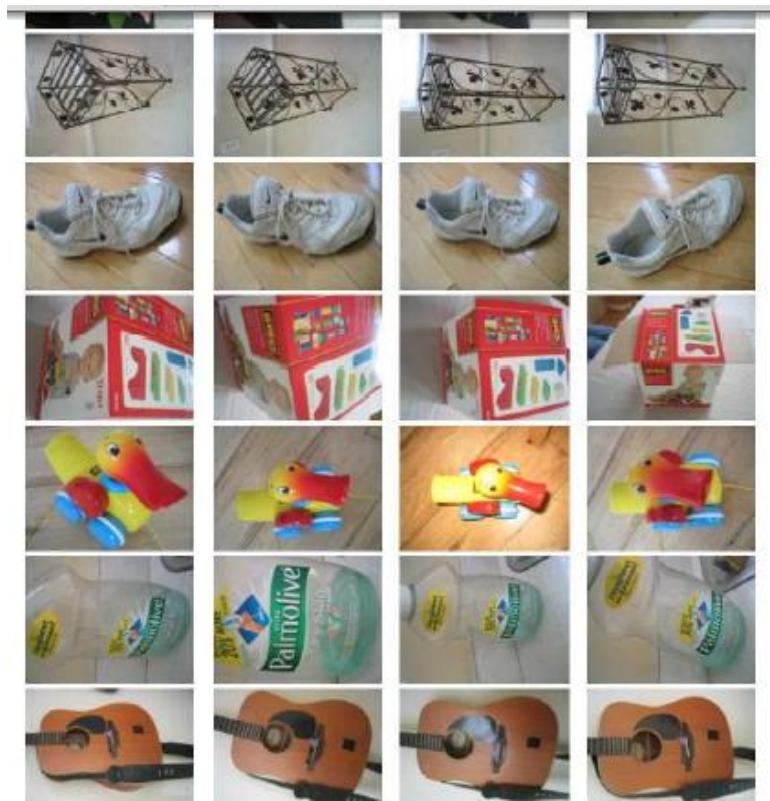
- How to summarize the content of an entire image? And gauge overall similarity?
- How large should the vocabulary be? How to perform quantization efficiently?
- Is having the same set of visual words enough to identify the object/scene? How to verify spatial agreement?
- How to score the retrieval results?

# Vocabulary size



*Influence on performance, sparsity*

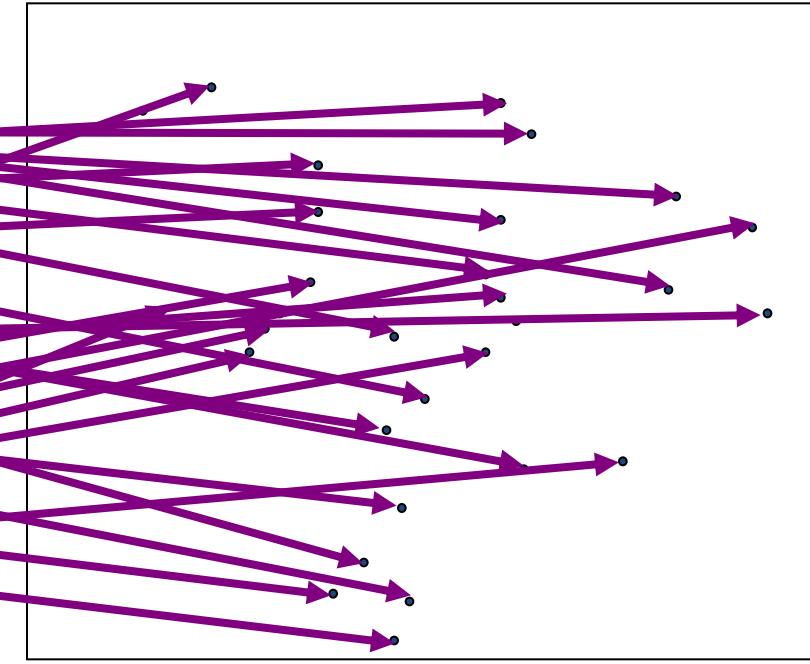
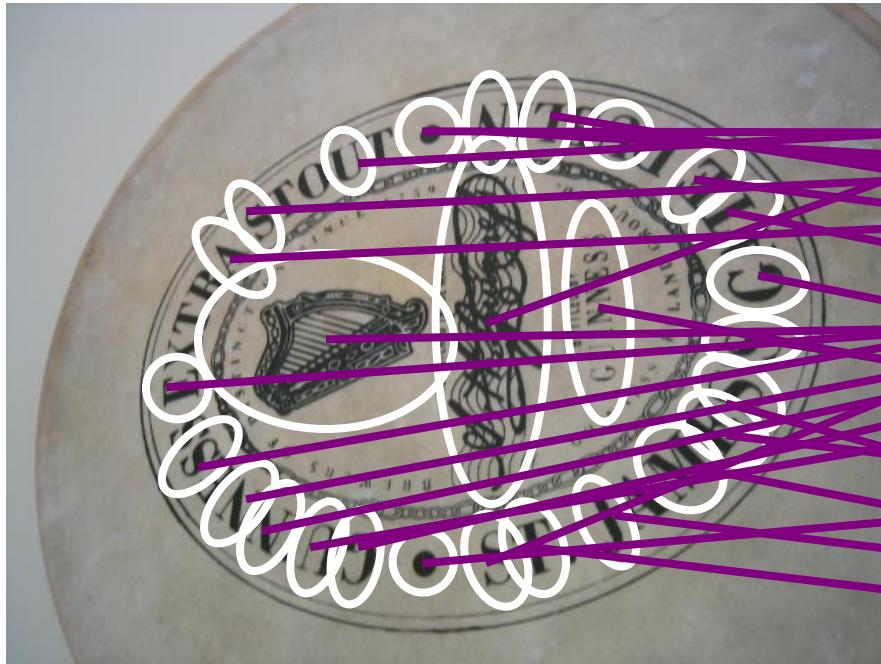
Results for recognition task  
with 6347 images

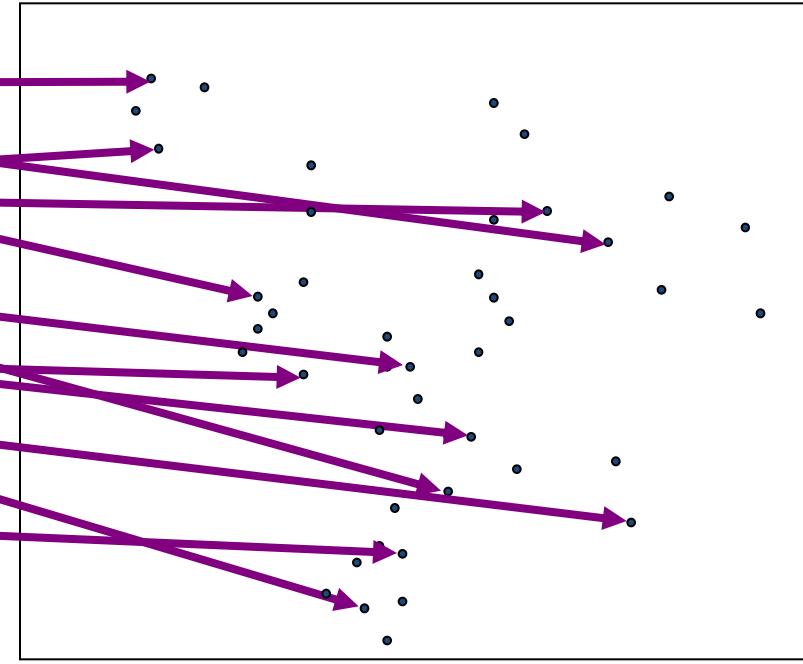
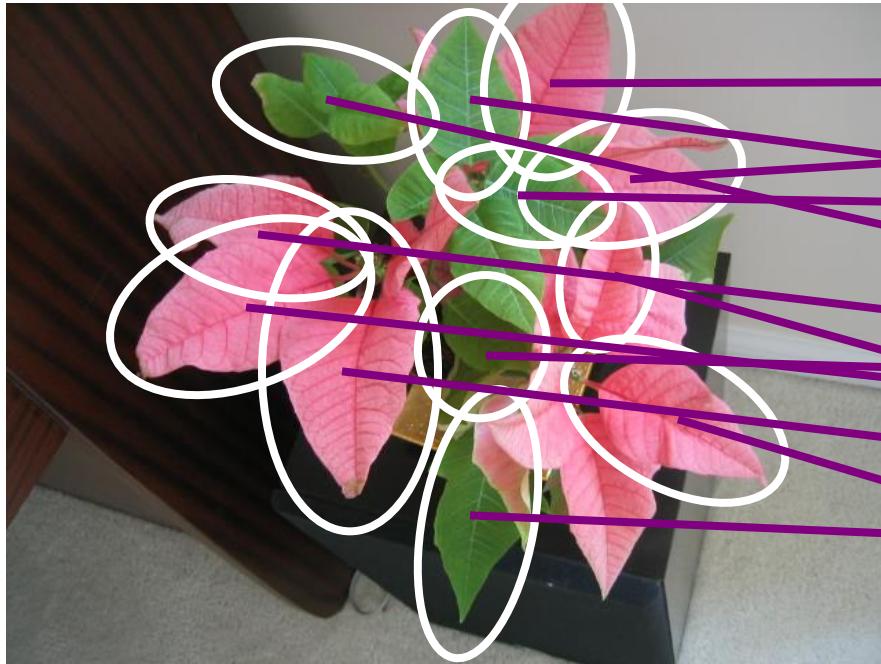


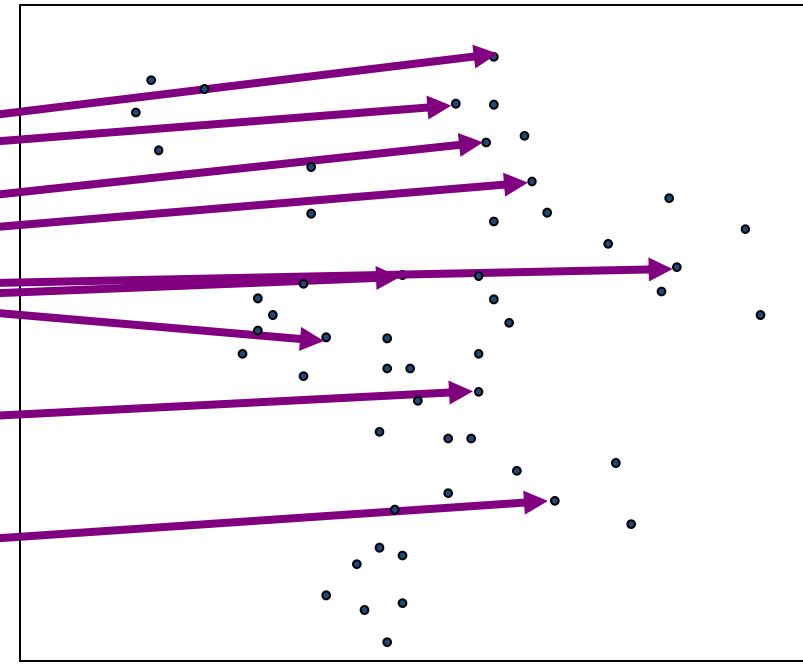
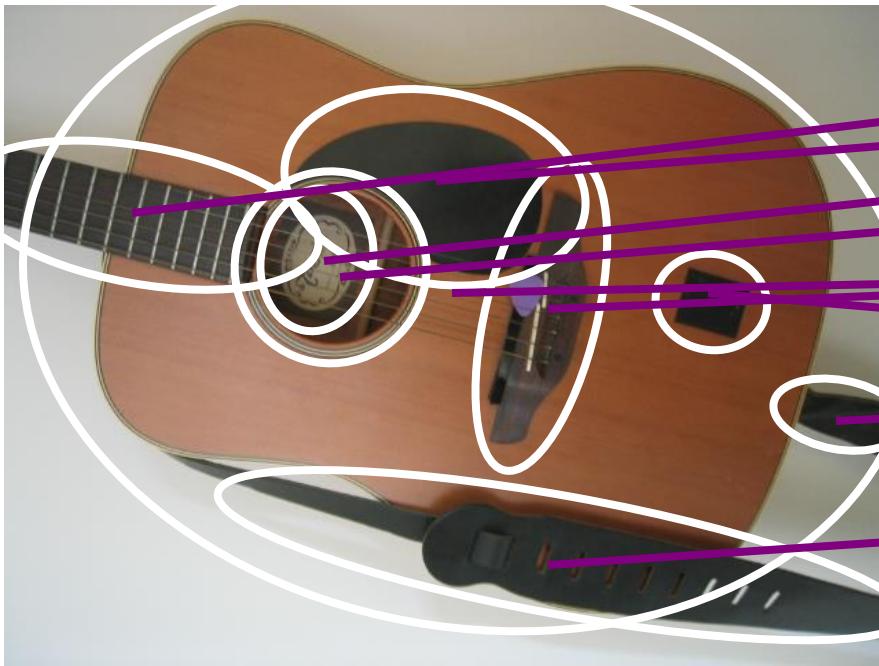
Nister & Stewenius, CVPR 2006  
Kristen Grauman

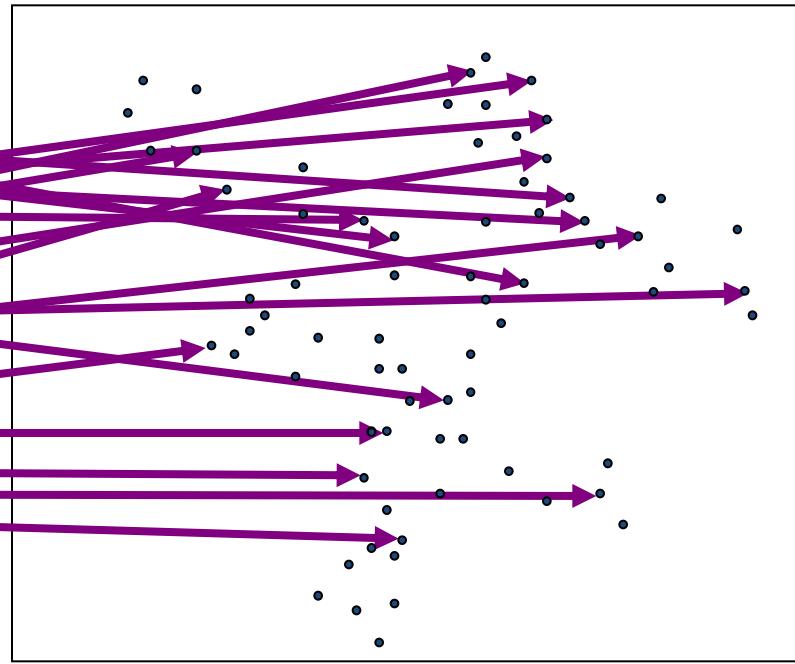
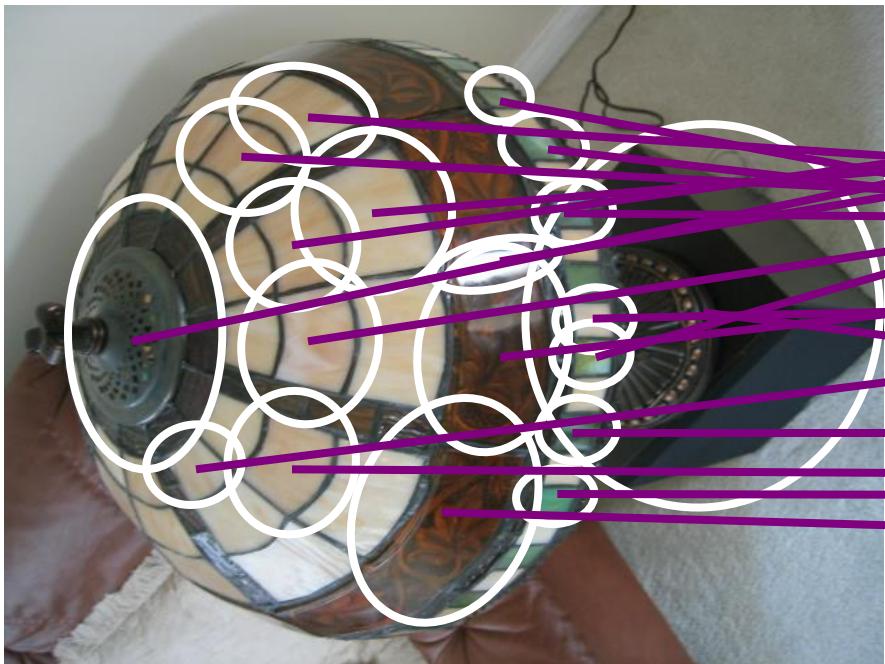
# Recognition with K-tree

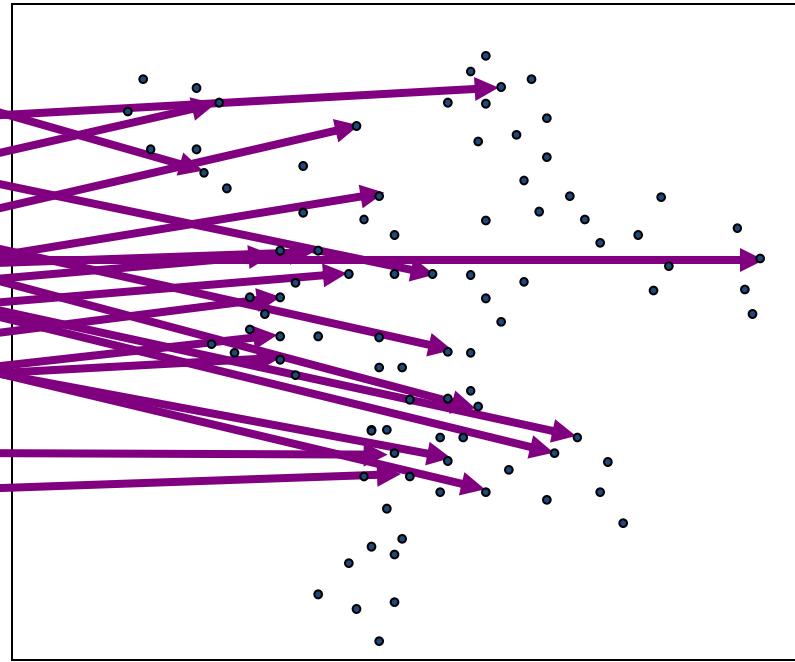
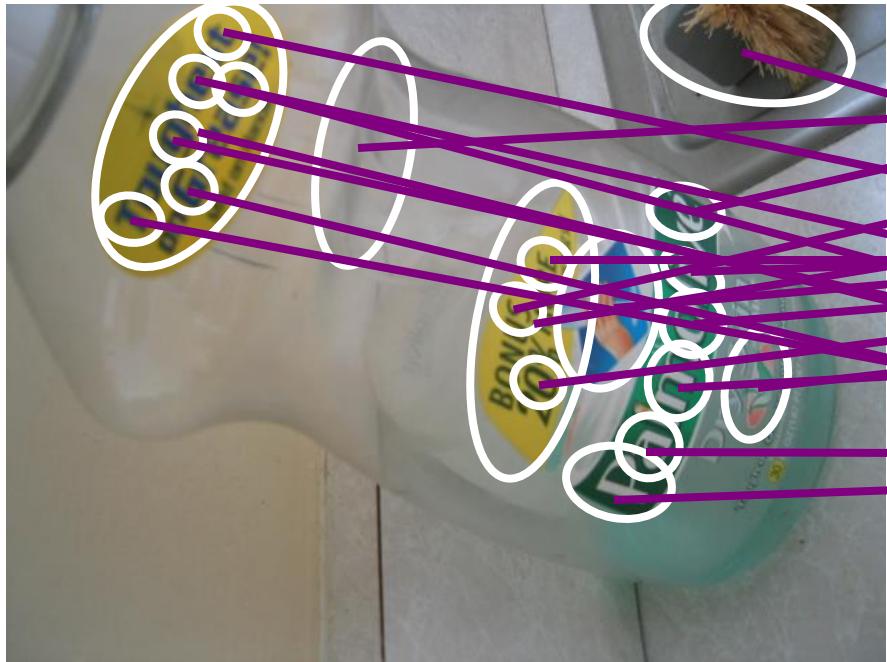
Following slides by David Nister (CVPR 2006)

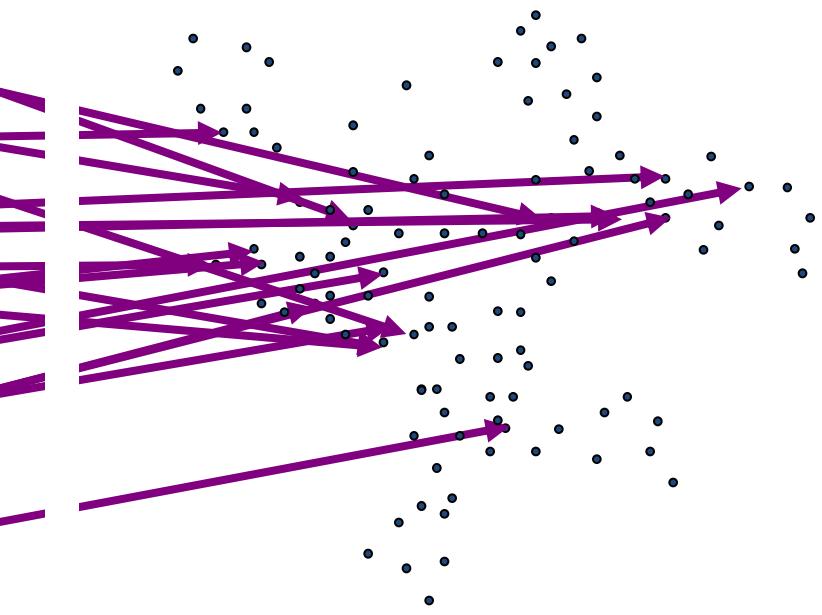


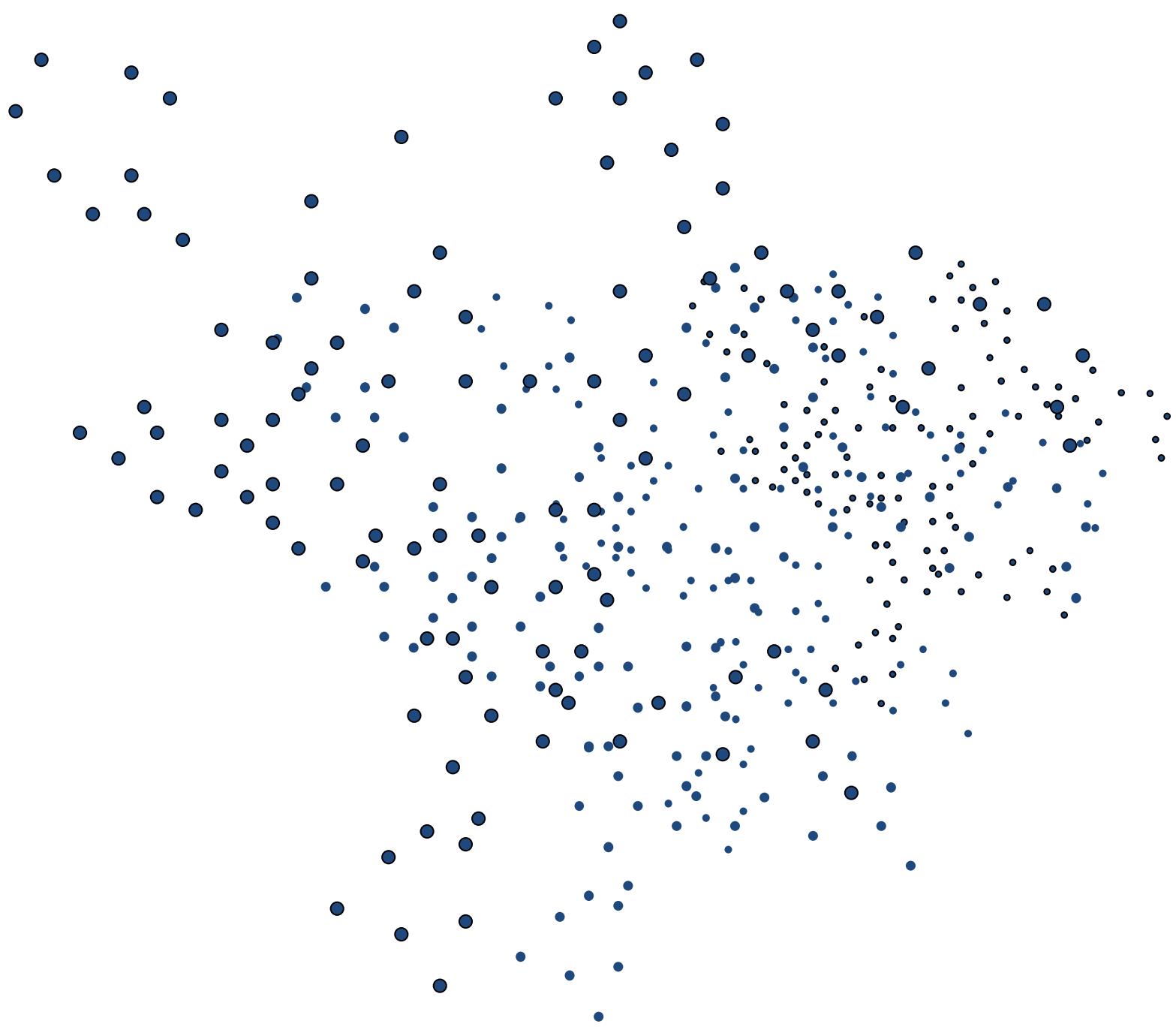


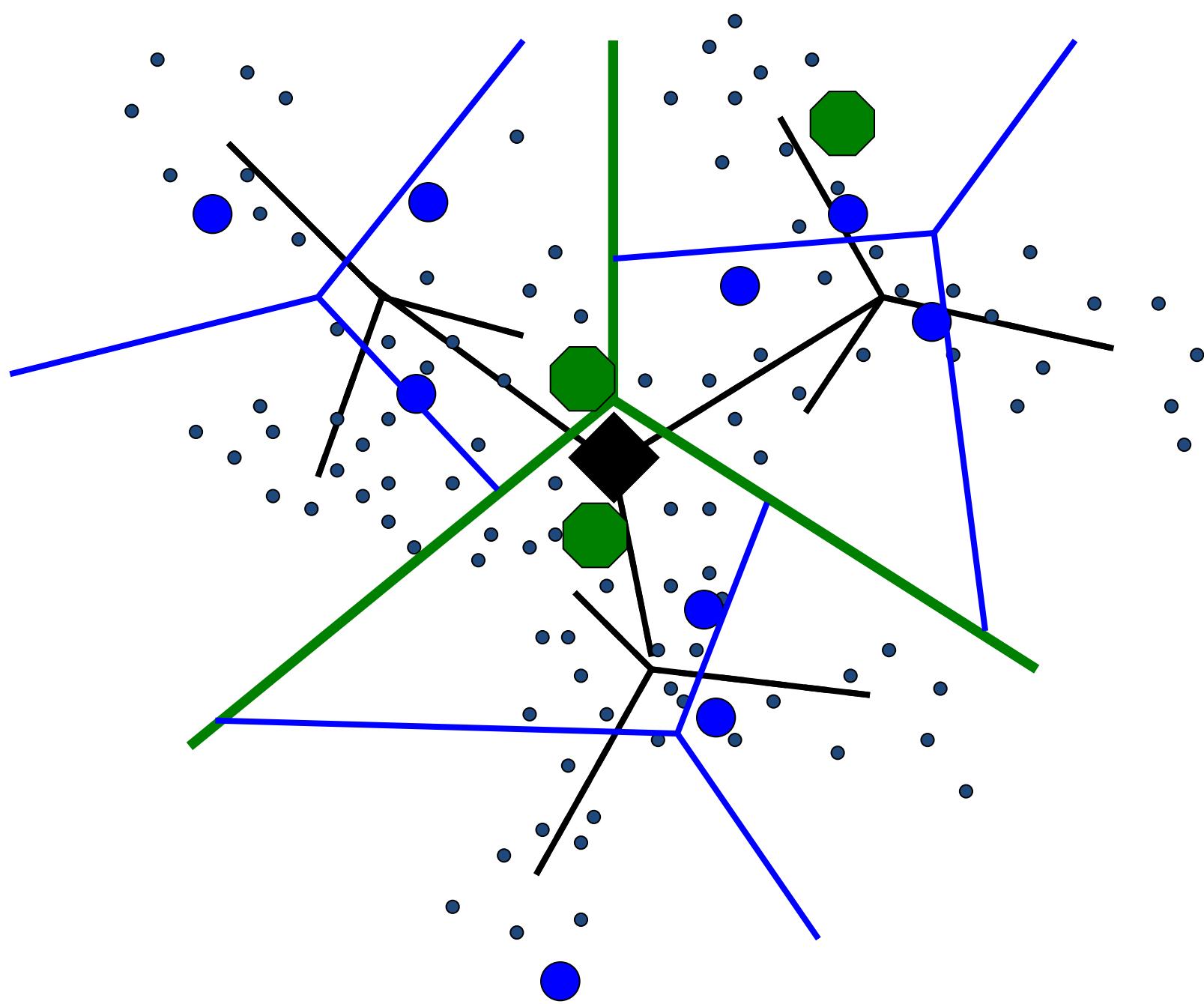


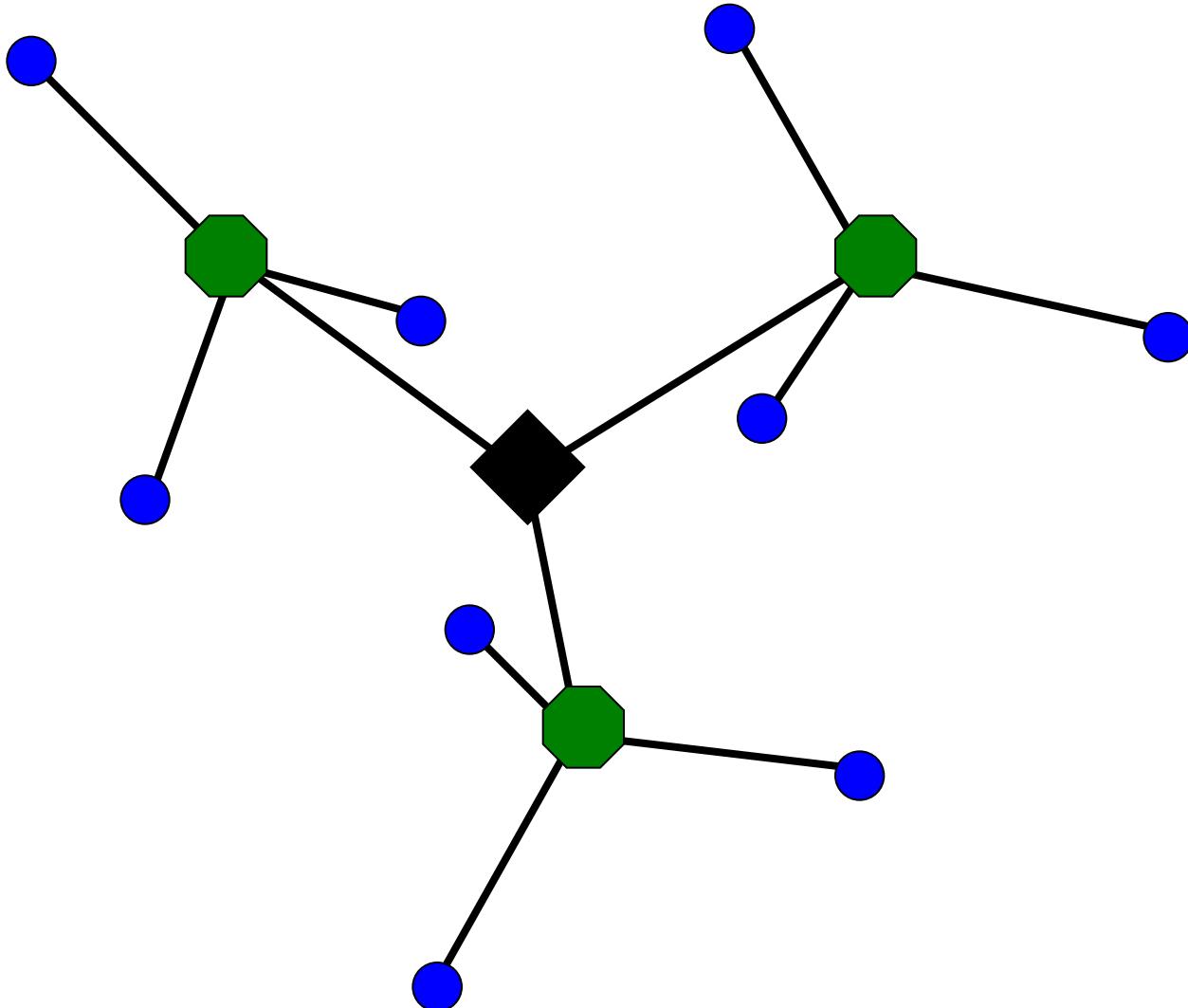


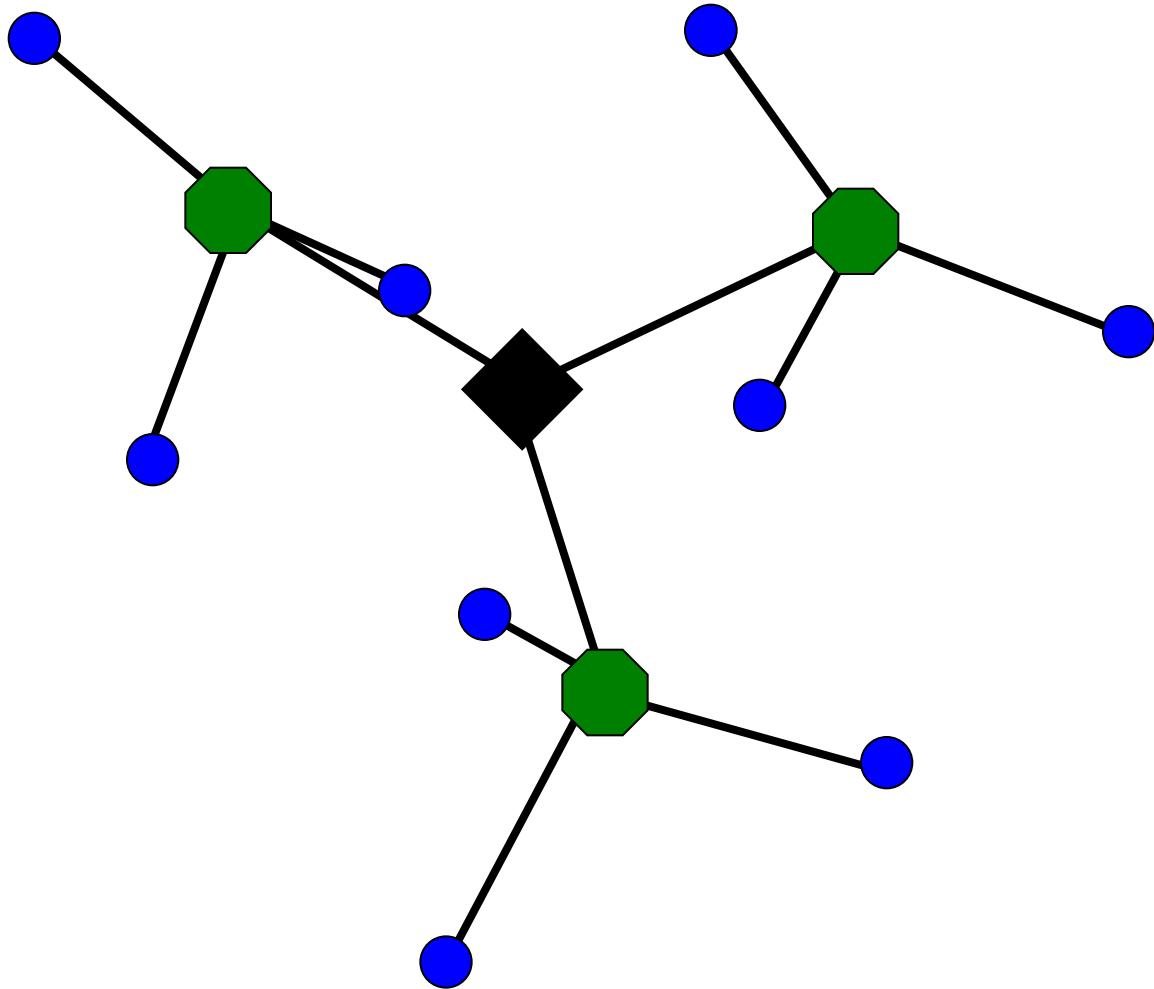


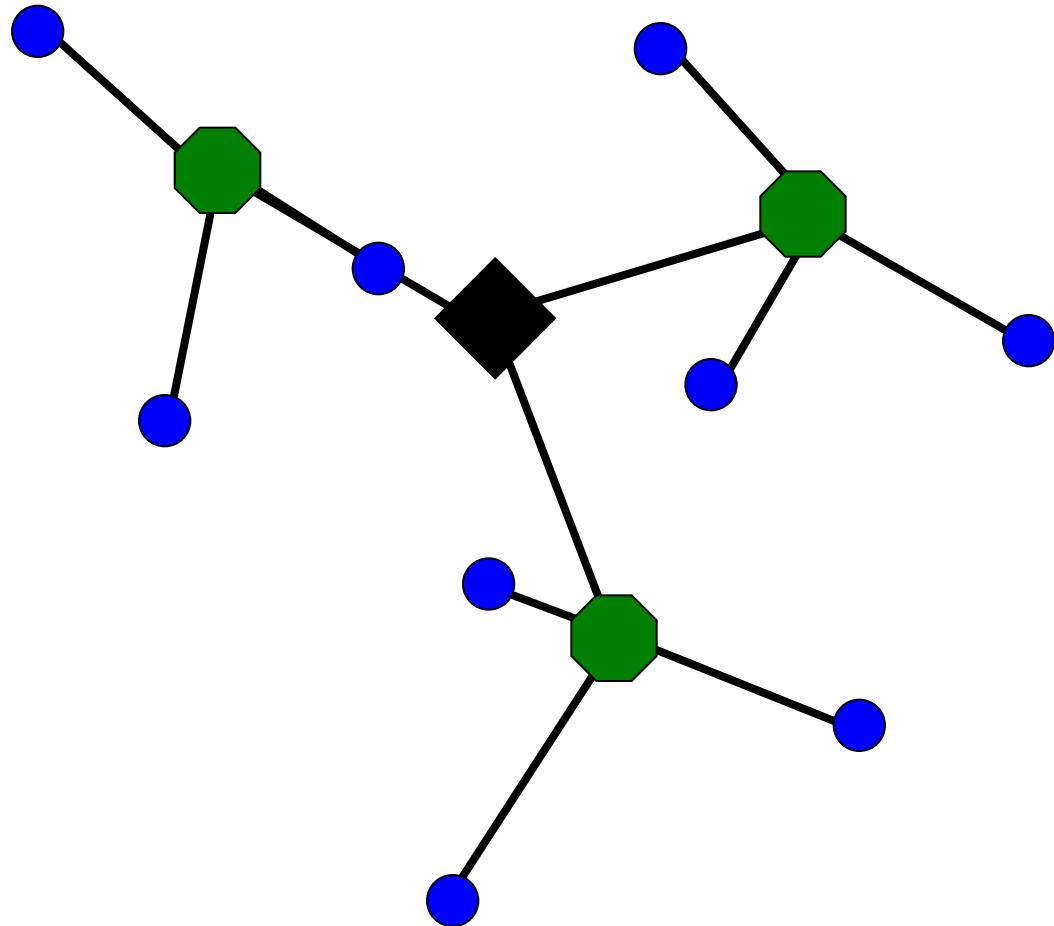


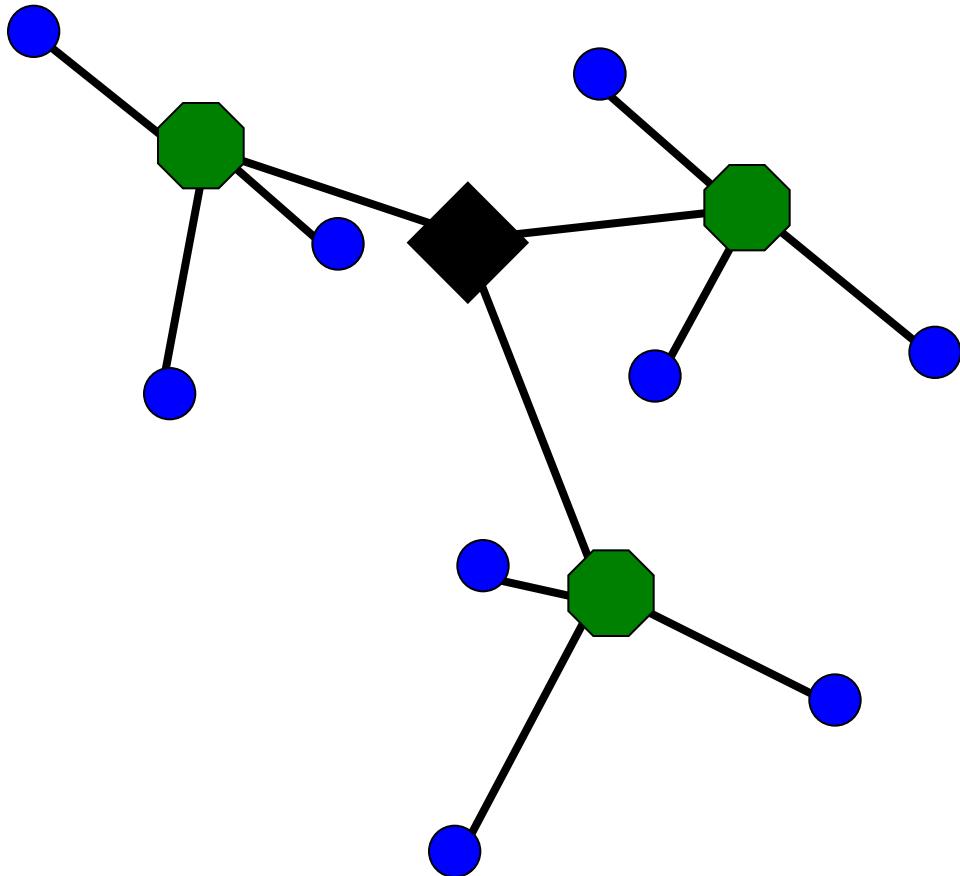


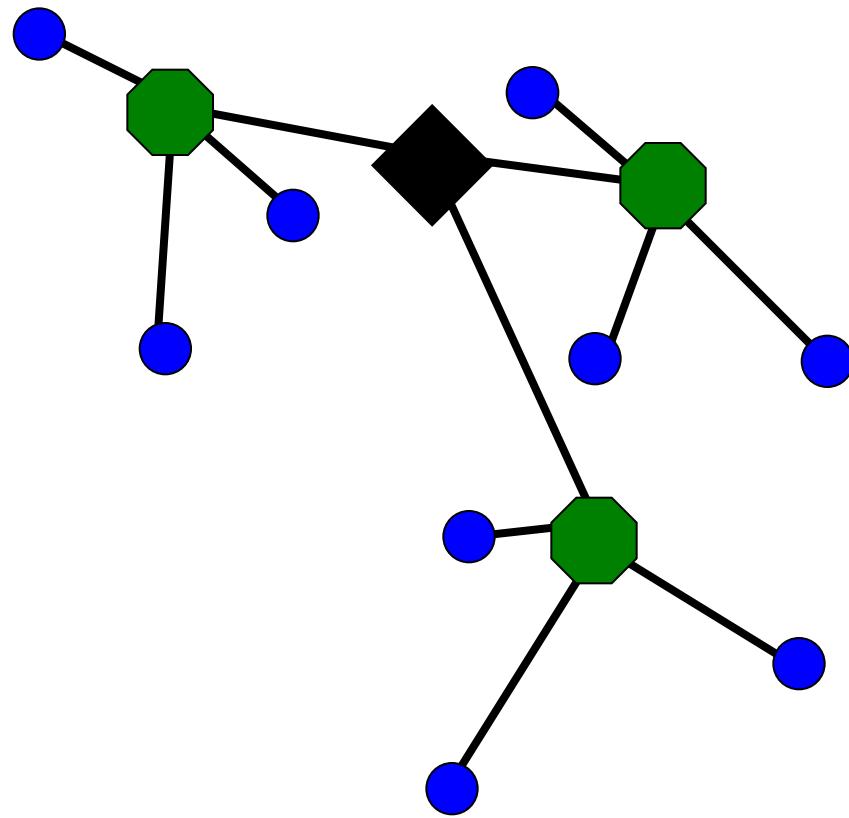


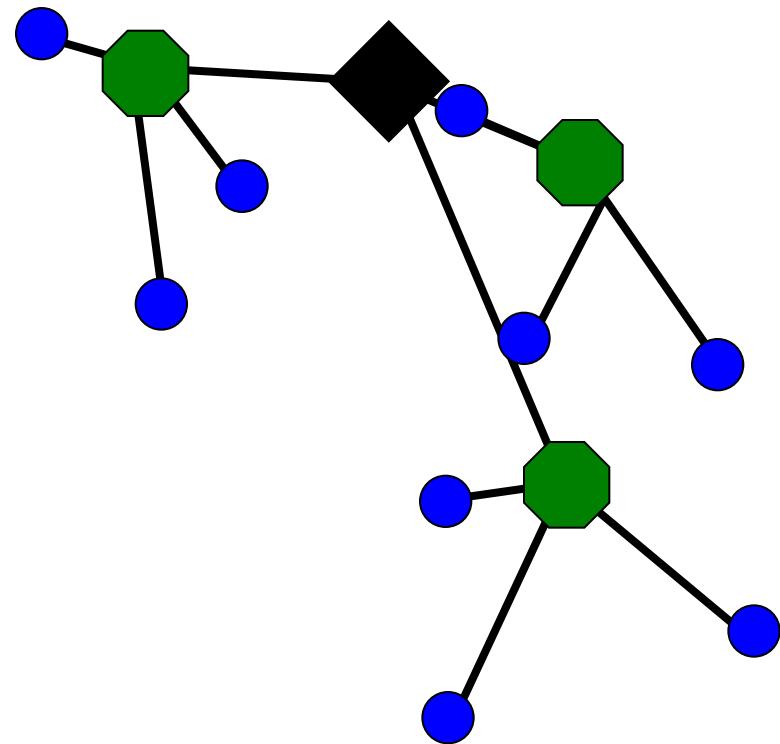


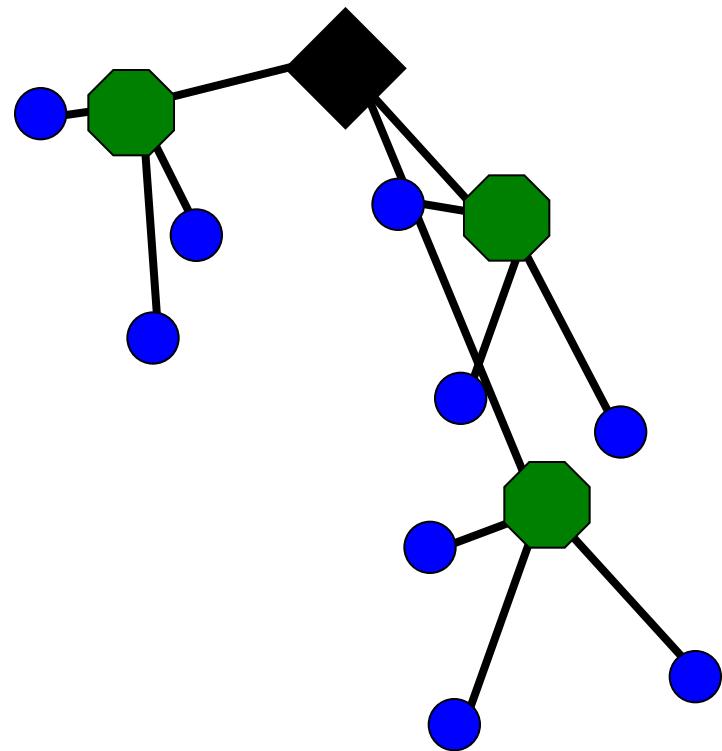


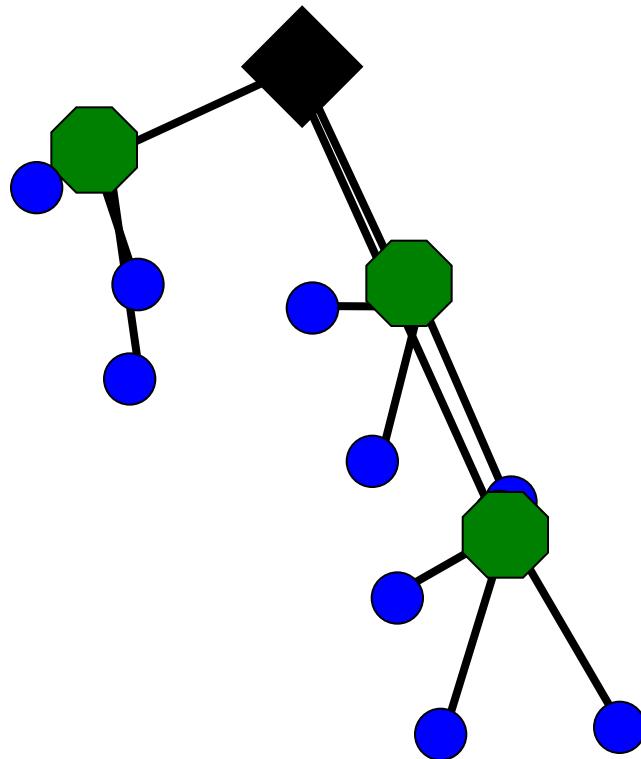


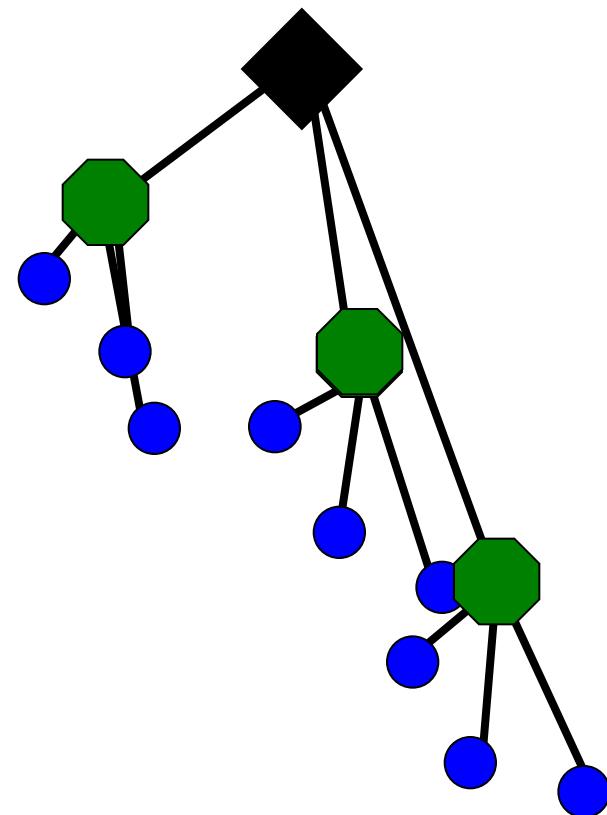


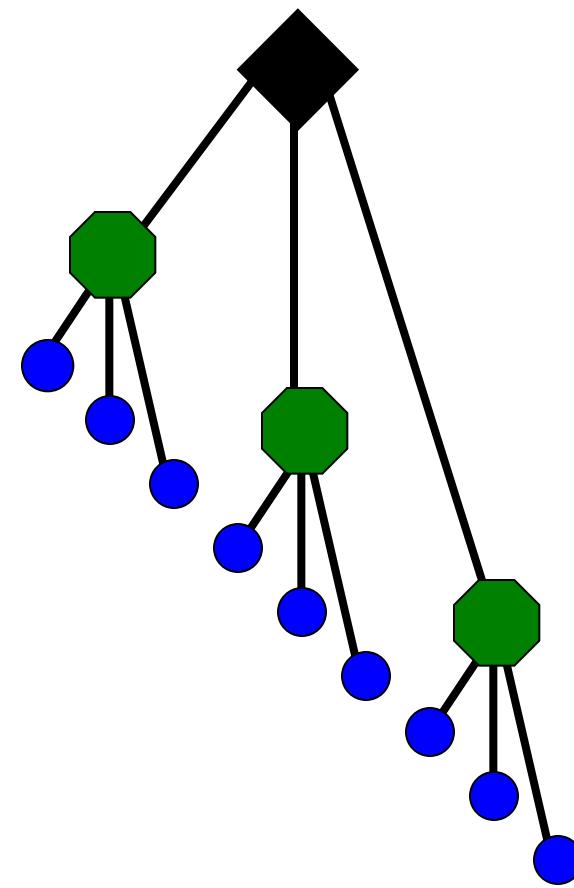


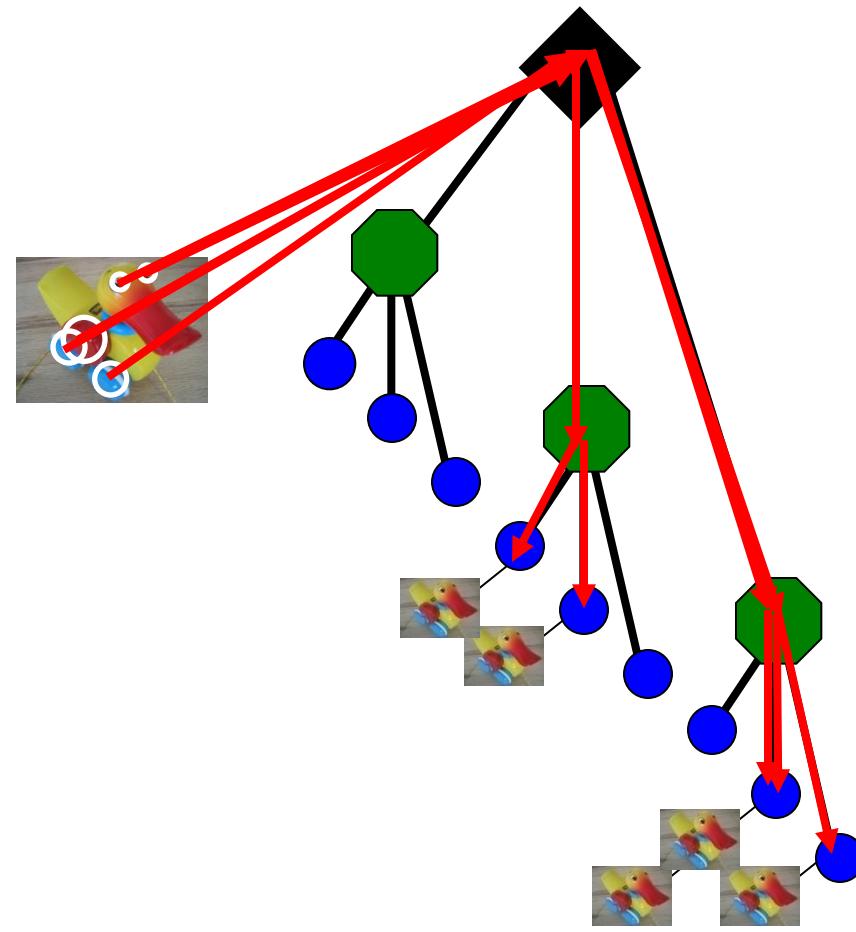


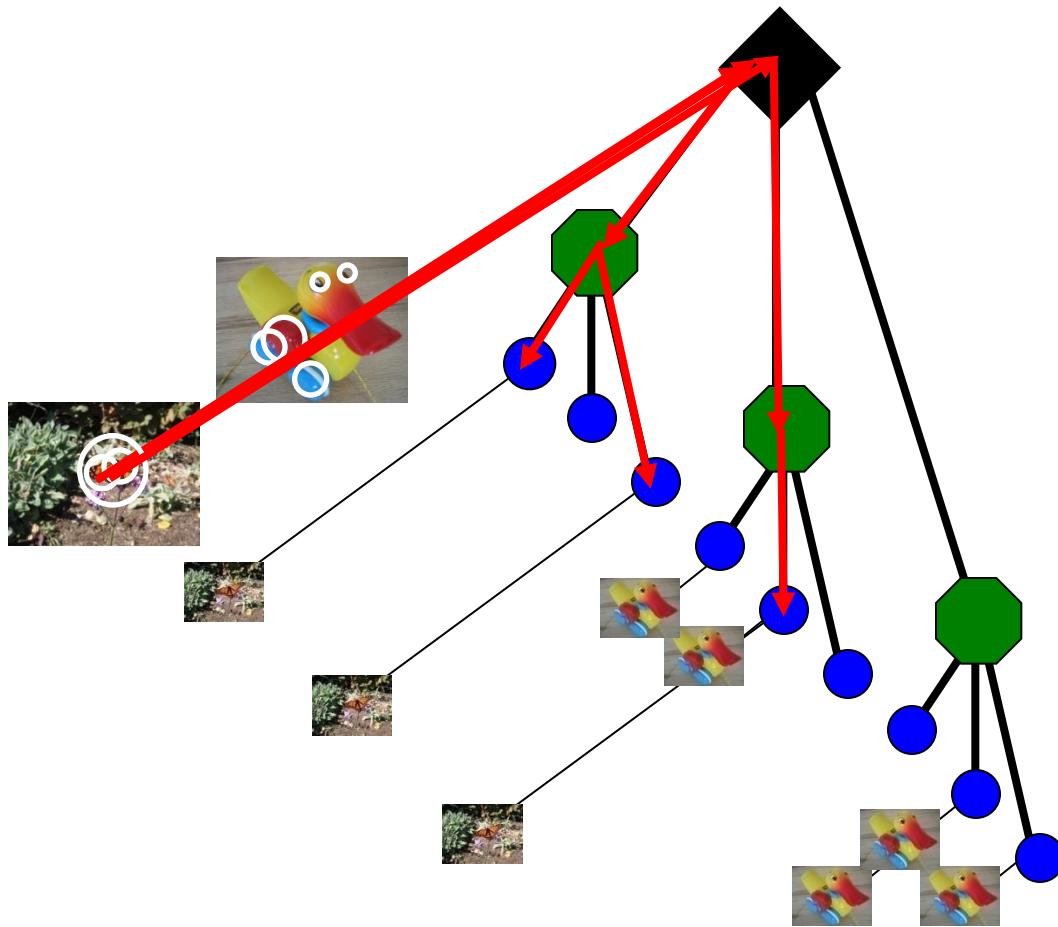


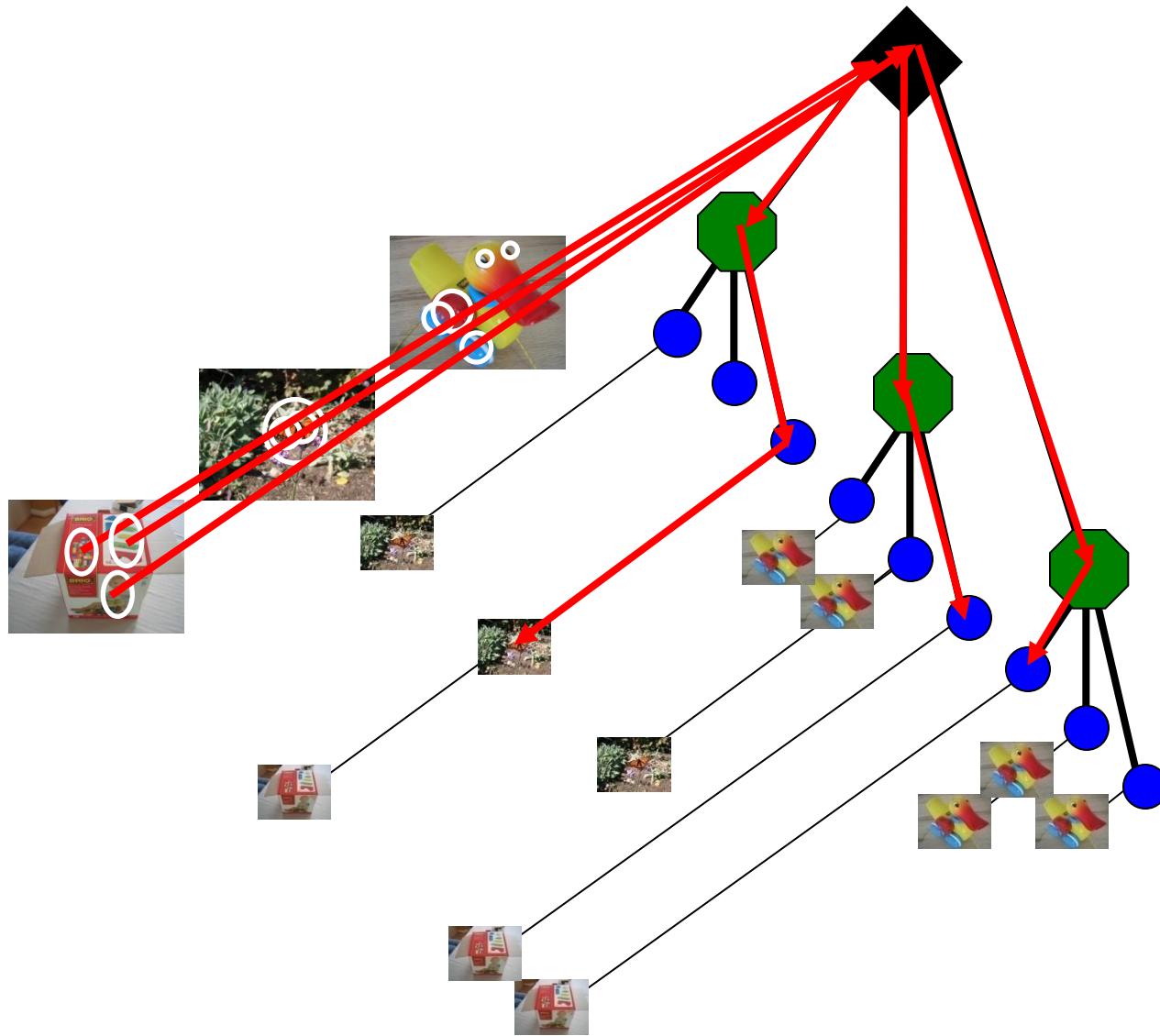


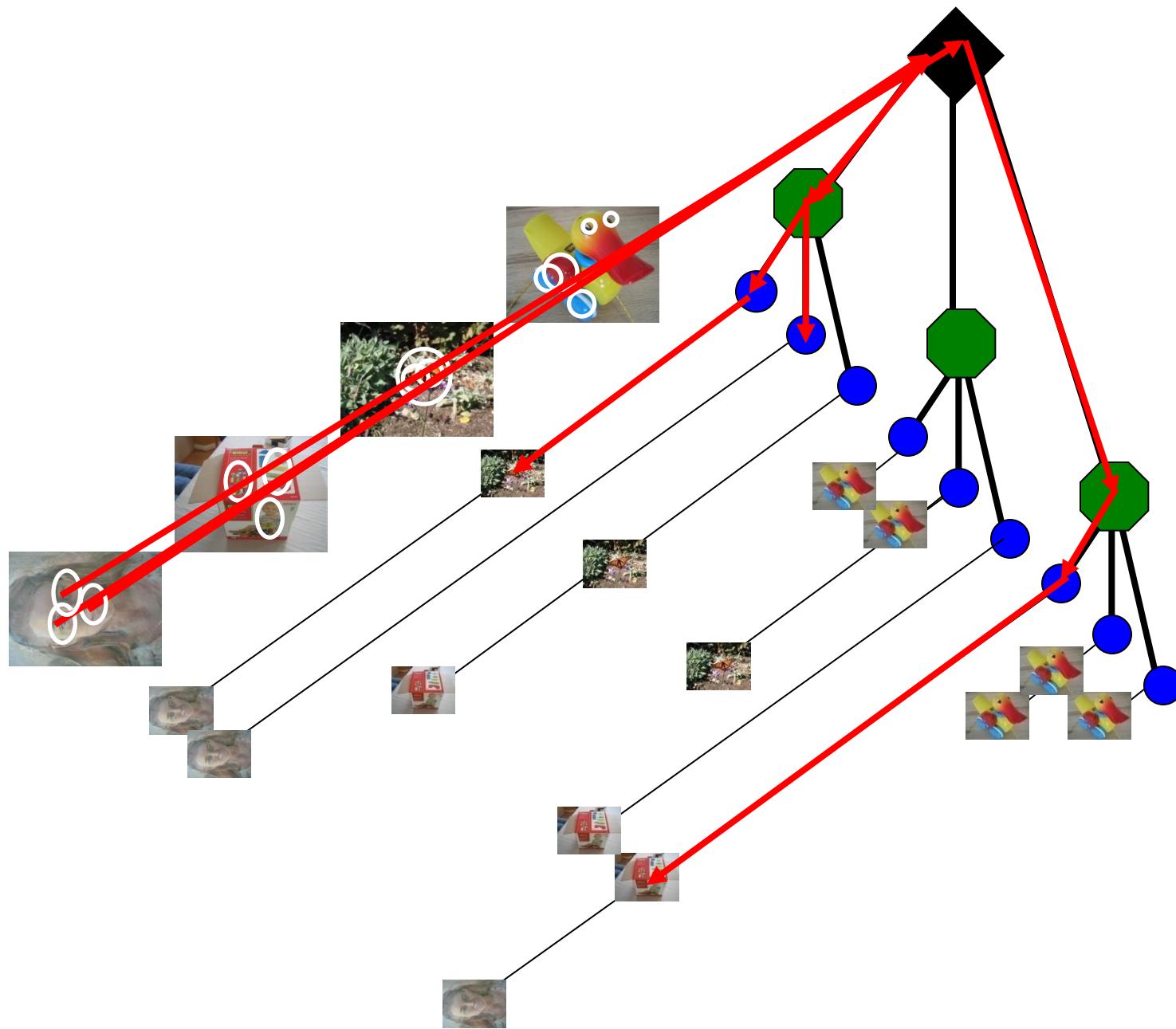


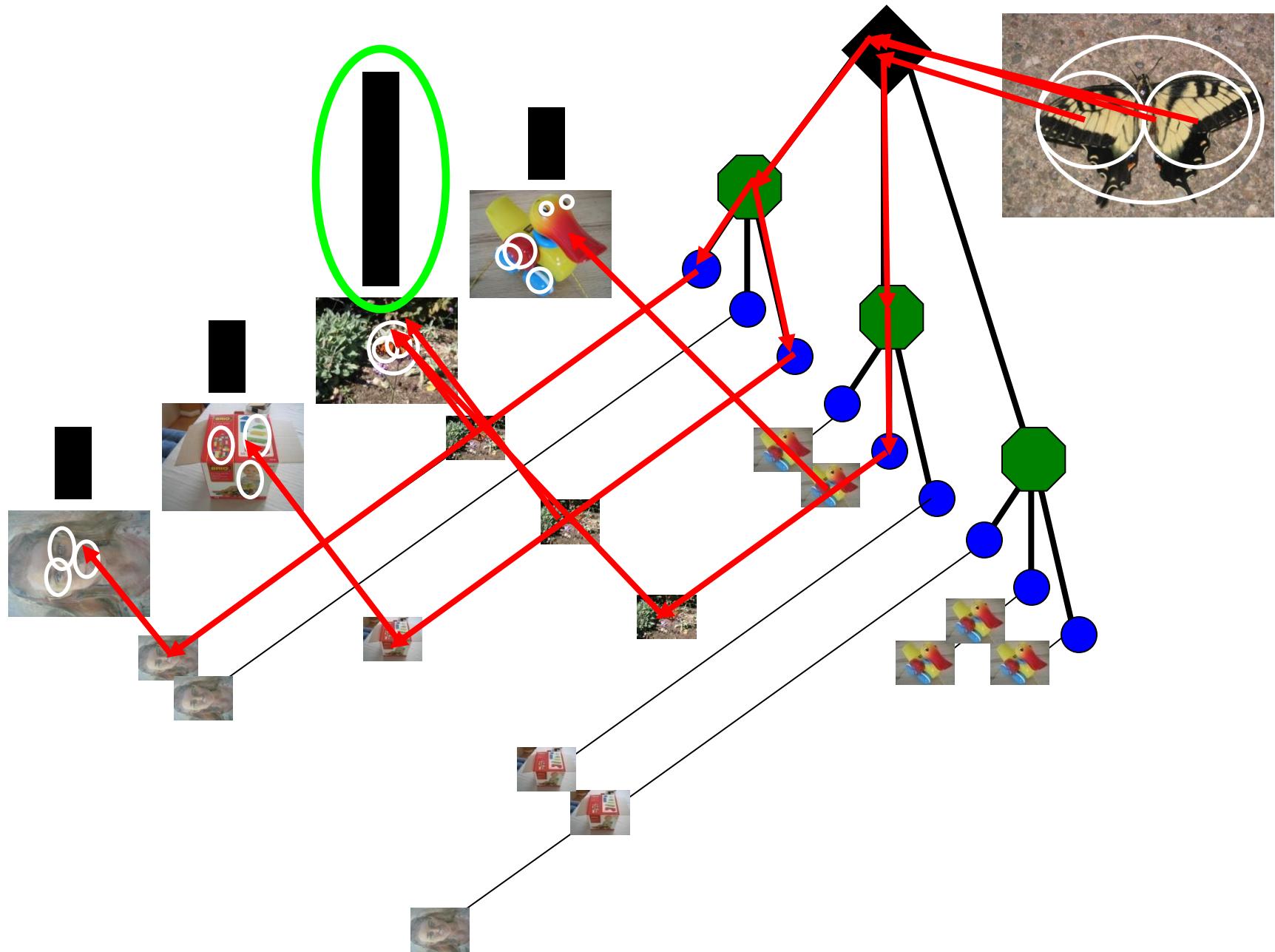












# Vocabulary trees: complexity

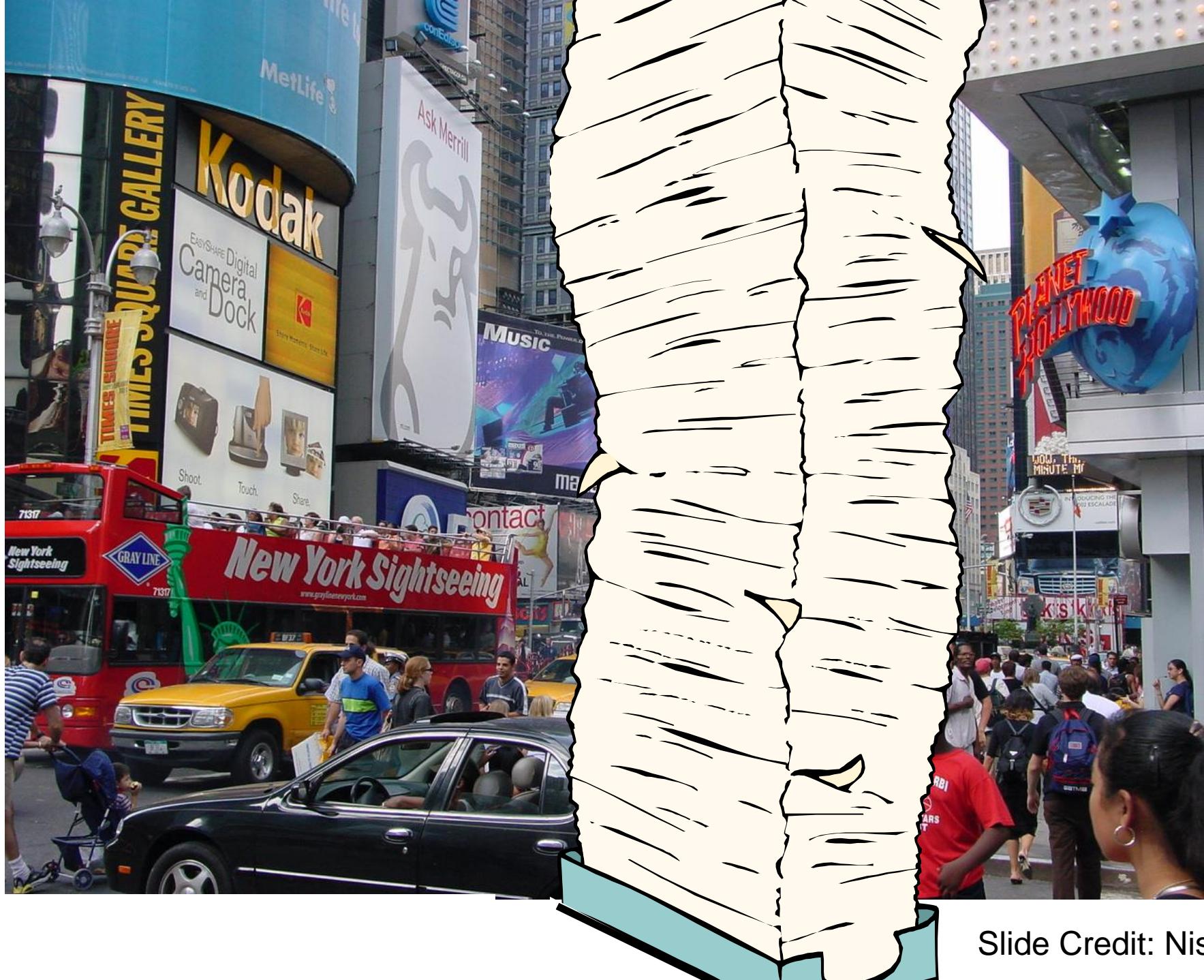
Number of words given tree parameters:  
branching factor and number of levels

Word assignment cost vs. flat vocabulary

110,000,000  
Images in  
5.8 Seconds



Slide Credit: Nister



Slide Credit: Nister



UK



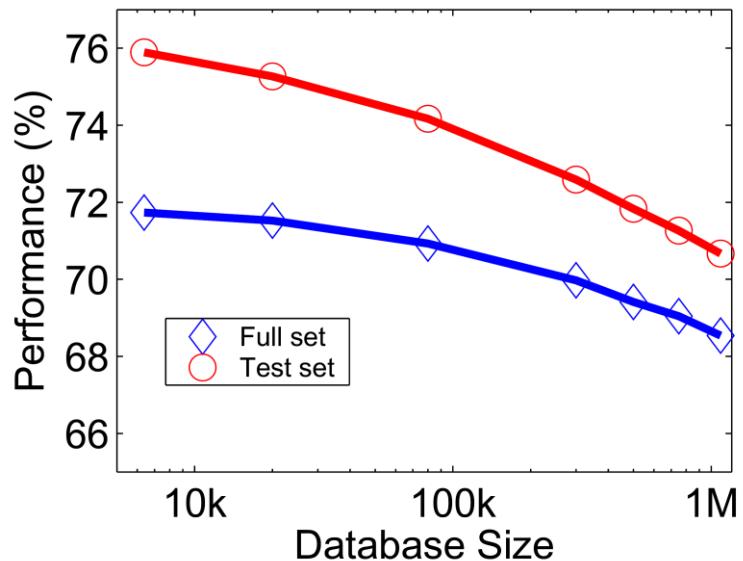
Center for  
Visualization & Virtual  
Environments

Slide Credit: Nister



Slide Credit: Nister

# Performance



## ImageSearch at the VizCentre

New query:  Browse... Send File

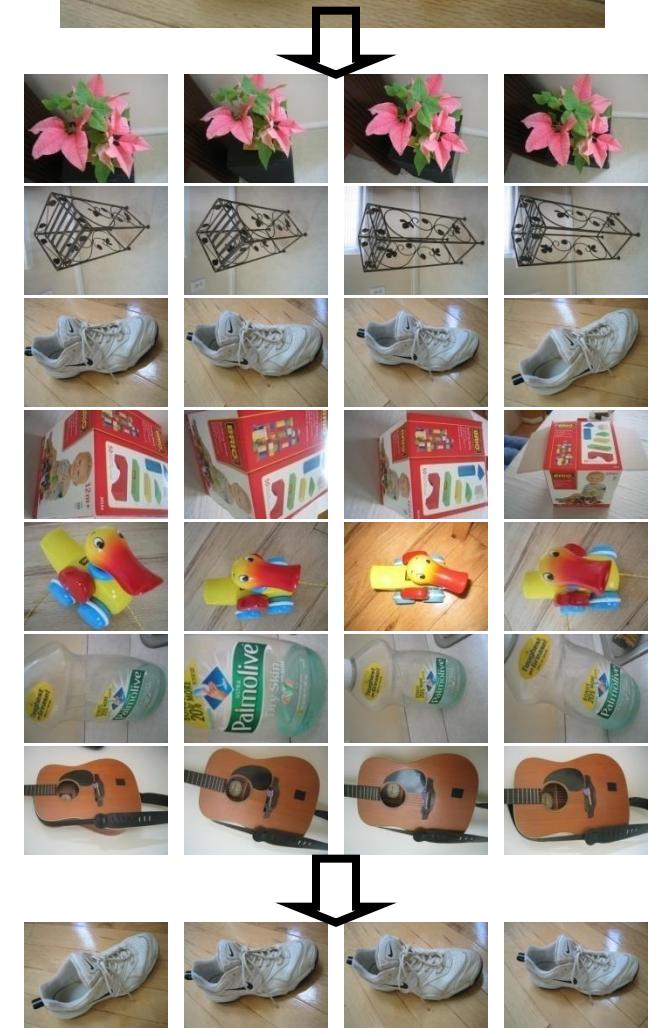
File is 500x320



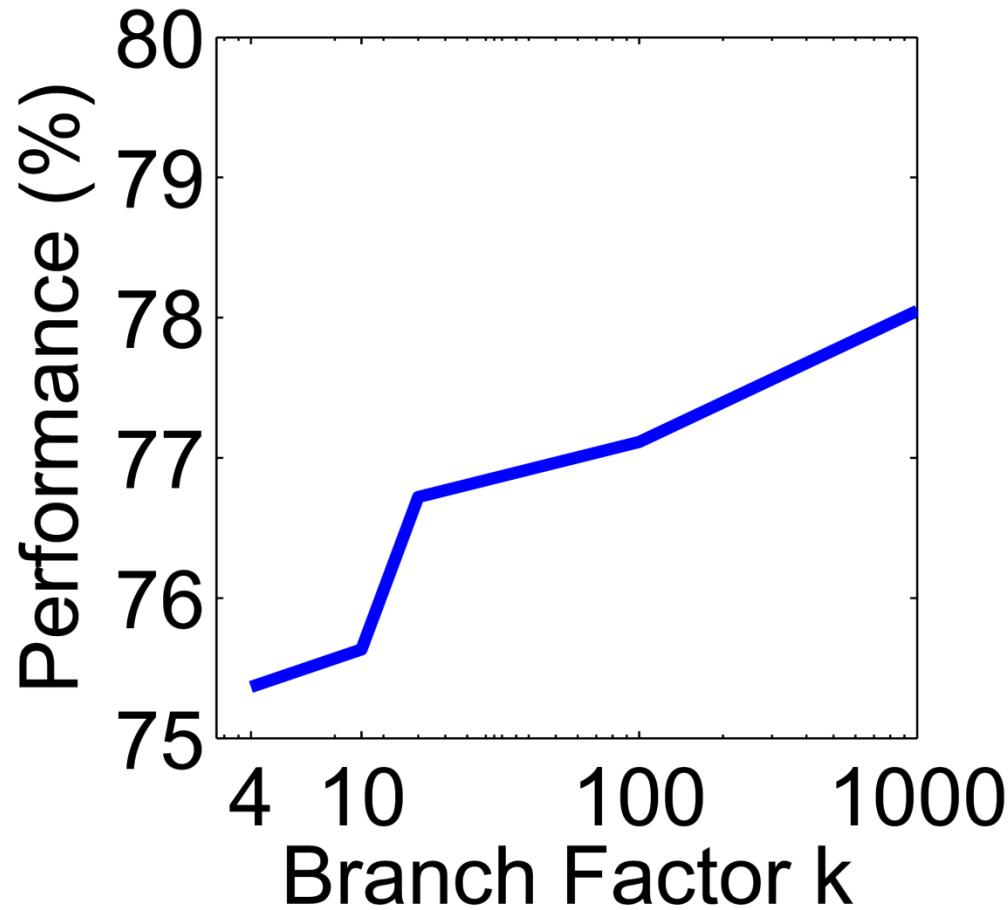
Top n results of your query.



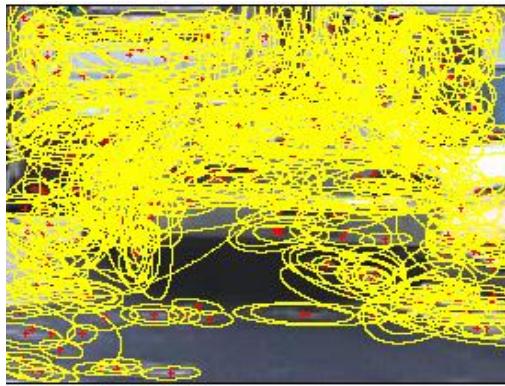
bourne/im1000043322.pgm bourne/im1000043323.pgm bourne/im1000043326.pgm bourne/im1000043327.pgm



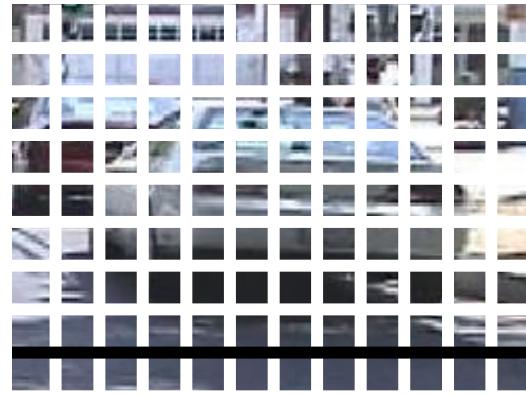
Higher branch factor works better  
(but slower)



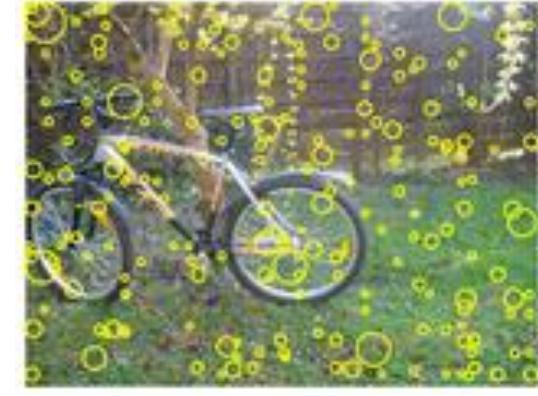
# Sampling strategies



Sparse, at interest points



Dense, uniformly



Randomly



Multiple interest operators

- To find specific, textured objects, sparse sampling from interest points often more reliable.
- Multiple complementary interest operators offer more image coverage.
- For object categorization, dense sampling offers better coverage.

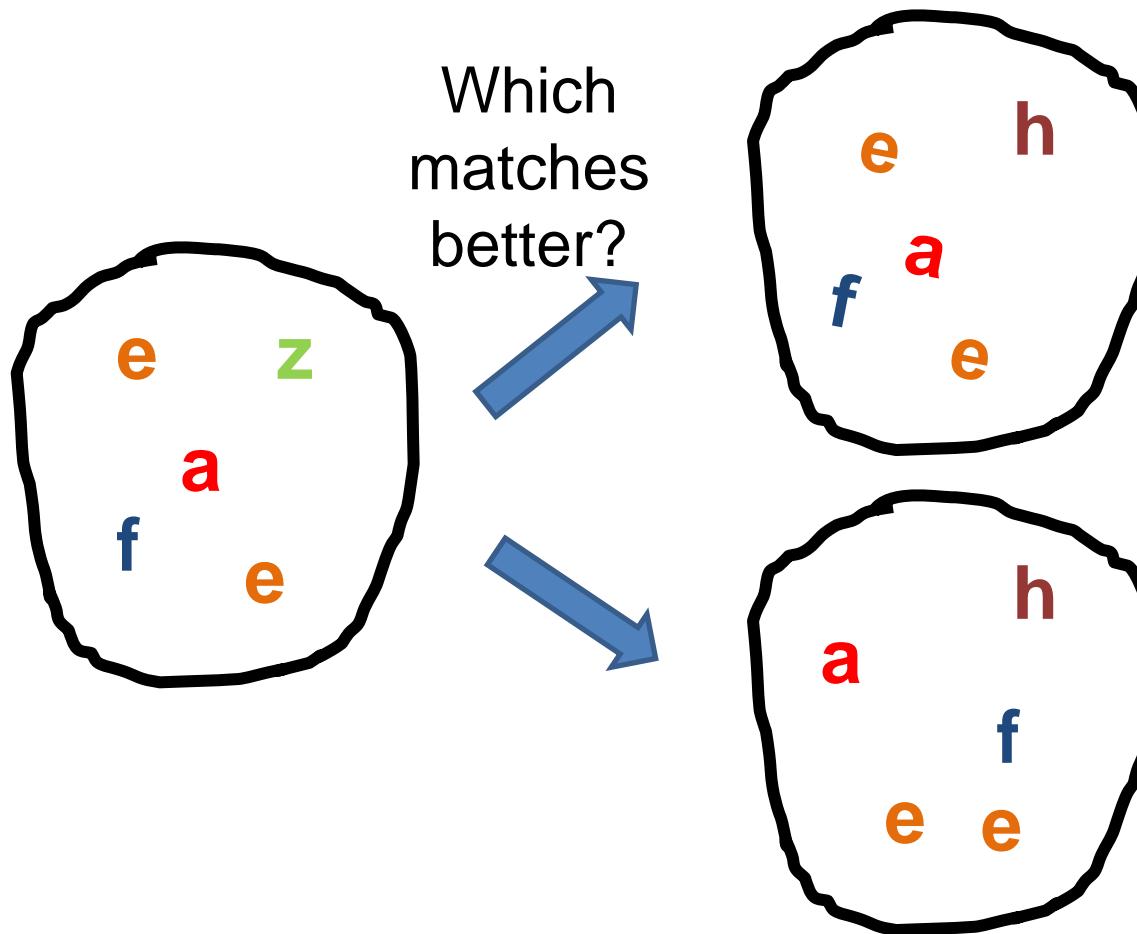
[See Nowak, Jurie & Triggs, ECCV 2006]

# Instance recognition: remaining issues

- How to summarize the content of an entire image? And gauge overall similarity?
- How large should the vocabulary be? How to perform quantization efficiently?
- Is having the same set of visual words enough to identify the object/scene? How to verify spatial agreement?
- How to score the retrieval results?

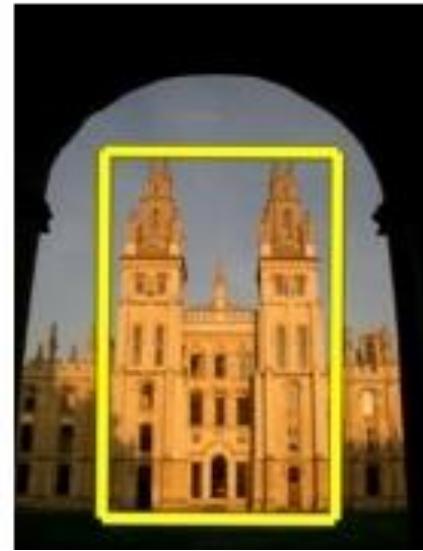
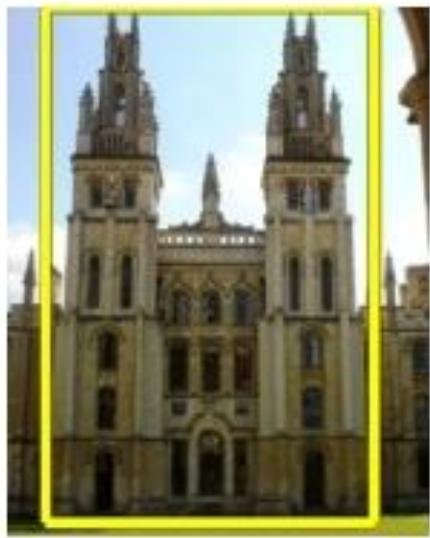
# Can we be more accurate?

So far, we treat each image as containing a “bag of words”, with no spatial information



# Can we be more accurate?

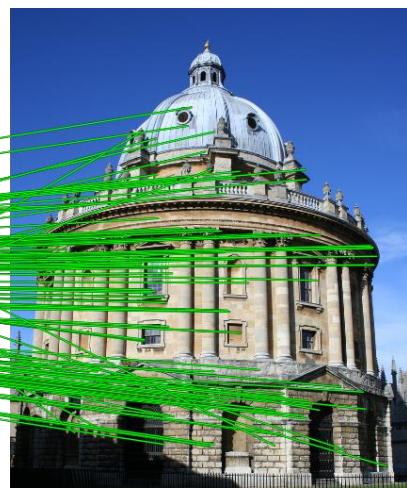
So far, we treat each image as containing a “bag of words”, with no spatial information



Real objects have consistent geometry

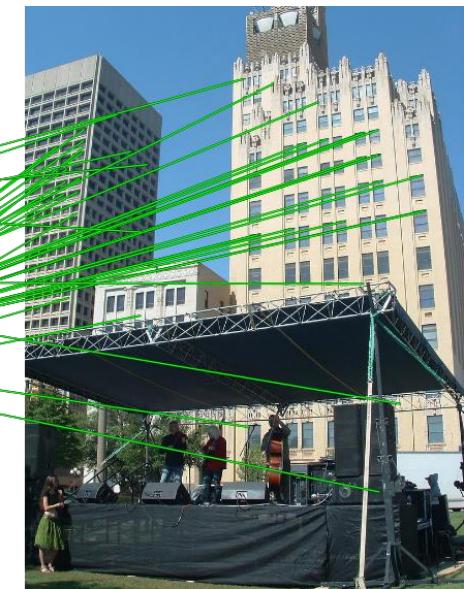
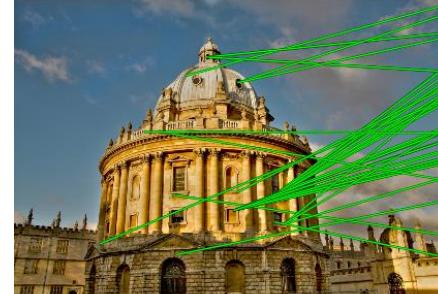
# Spatial Verification

Query



DB image with high BoW  
similarity

Query

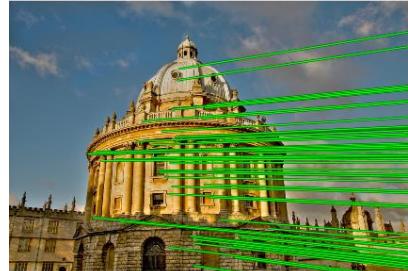


DB image with high BoW  
similarity

Both image pairs have many visual words in common.

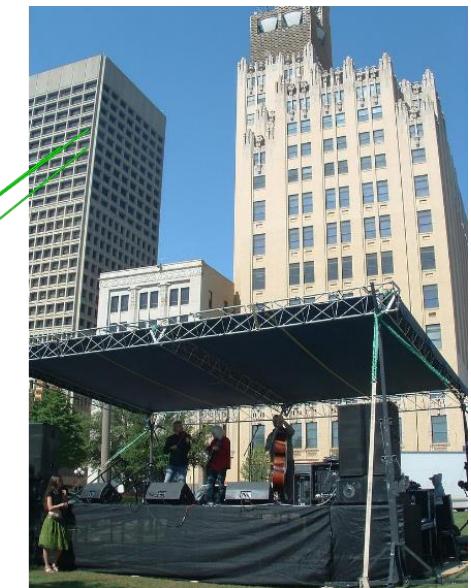
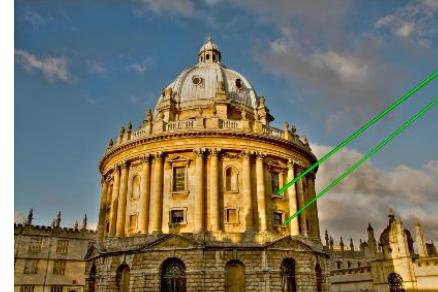
# Spatial Verification

Query



DB image with high BoW  
similarity

Query



DB image with high BoW  
similarity

Only some of the matches are mutually consistent

# Instance recognition: remaining issues

- How to summarize the content of an entire image? And gauge overall similarity?
- How large should the vocabulary be? How to perform quantization efficiently?
- Is having the same set of visual words enough to identify the object/scene? How to verify spatial agreement?
- How to score the retrieval results?

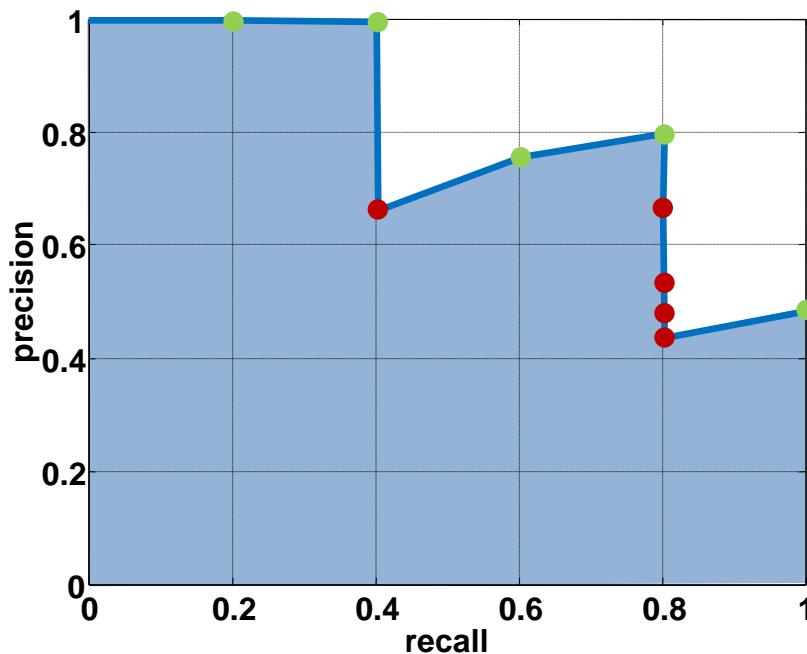
# Scoring retrieval quality



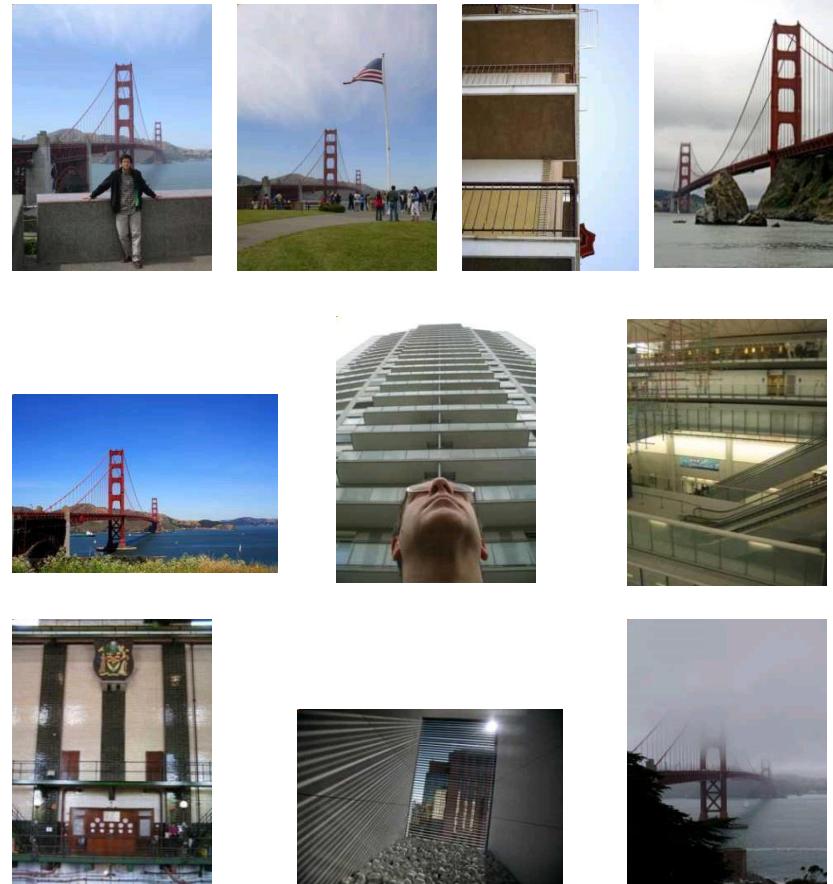
Query

Database size: 10 images  
Relevant (total): 5 images

$$\text{precision} = \frac{\text{#relevant}}{\text{#returned}}$$
$$\text{recall} = \frac{\text{#relevant}}{\text{#total relevant}}$$



Results (ordered):



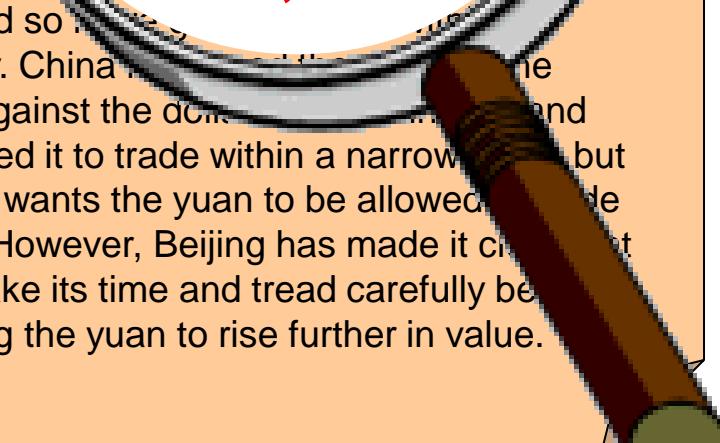
# What else can we borrow from text retrieval?

## Index

"Along I-75," From Detroit to Florida; <i>inside back cover</i>	Butterfly Center, McGuire; 134
"Drive I-95," From Boston to Florida; <i>inside back cover</i>	CAA (see AAA)
1929 Spanish Trail Roadway;	CCC, The; 111,113,115,135,142
101-102,104	Ca d'Zan; 147
511 Traffic Information; 83	Caloosahatchee River; 152
A1A (Barrier Isl) - I-95 Access; 86	Name; 150
AAA (and CAA); 83	Canaveral Natnl Seashore; 173
AAA National Office; 88	Cannon Creek Airpark; 130
Abbreviations,	Canopy Road; 106,169
Colored 25 mile Maps; cover	Cape Canaveral; 174
Exit Services; 196	Castillo San Marcos; 169
Travelogue; 85	Cave Diving; 131
Africa; 177	Cayo Costa, Name; 150
Agricultural Inspection Stns; 126	Celebration; 93
Ah-Tah-Thi-Ki Museum; 160	Charlotte County; 149
Air Conditioning, First; 112	Charlotte Harbor; 150
Alabama; 124	Chautauqua; 116
Alachua; 132	Chipley; 114
County; 131	Name; 115
Alafia River; 143	Choctawhatchee, Name; 115
Alapaha, Name; 126	Circus Museum, Ringling; 147
Alfred B MacIay Gardens; 106	Citrus; 88,97,130,136,140,180
Alligator Alley; 154-155	CityPlace, W Palm Beach; 180
Alligator Farm, St Augustine; 169	City Maps,
Alligator Hole (definition); 157	Ft Lauderdale Expwys; 194-195
Alligator, Buddy; 155	Jacksonville; 163
Alligators; 100,135,138,147,156	Kissimmee Expwys; 192-193
Anastasia Island; 170	Miami Expressways; 194-195
Anhinga; 108-109,146	Orlando Expressways; 192-193
Apalachicola River; 112	Pensacola; 26
Appleton Mus of Art; 136	Tallahassee; 191
Aquifer; 102	Tampa-St. Petersburg; 63
Arabian Nights; 94	St. Augustine; 191
Art Museum, Ringling; 147	Civil War; 100,108,127,138,141
Aruba Beach Cafe; 183	Clearwater Marine Aquarium; 187
Aucilla River Project; 106	Collier County; 154
Babcock-Web WMA; 151	Collier, Barron; 152
Bahia Mar Marina; 184	Colonial Spanish Quarters; 168
Baker County; 99	Columbia County; 101,128
Barefoot Mailmen; 182	Coquina Building Material; 165
Barge Canal; 137	Corkscrew Swamp, Name; 154
Bee Line Expy; 80	Cowboys; 95
Belz Outlet Mall; 89	Crab Trap II; 144
Bernard Castro; 136	Cracker, Florida; 88,95,132
Big "I"; 165	Crossstown Expy; 11,35,98,143
Big Cypress; 155,158	Cuban Bread; 184
Big Foot Monster; 105	Dade Battlefield; 140
Big Pine Key; 184	Dade, Maj, Francis; 139-140,161
Big Pine Key; 184	Dania Beach Hurricane; 184
Big Pine Key; 184	Spur SR91; 76
Big Pine Key; 184	Spur SR91; 76

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% rise in exports to \$750bn, compared with \$660bn. The US is annoyed that China's exports are increasing so rapidly, and has deliberately agreed to increase its imports of Chinese goods. The Chinese government also needs to encourage domestic demand so that it can meet the growing demand in the country. China's central bank has been allowed to let the yuan against the dollar rise, but it has not yet permitted it to trade within a narrow range. The US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

**China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value**



# *tf-idf* weighting

- Term frequency – inverse document frequency
- Describe frame by frequency of each word within it, downweight words that appear often in the database
- (Standard weighting for text retrieval)

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

Number of occurrences of word i in document d

Number of words in document d

Total number of documents in database

Number of documents word i occurs in, in whole database

# Query expansion

Query: *golf green*

Results:

- How can the grass on the *greens* at a *golf* course be so perfect?
- For example, a skilled *golfer* expects to reach the *green* on a par-four hole in ...
- Manufactures and sells synthetic *golf* putting *greens* and mats.

Irrelevant result can cause a 'topic drift':

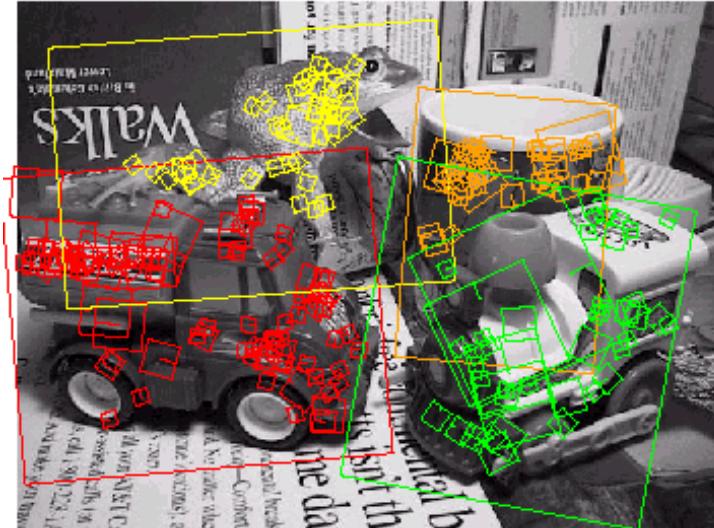
- Volkswagen *Golf*, 1999, *Green*, 2000cc, petrol, manual, , hatchback, 94000miles, 2.0 GTi, 2 Registered Keepers, HPI Checked, Air-Conditioning, Front and Rear Parking Sensors, ABS, Alarm, Alloy

# Query Expansion



# Things to remember

- Object instance recognition
  - Find keypoints, compute descriptors
  - Match descriptors
  - Vote for / fit affine parameters
  - Return object if # inliers > T



- Keys to efficiency
  - Visual words
    - Used for many applications
  - Inverse document file
    - Used for web-scale search

