

Computer Vision

Instance Recognition

Matching local features



Matching local features

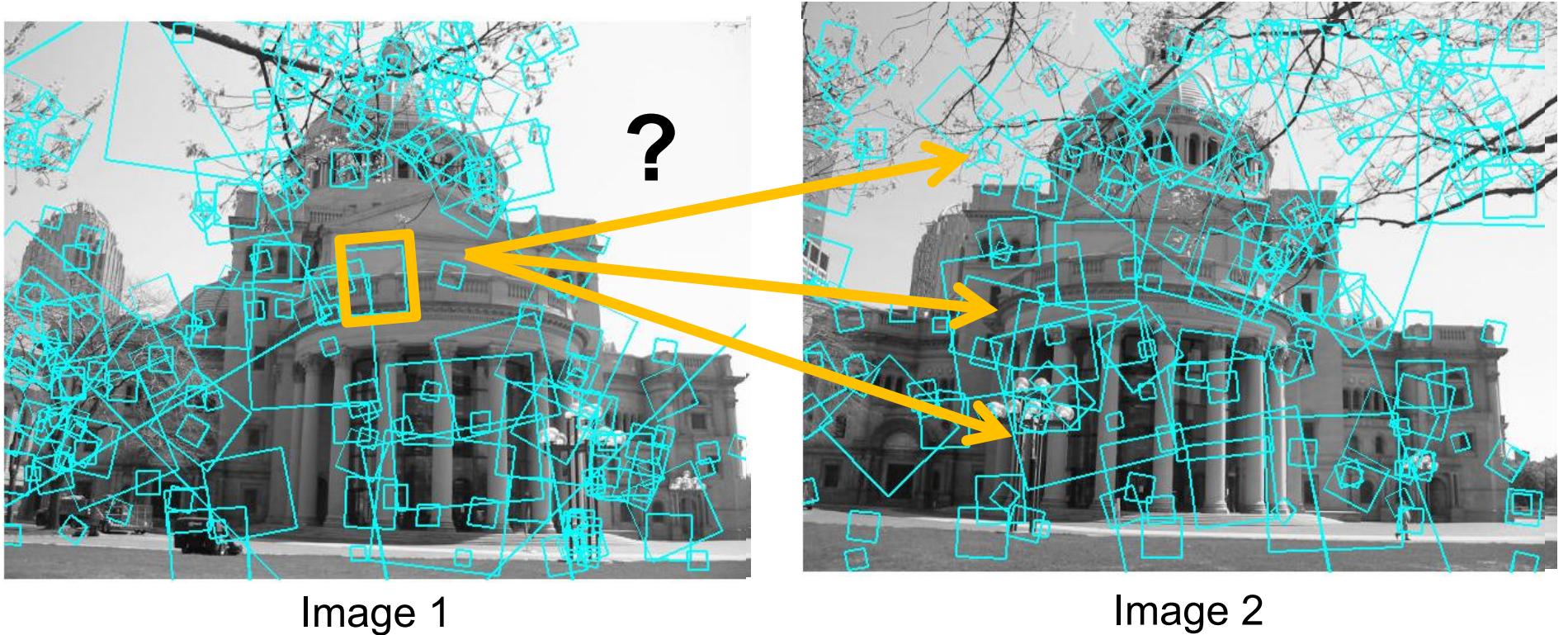


Image 1

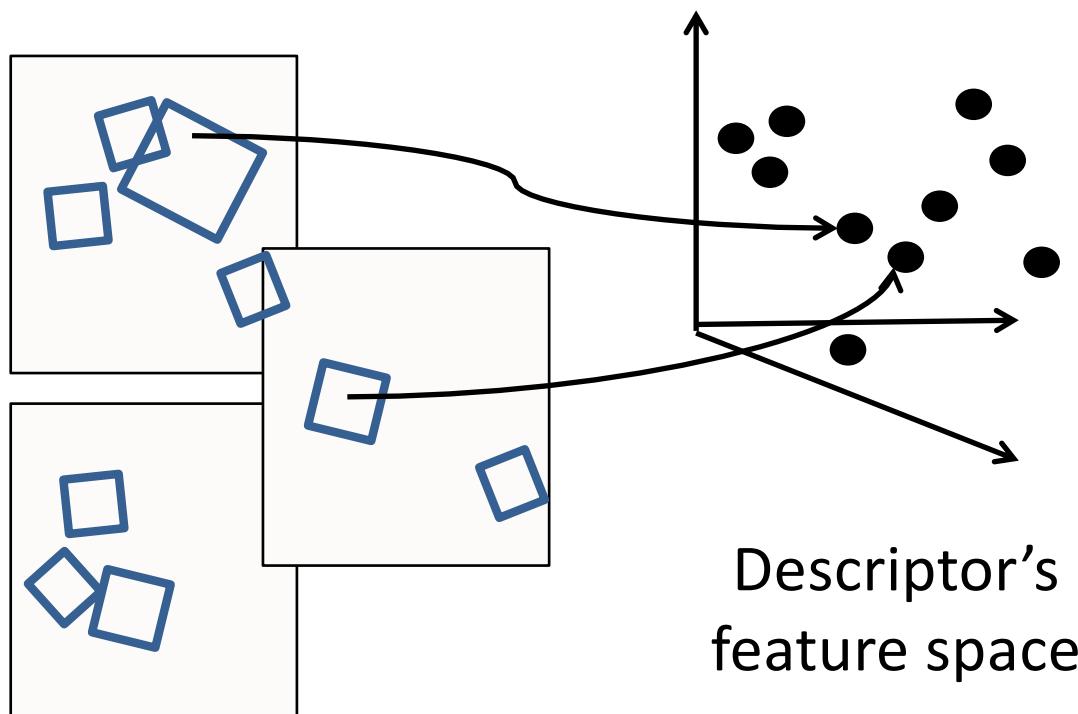
Image 2

To generate candidate matches, find patches that have the most similar appearance (e.g., lowest SSD)

Simplest approach: compare them all, take the closest (or closest k , or within a thresholded distance)

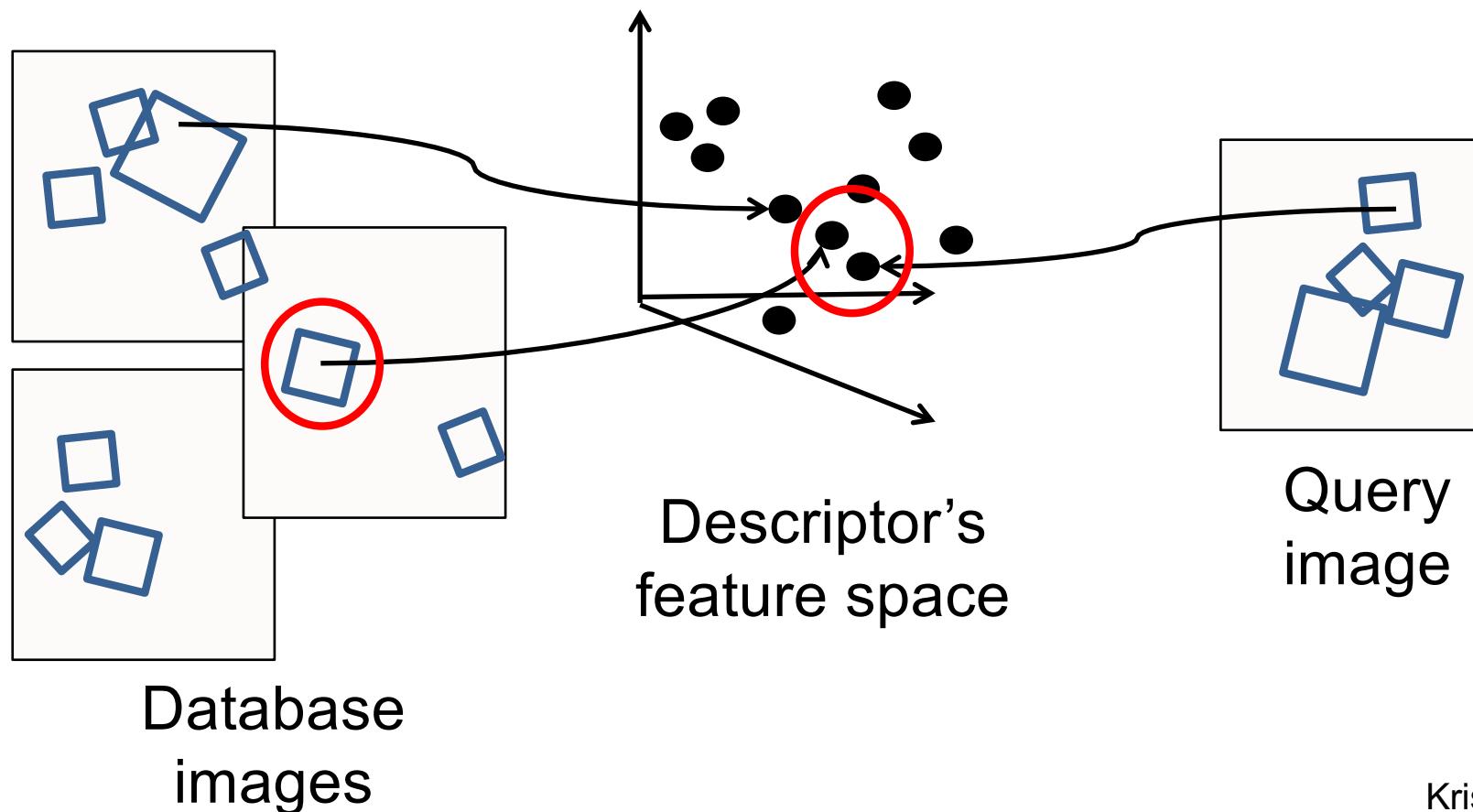
Indexing local features

- 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.

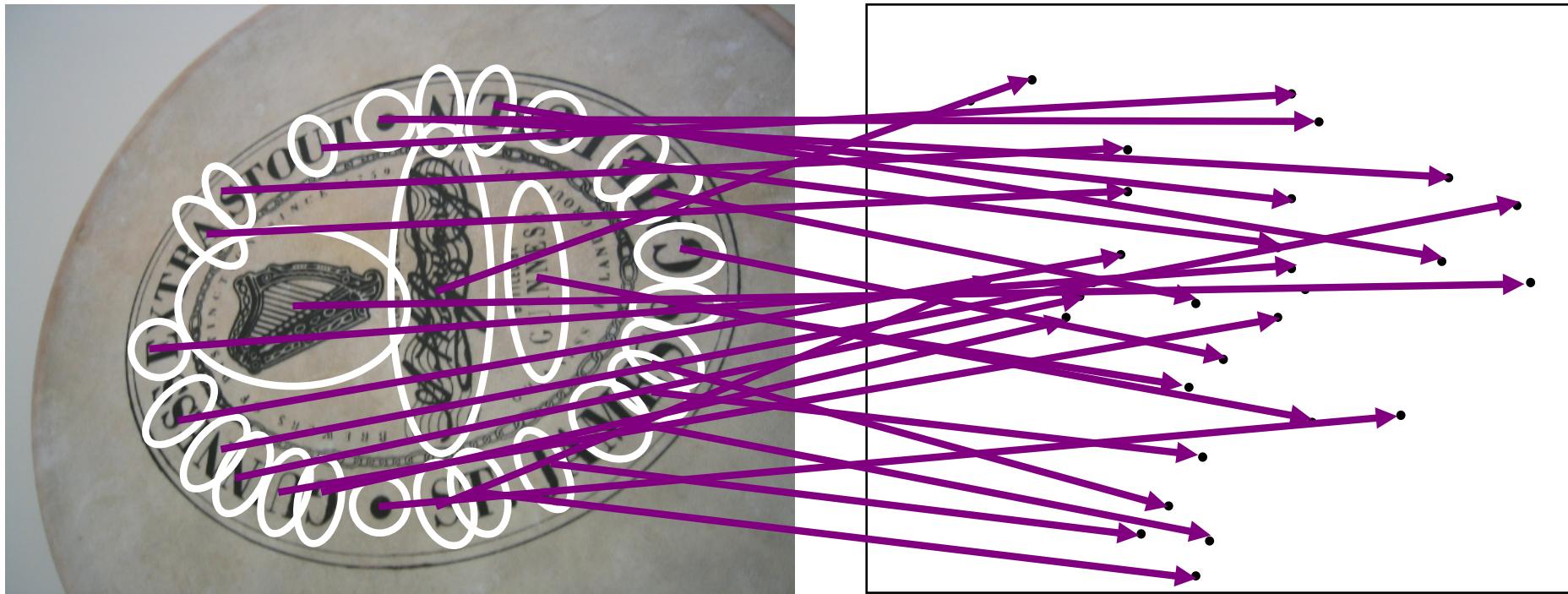


Indexing local features

- With potentially thousands of features per image, and hundreds to millions of images to search, how to efficiently find those that are relevant to a new image?

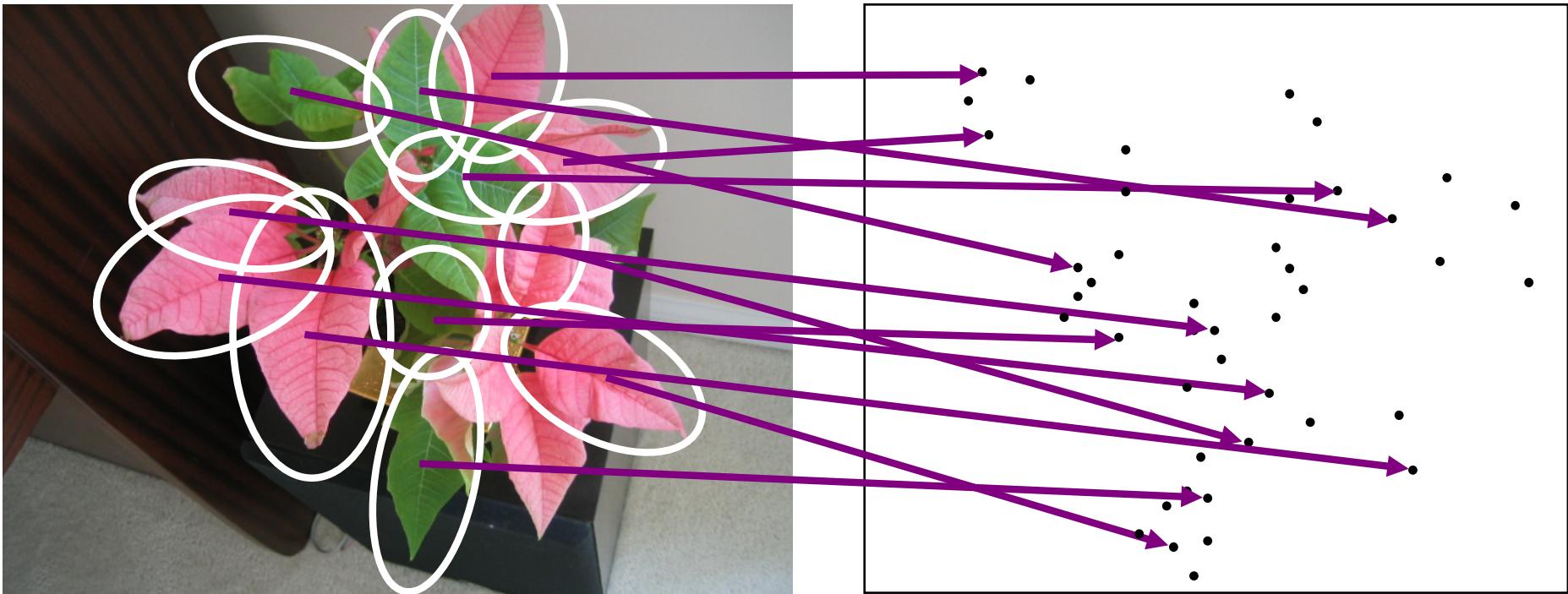
Visual words: main idea

Extract some local features from a number of images ...

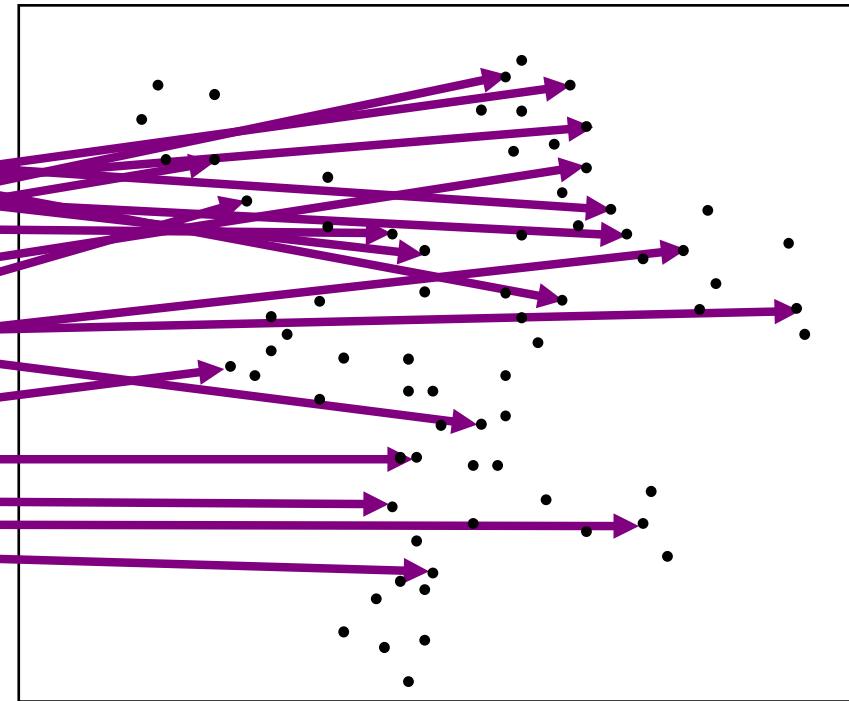
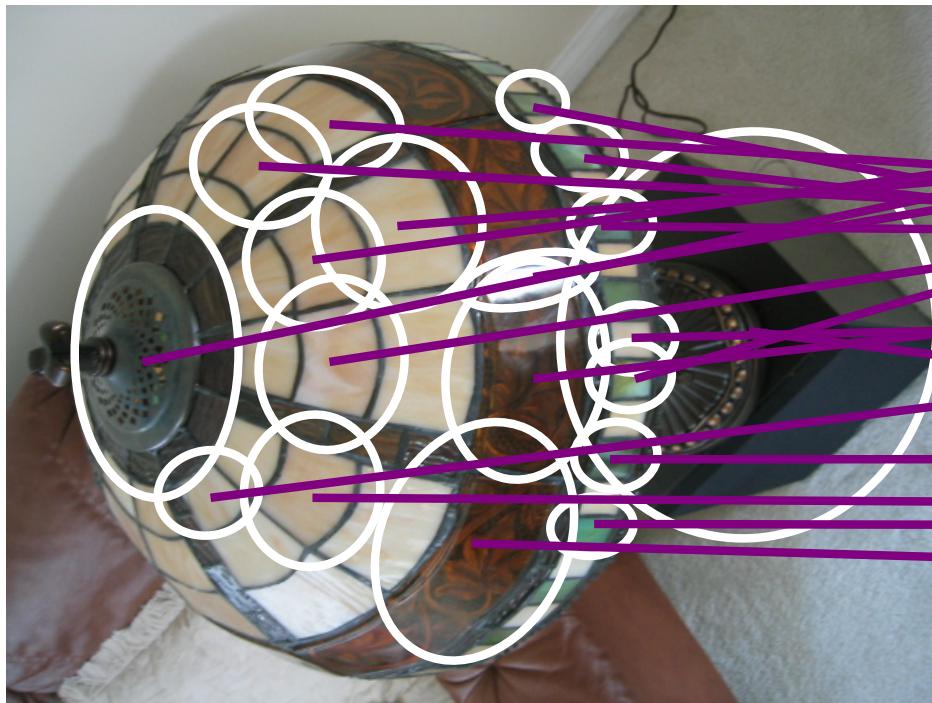


e.g., SIFT descriptor space: each point is 128-dimensional

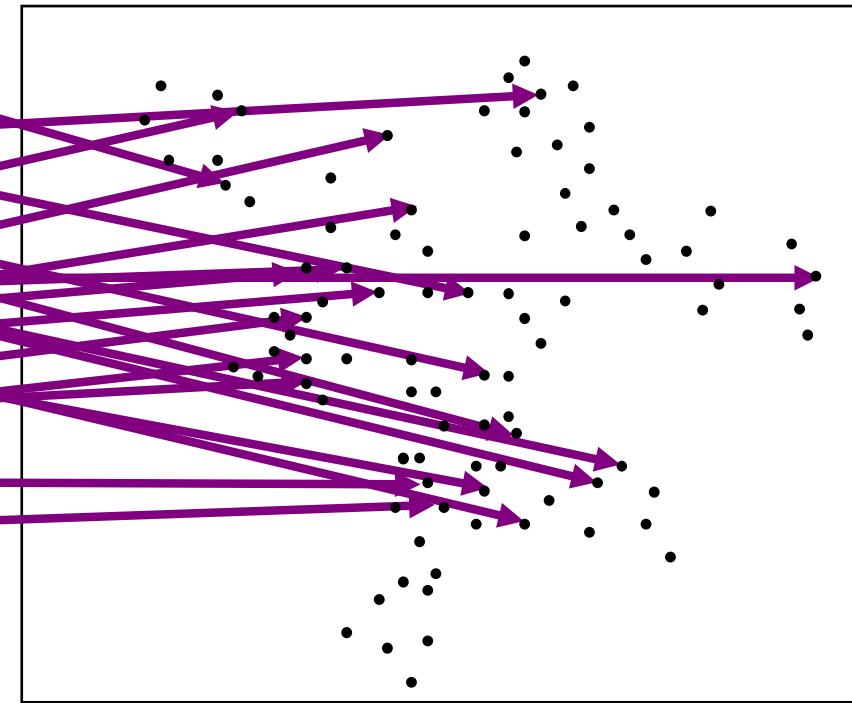
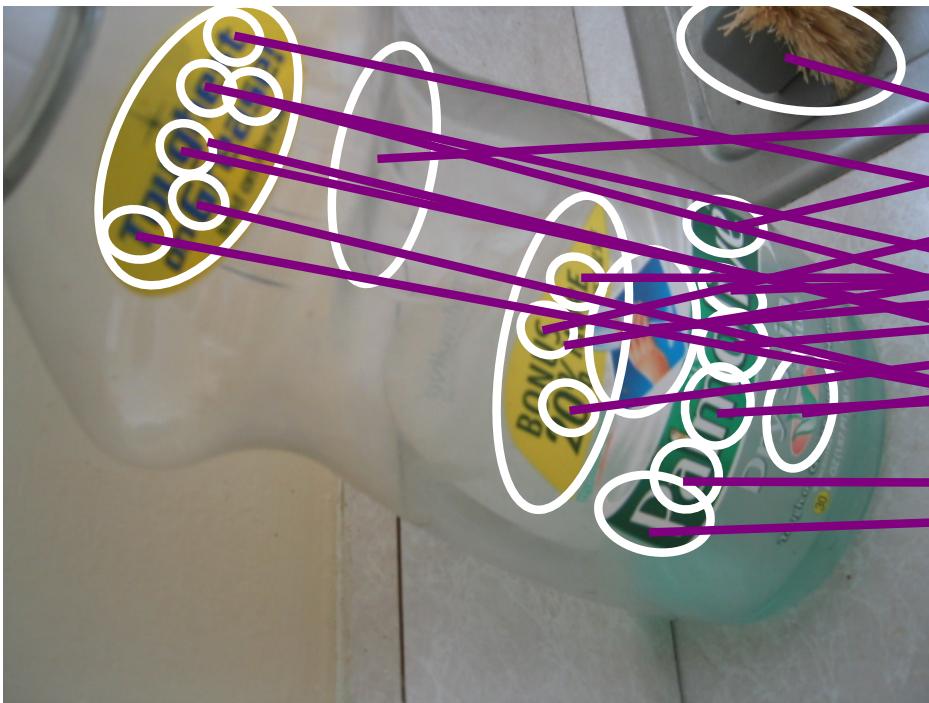
Visual words: main idea



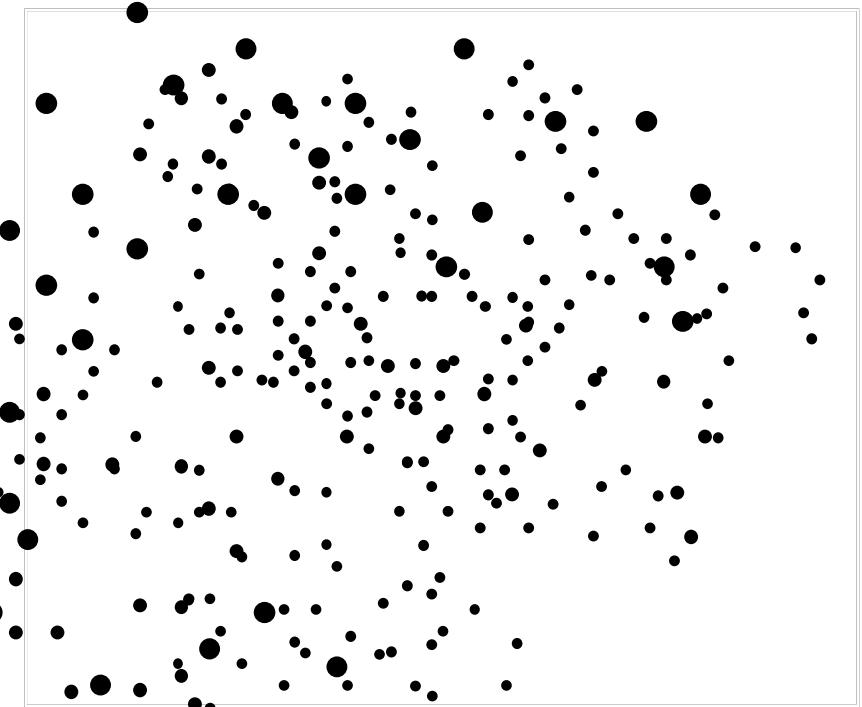
Visual words: main idea

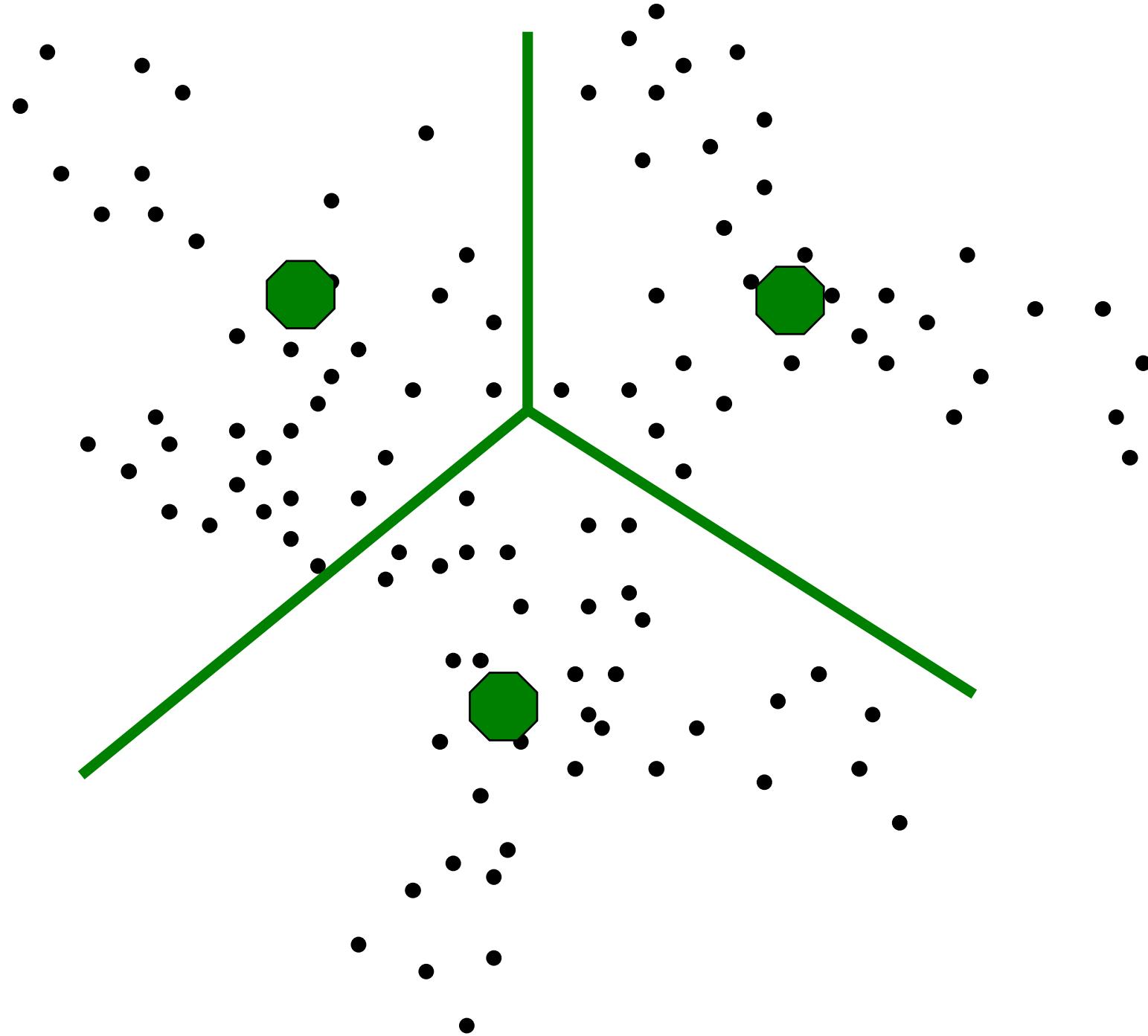


Visual words: main idea



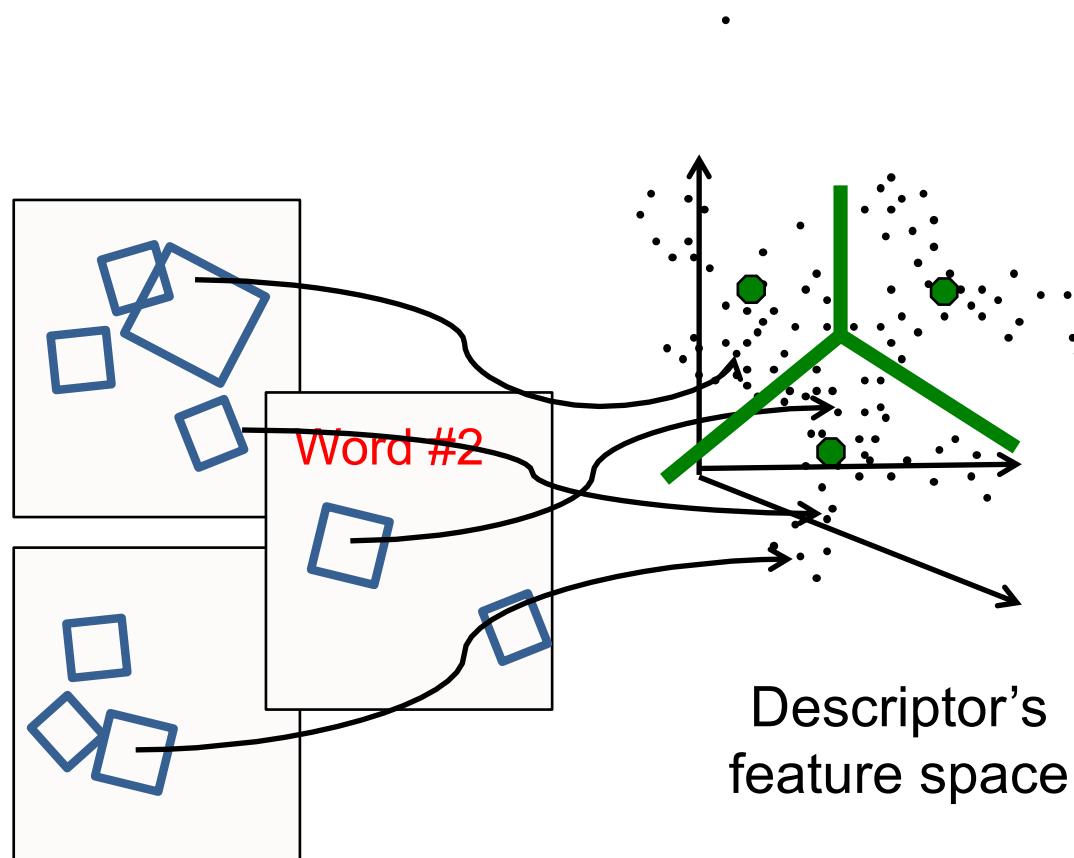
Each point is a local descriptor, e.g. SIFT vector.





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

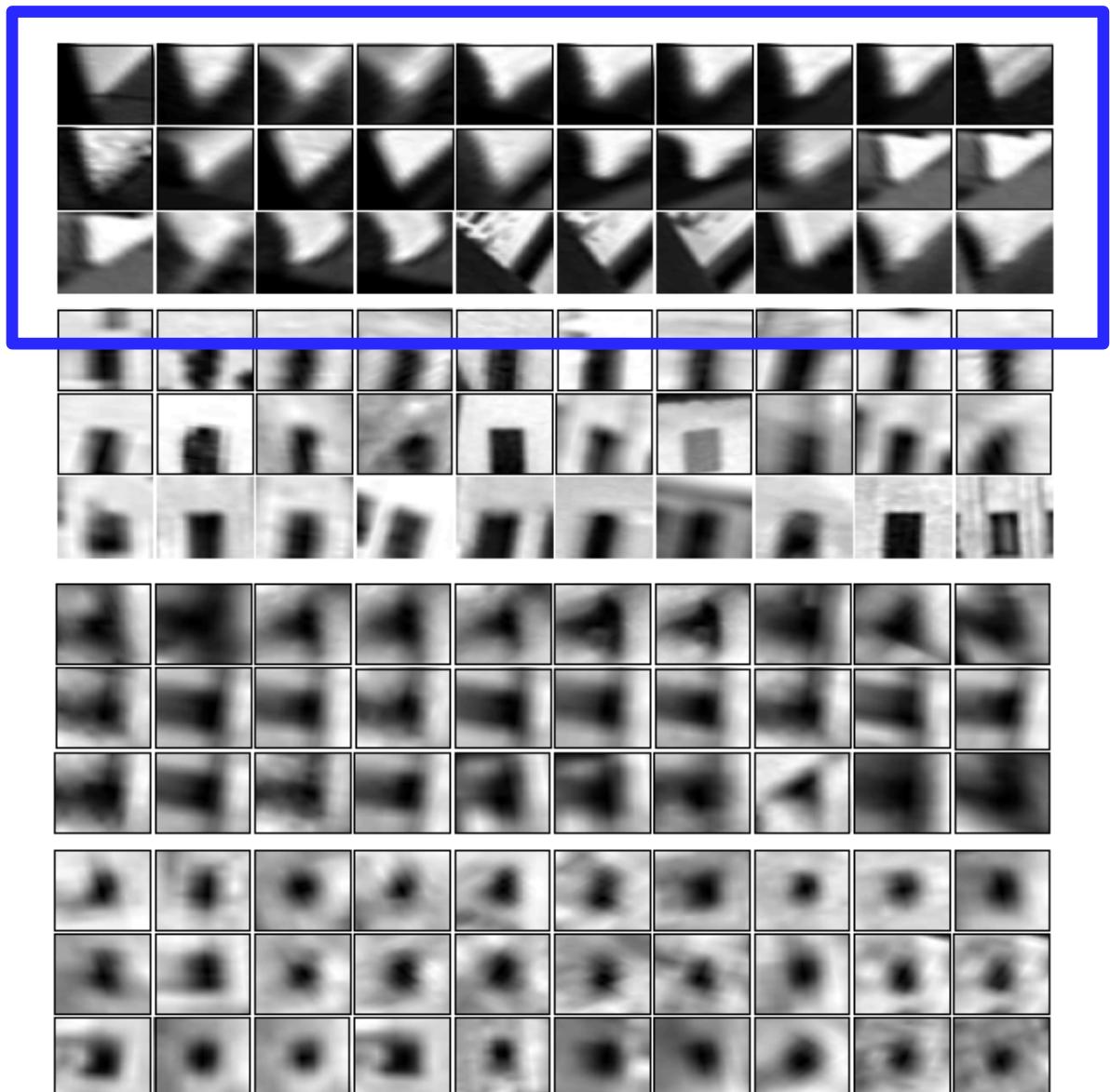
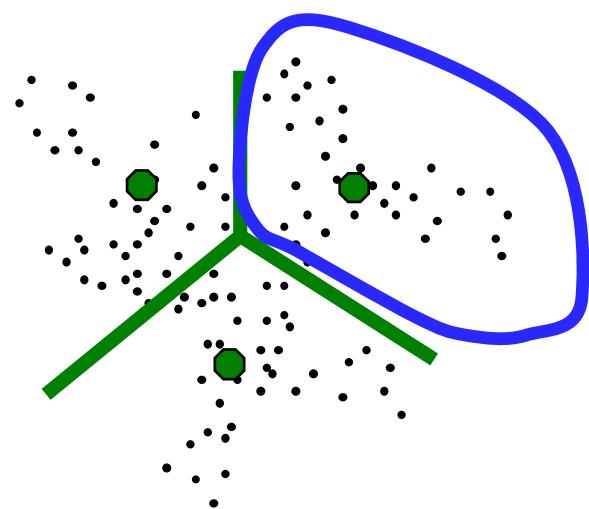


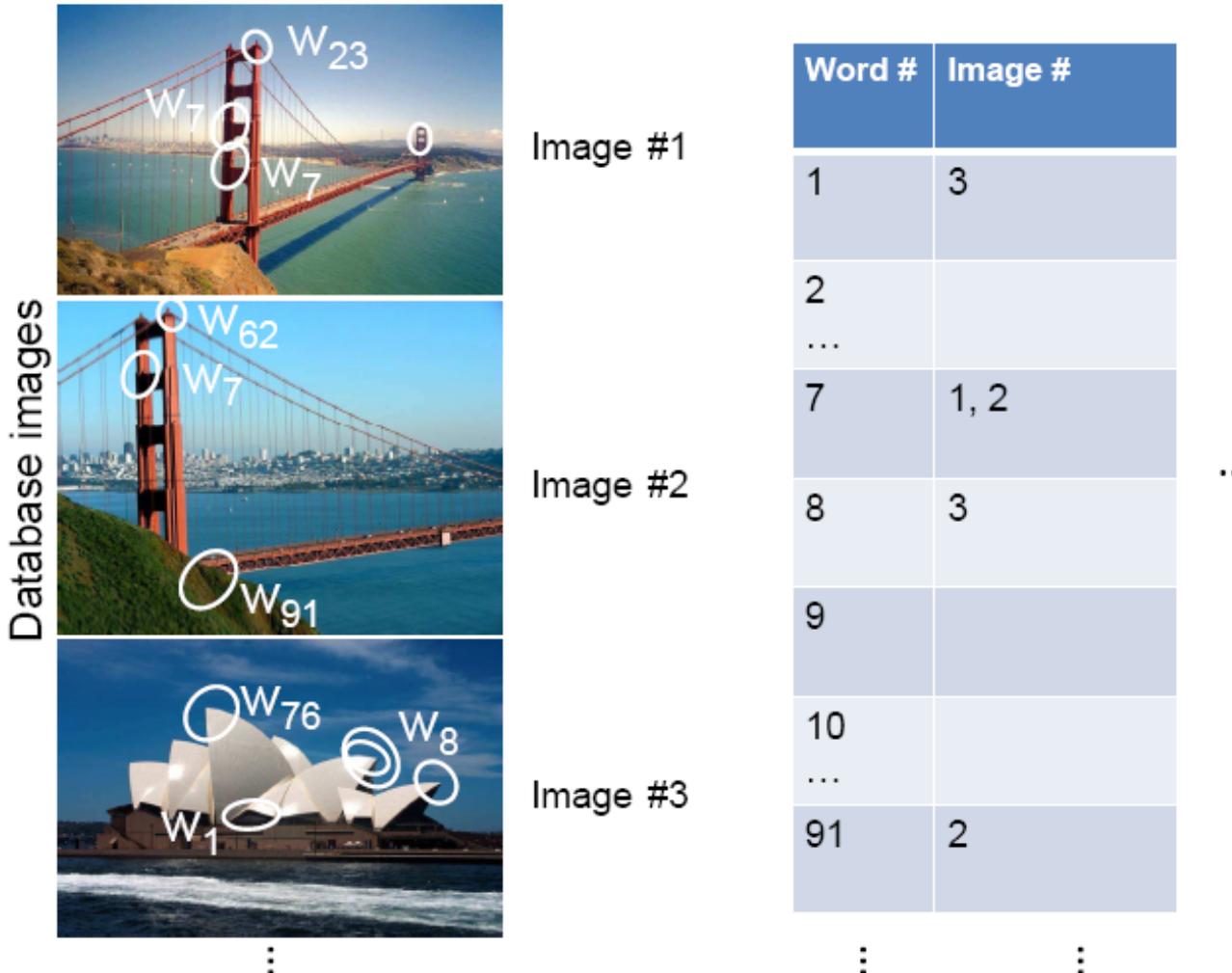
Figure from Sivic & Zisserman, ICCV 2003

Indexing local features

| 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 Name; 150 |
| AAA (and CAA); 83 | Canaveral Natnl Seashore; 173 |
| AAA National Office; 88 | Cannon Creek Airpark; 130 |
| Abbreviations, Colored 25 mile Maps; cover | Canopy Road; 106,169 |
| 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; 180 | Celebration; 93 |
| Air Conditioning, First; 112 | Charlotte County; 149 |
| Alabama; 124 | Charlotte Harbor; 150 |
| Alachua; 132 County; 131 | Chautauqua; 116 |
| Alafia River; 143 | Chipley; 114 Name; 115 |
| Alapaha, Name; 126 | Choctawhatchee, Name; 115 |
| Alfred B Maclay 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, 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 |
| Anhailca; 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 | Colonial Spanish Quarters; 168 |
| Bahia Mar Marina; 184 | Columbia County; 101,128 |
| Baker County; 99 | Coquina Building Material; 165 |
| Barefoot Mallmen; 182 | Corkscrew Swamp, Name; 154 |
| Barge Canal; 137 | Cowboys; 95 |
| Bee Line Expy; 80 | Crab Trap II; 144 |
| <small>Books Printed 1998-2000</small> | Cracker, Florida; 88,95,132 |
| | <small>Copyright 1998-2000, 2002</small> |
| | 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; 168 |
| | 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 |
| | NETT; 78,104,106 |

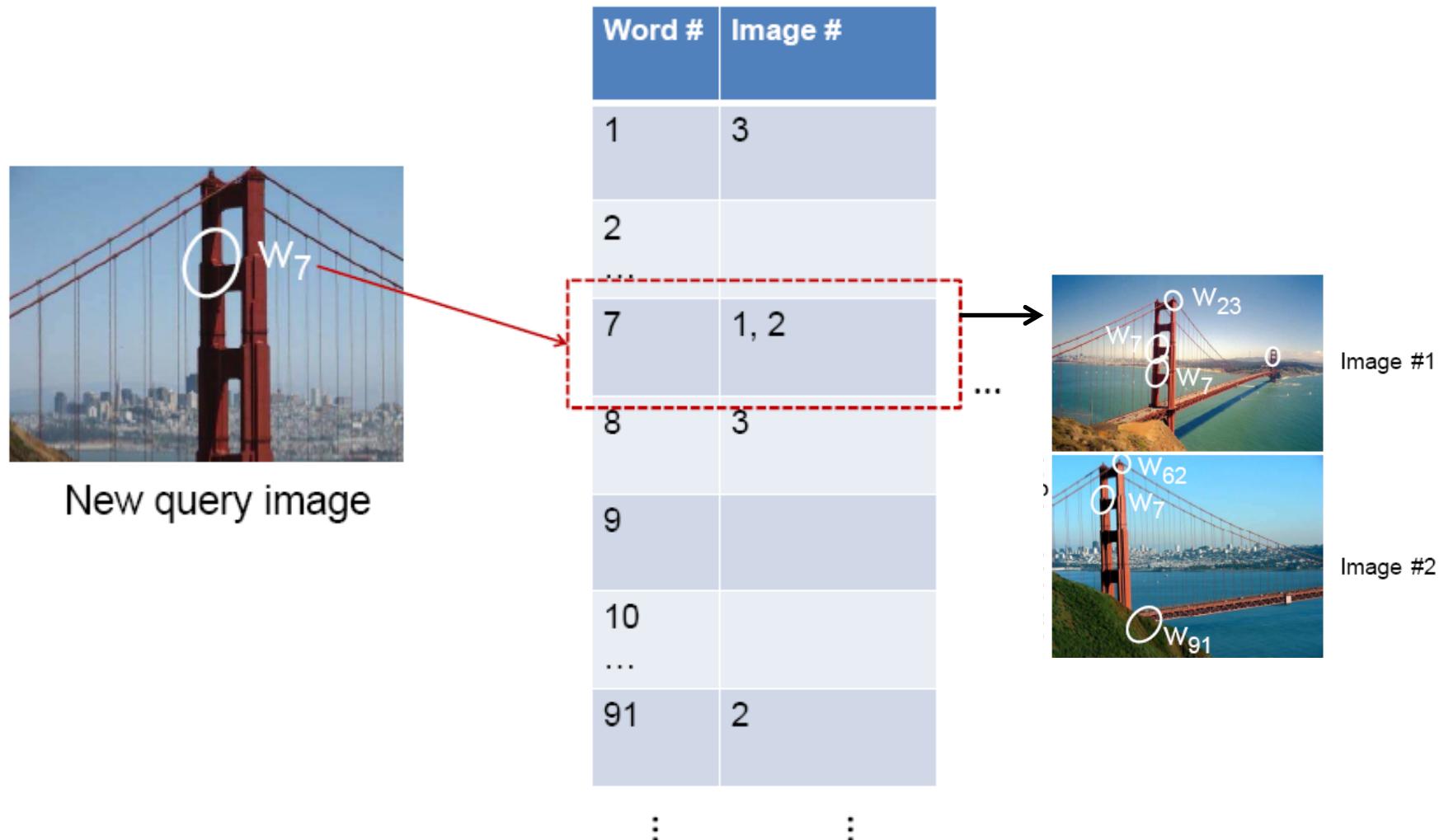
- **Inverted file index**
- For text documents, an efficient way to find all *pages* on which a *word* occurs is to use an index...
- We want to find all *images* in which a *feature* occurs.
- To use this idea, we'll need to map our features to “visual words”.

Inverted file index



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

Inverted file index



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?

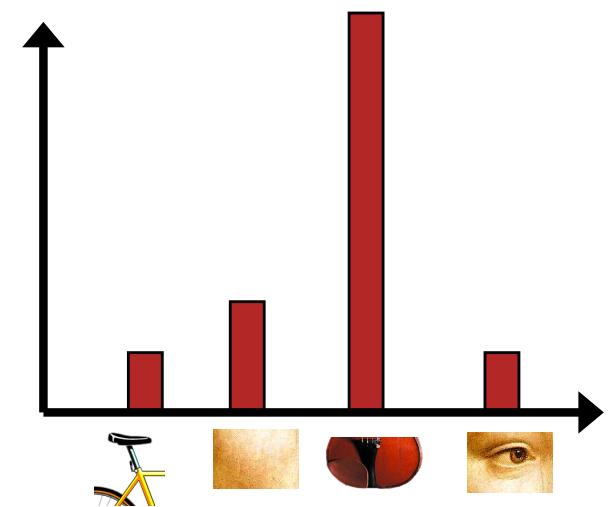
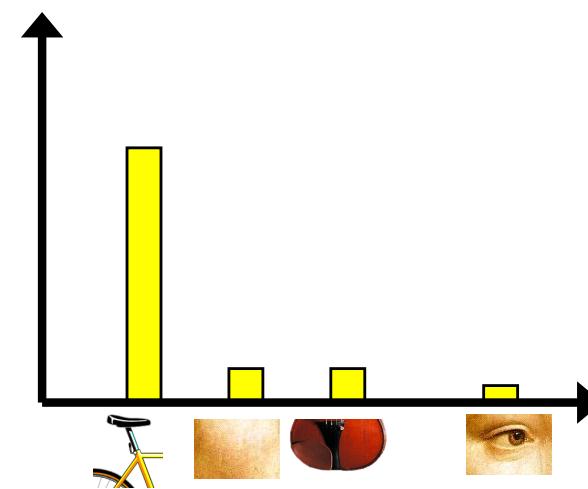
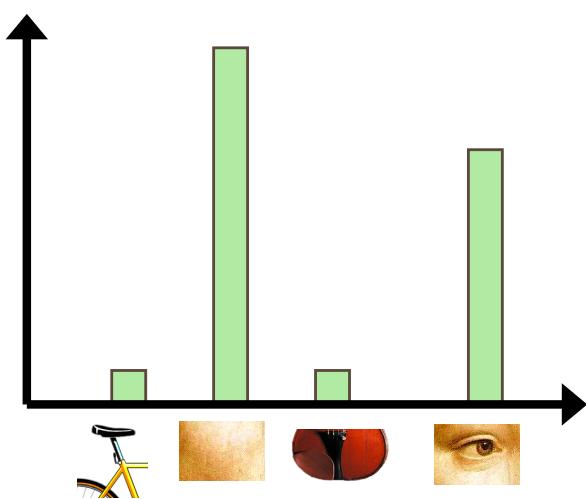
Analogy to documents

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 our eyes send to the brain from our eyes. For example, it has been thought that the retina sends messages directly to the brain; this was first suggested by the point by point analysis of the visual system by Hubel and Wiesel. In the 1960s, Hubel and Wiesel discovered that the visual system does not know the whole story. They found that the perception of a visual scene depends considerably on the context of other visual events. By following the visual pathways along their path through the brain to the layers of the optical cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a step-wise analysis by a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.

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

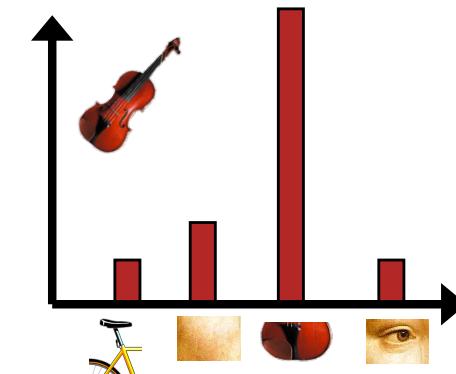
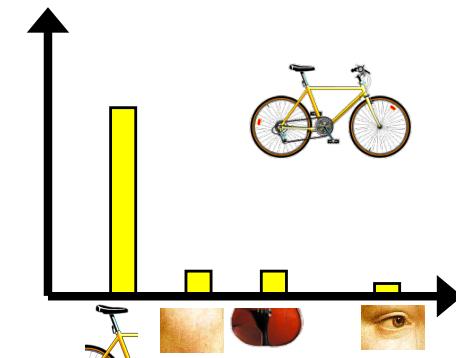
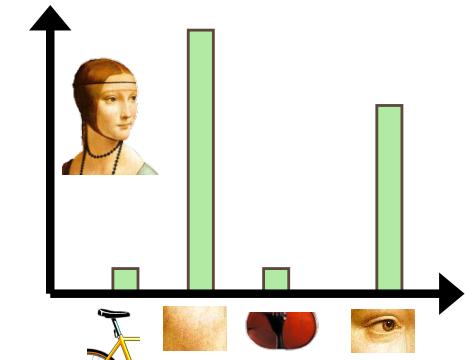
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a projected 30% jump in exports to the US, which would be a 18% rise in value. China's foreign reserves are likely to grow further as the country has long complained of unfair trading practices by the US. The Chinese government is under pressure to allow the yuan to appreciate, but it has been holding the surplus at only one level. Zhou Xiaochuan, governor of the central bank, said that the yuan needed to appreciate to meet the growing demand so many people have for it in the country. China increased the value of the yuan against the dollar by 2.1% in March 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



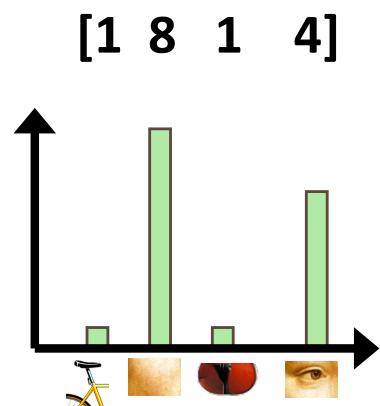
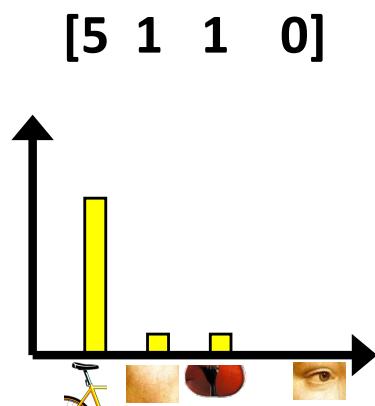
Bags of visual words

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



Bags of visual words

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

 \vec{d}_j  \vec{q}

$$\begin{aligned} sim(d_j, q) &= \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|} \\ &= \frac{\sum_{i=1}^V d_j(i) * q(i)}{\sqrt{\sum_{i=1}^V d_j(i)^2} * \sqrt{\sum_{i=1}^V q(i)^2}} \end{aligned}$$

for vocabulary of V words

What else can we borrow from text retrieval?

Index

"Along I-75," From Detroit to Florida; *inside back cover*
"Drive I-95," From Boston to Florida; *inside back cover*
1929 Spanish Trail Roadway; 101-102,104
511 Traffic Information; 83
A1A (Barrier Isl) - I-95 Access; 86
AAA (and CAA); 83
AAA National Office; 88
Abbreviations.
 Colored 25 mile Maps; cover
 Exit Services; 196
 Travelogue; 85
Africa; 177
Agricultural Inspection Stns; 126
Ah-Tah-Thi-Ki Museum; 160
Air Conditioning, First; 112
Alabama; 124
Alachua; 132
 County; 131
Alafia River; 143
Alapaha, Name; 126
Alfred B MacIay Gardens; 106
Alligator Alley; 154-155
Alligator Farm, St Augustine; 169
Alligator Hole (definition); 157
Alligator, Buddy; 155
Alligators; 100,135,138,147,156
Anastasia Island; 170
Anhaica; 108-109,146
Apalachicola River; 112
Appleton Mus of Art; 136
Aquifer; 102
Arabian Nights; 94
Art Museum, Ringling; 147
Aruba Beach Cafe; 183
Aucilla River Project; 106
Babcock-Web WMA; 151
Bahia Mar Marina; 184
Baker County; 99
Barefoot Mailmen; 182
Barge Canal; 137
Bee Line Expy; 80
Belz Outlet Mall; 89
Bernard Castro; 136
Big "I"; 165
Big Cypress; 155,158
Big Foot Monster; 105
Billie Jean King; 100
Butterfly Center, McGuire; 134
CAA (see AAA)
CCC, The; 111,113,115,135,142
Ca d'Zan; 147
Caloosahatchee River; 152
 Name; 150
Canaveral Natnl Seashore; 173
Cannon Creek Airpark; 130
Canopy Road; 106,160
Cape Canaveral; 174
Castillo San Marcos; 169
Cave Diving; 131
Cayo Costa, Name; 150
Celebration; 93
Charlotte County; 149
Charlotte Harbor; 150
Chautauqua; 116
Chipley; 114
 Name; 115
Choctawatchee, Name; 115
Circus Museum, Ringling; 147
Citrus; 88,97,130,136,140,180
CityPlace, W Palm Beach; 180
City Maps,
 Ft Lauderdale Expyws; 194-195
 Jacksonville; 163
 Kissimmee Expyws; 192-193
 Miami Expressways; 194-195
 Orlando Expressways; 192-193
 Pensacola; 26
 Tallahassee; 191
 Tampa-St. Petersburg; 63
 St. Augusitne; 191
Civil War; 100,108,127,138,141
Clearwater Marine Aquarium; 187
Collier County; 154
Collier, Barron; 152
Colonial Spanish Quarters; 168
Columbia County; 101,128
Coquina Building Material; 165
Corkscrew Swamp, Name; 154
Cowboys; 95
Crab Trap II; 144
Cracker, Florida; 88,95,132
Crosstown Expy; 11,35,98,143
Cuban Bread; 184
Dade Battlefield; 140
Dade, Maj. Francis; 139-140,161
Dania Beach Hurricane; 184
Dade, Maj. Francis; 139-140,161
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; 168
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
 Toll Roads; 189

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 30% jump in exports to the US, which has a

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

Zhou Xiaochuan, central bank chief, needed to do more to encourage demand so more Chinese stay in the country. China increased the yuan against the dollar by 2.1% in March 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.



Weighting the words

- Just as with text, some visual words are more discriminative than others

the, and, or vs. *cow, AT&T, Cher*

- the bigger fraction of the documents a word appears in, the less useful it is for matching
 - e.g., a word that appears in *all* documents is not helping us

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

Large-scale image search



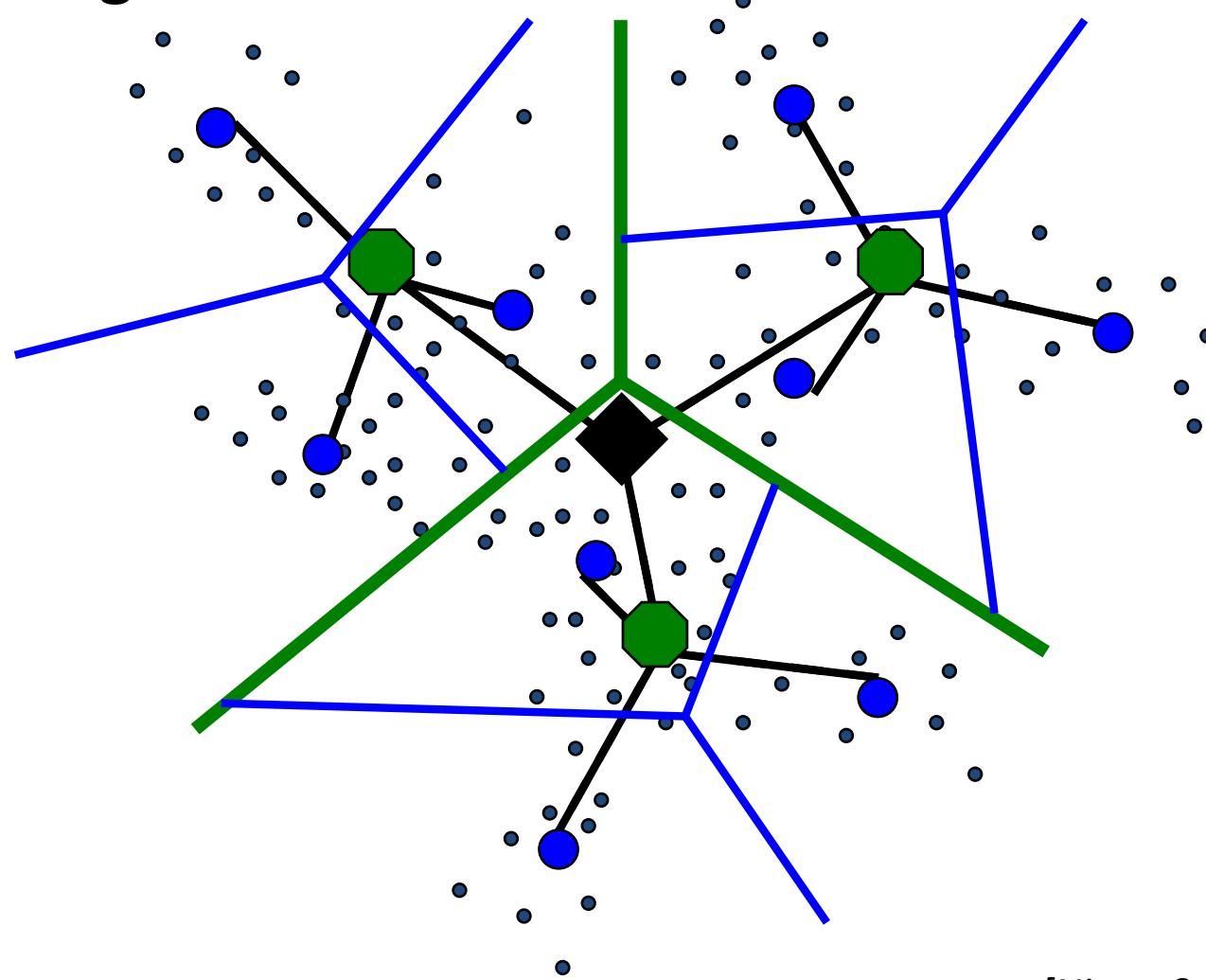
- Build the database:
 - Extract features from the database images
 - Learn a vocabulary using k-means (typical k: 100,000)
 - Compute *weights* for each word
 - Create an inverted file mapping words → images

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?

Vocabulary Tree

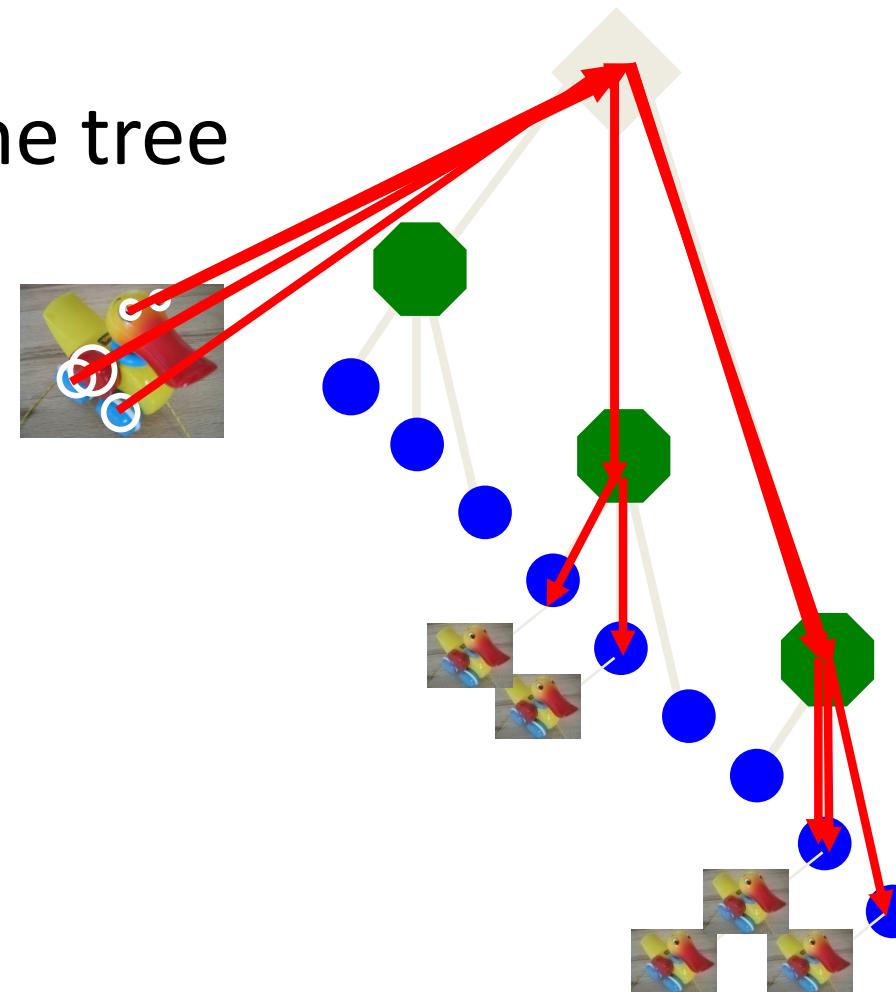
- Large vocabularies can be improved with hierarchical clustering



[Nister & Stewenius, CVPR'06]

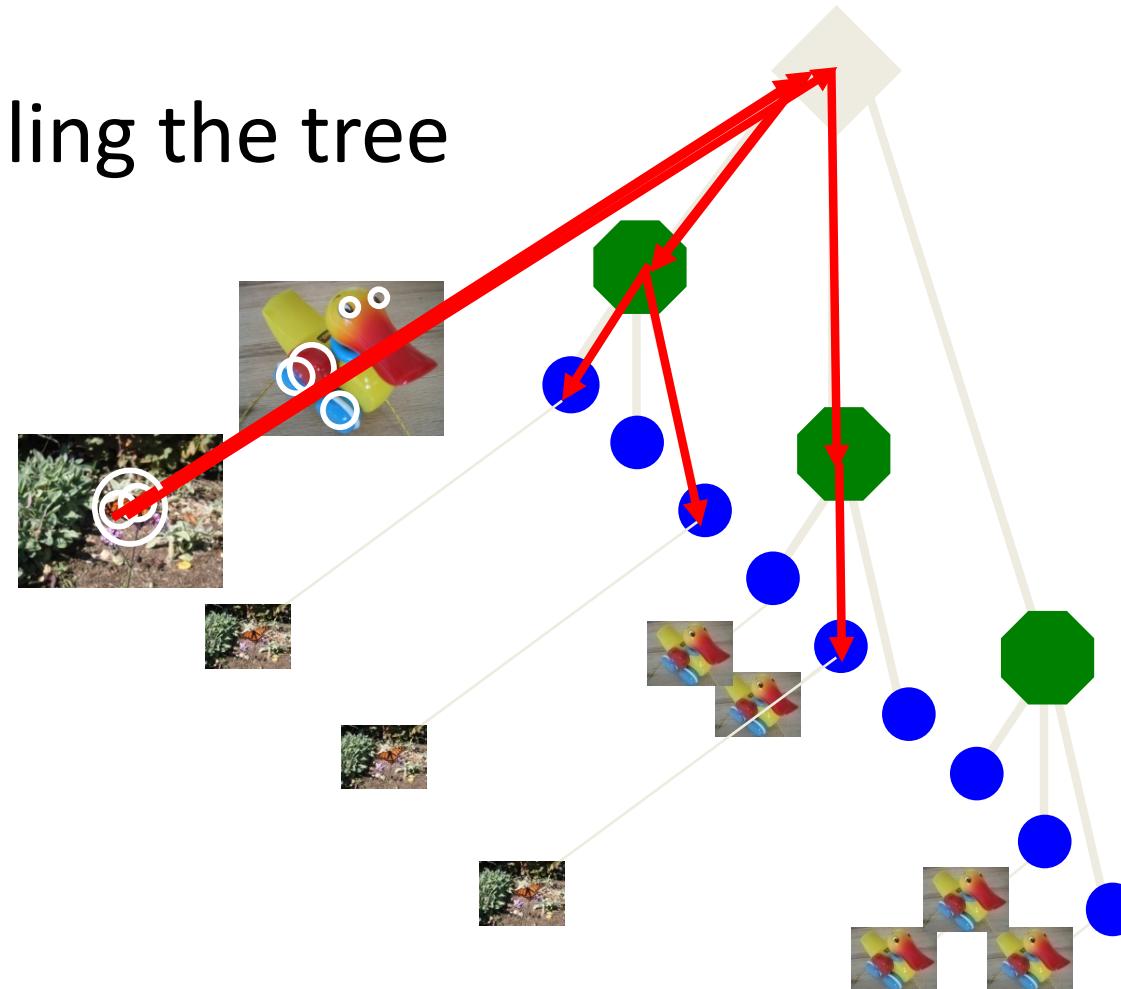
Vocabulary Tree

- Training: Filling the tree



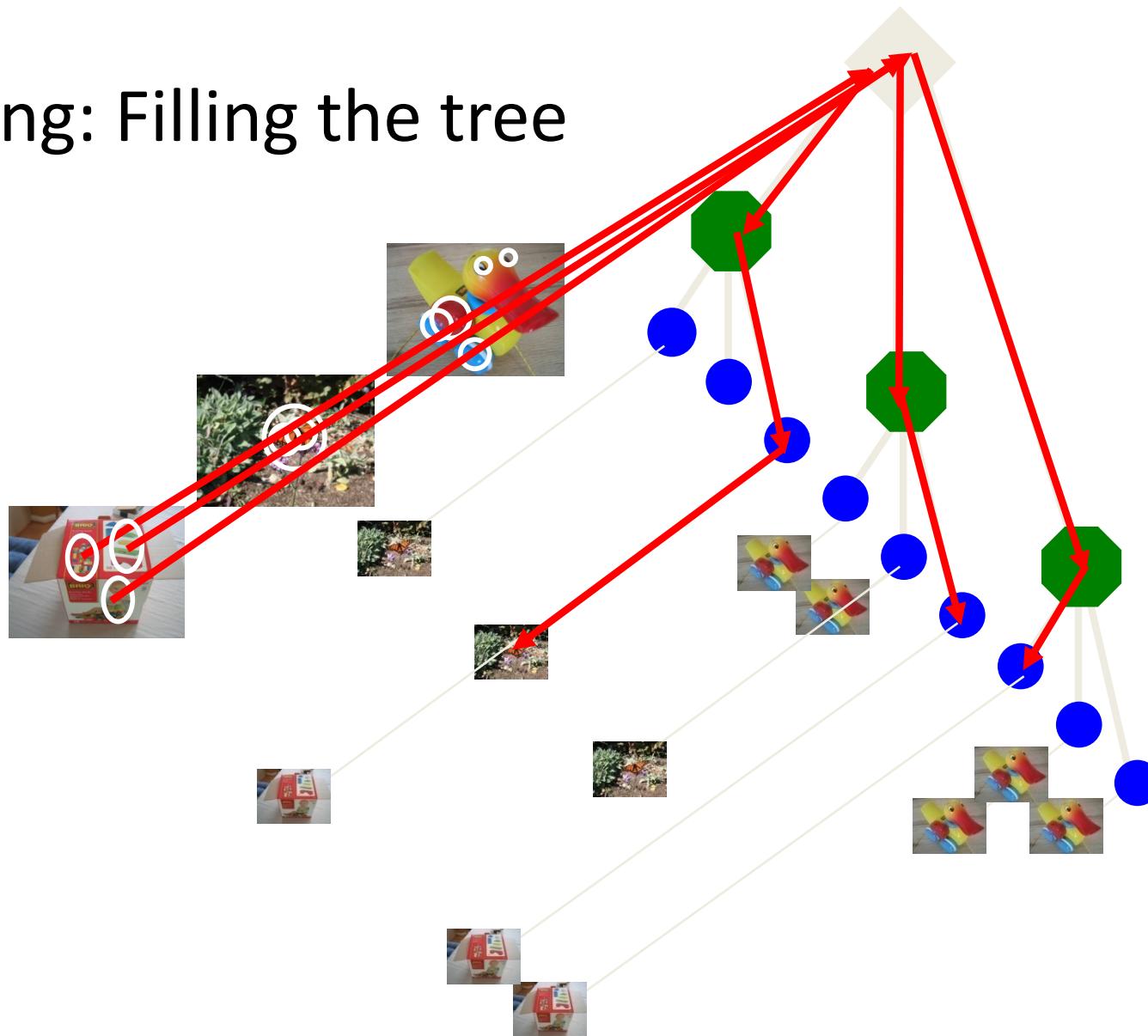
Vocabulary Tree

- Training: Filling the tree



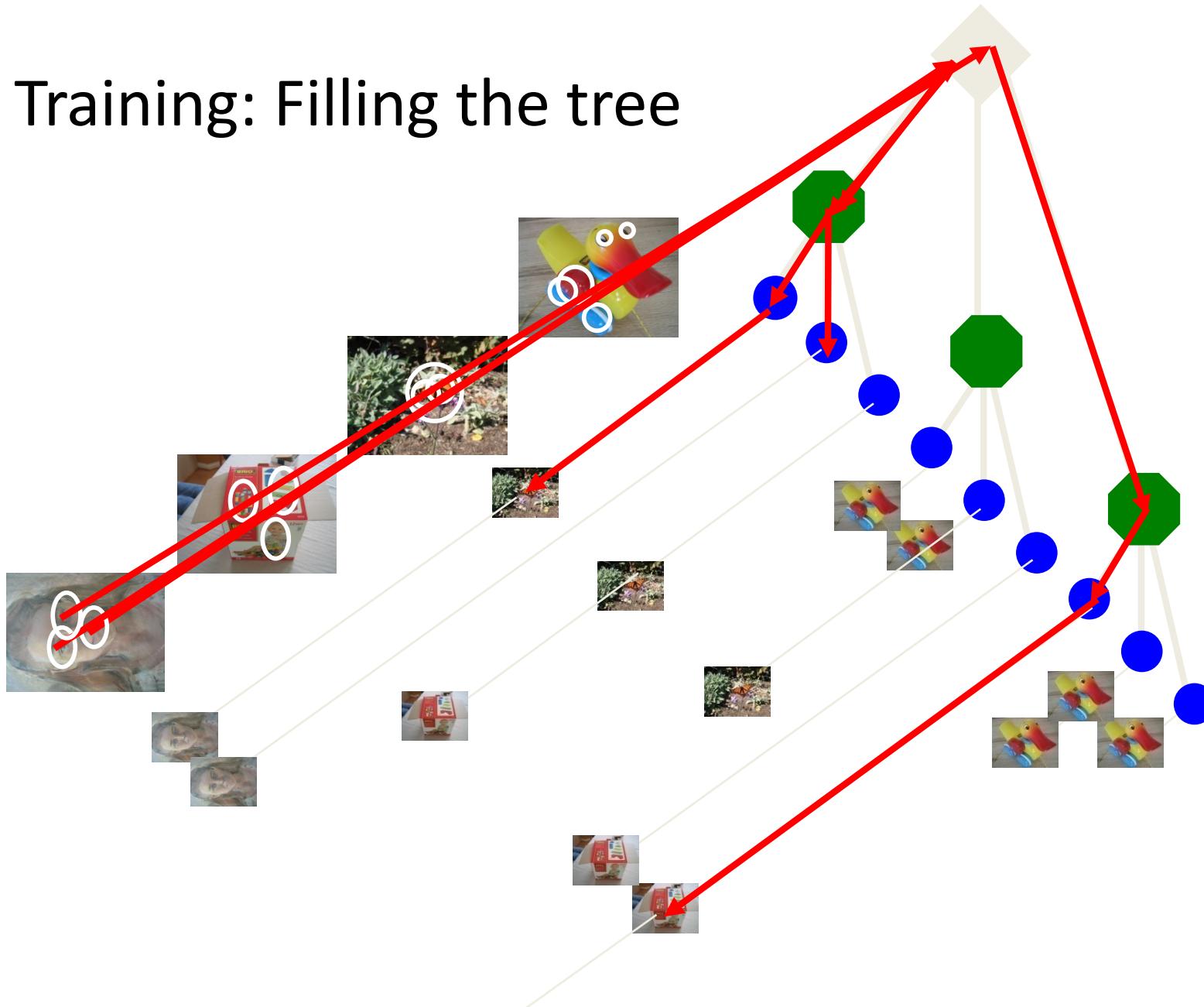
Vocabulary Tree

- Training: Filling the tree



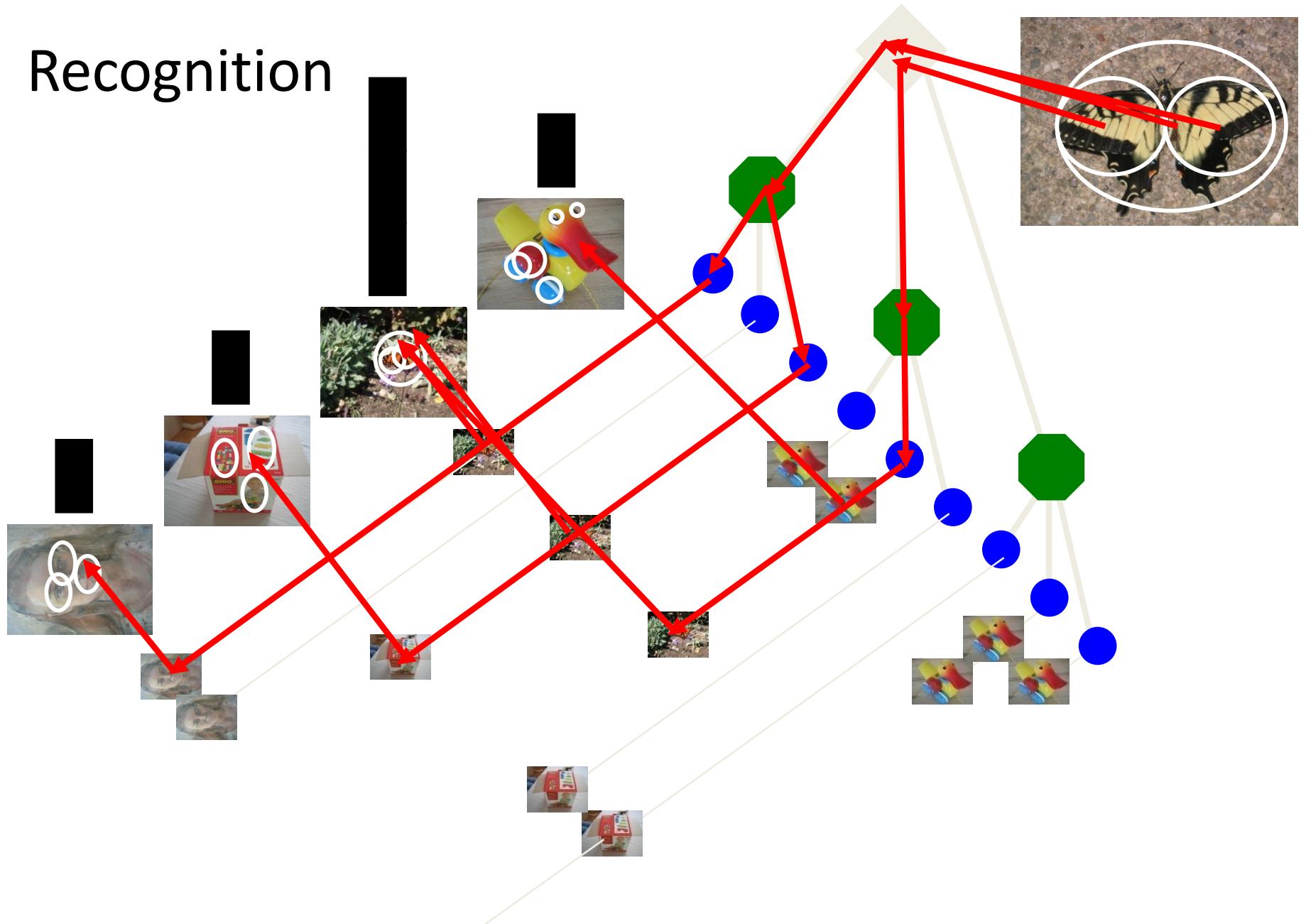
Vocabulary Tree

- Training: Filling the tree

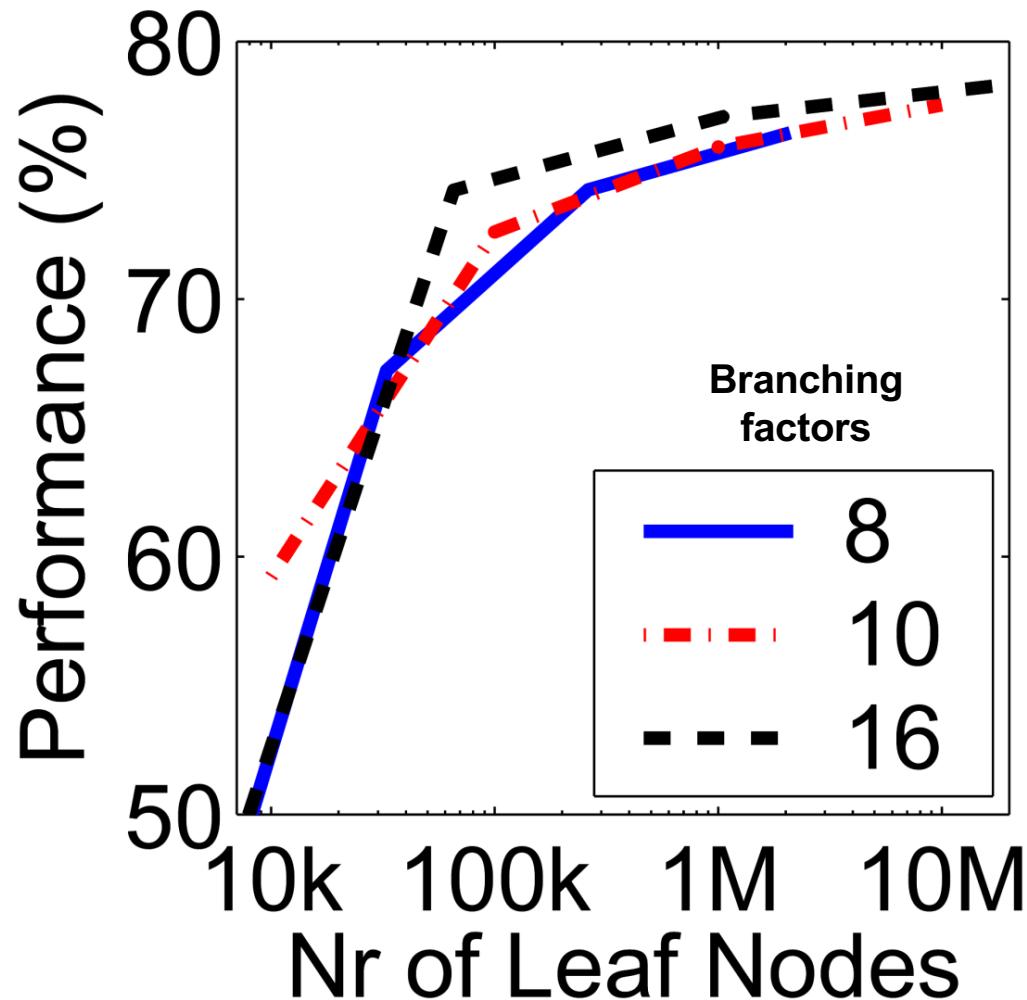


Vocabulary Tree

- Recognition

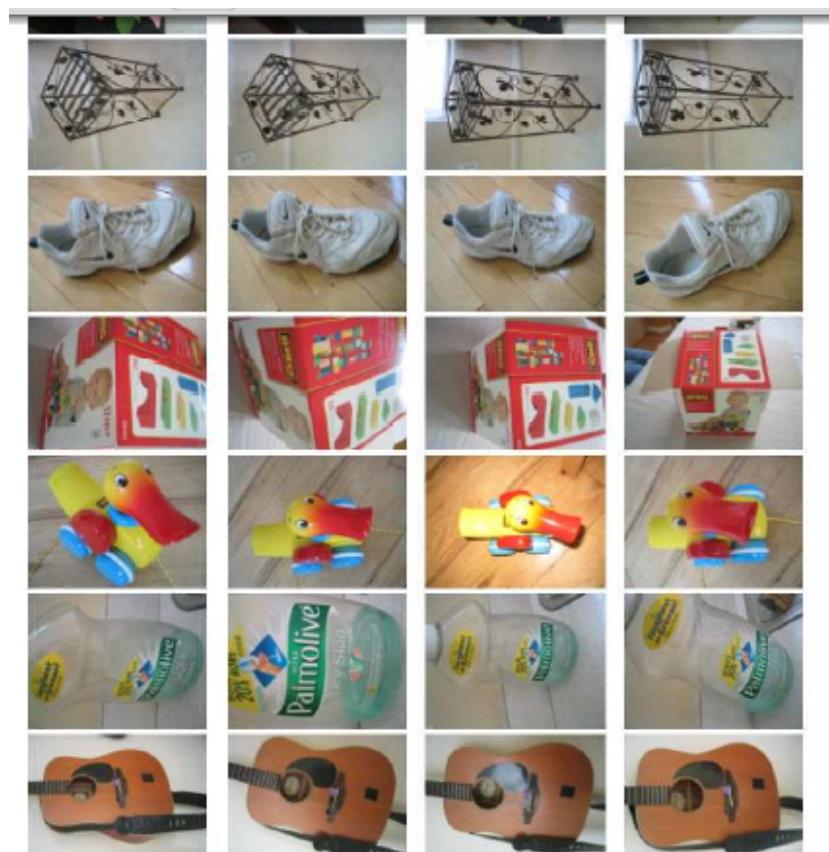


Vocabulary size



Influence on performance, sparsity

Results for recognition task
with 6347 images



Nister & Stewenius, CVPR 2006

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?)

Visual words/bags of words

- + flexible to geometry / deformations / viewpoint
- + compact summary of image content
- + provides vector representation for sets
- + very good results in practice

- background and foreground mixed when bag covers whole image
- optimal vocabulary formation remains unclear
- basic model ignores geometry – must verify afterwards, or encode via features

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?**

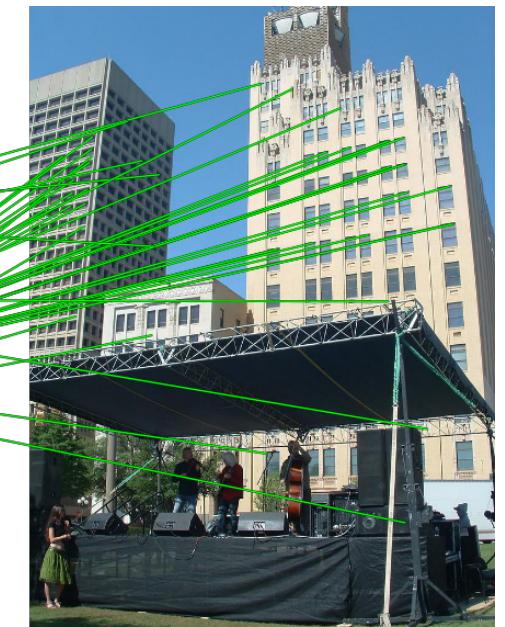
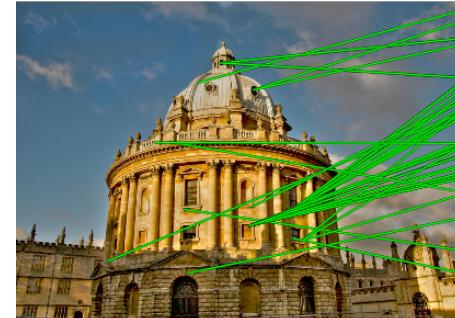
Spatial Verification

Query



DB image with high
BoW similarity

Query



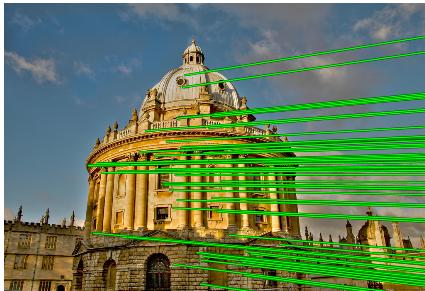
DB image with high
BoW similarity

Both image pairs have many visual words in common.

Slide credit: Ondrej Chum

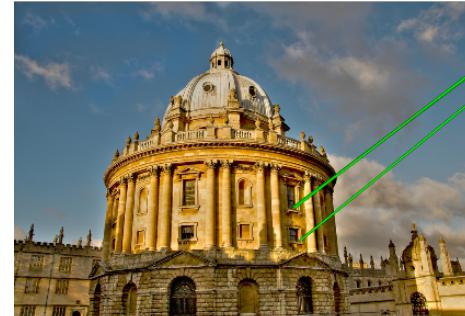
Spatial Verification

Query



DB image with high
BoW similarity

Query



DB image with high
BoW similarity

Only some of the matches are mutually consistent

Slide credit: Ondrej Chum

Outliers

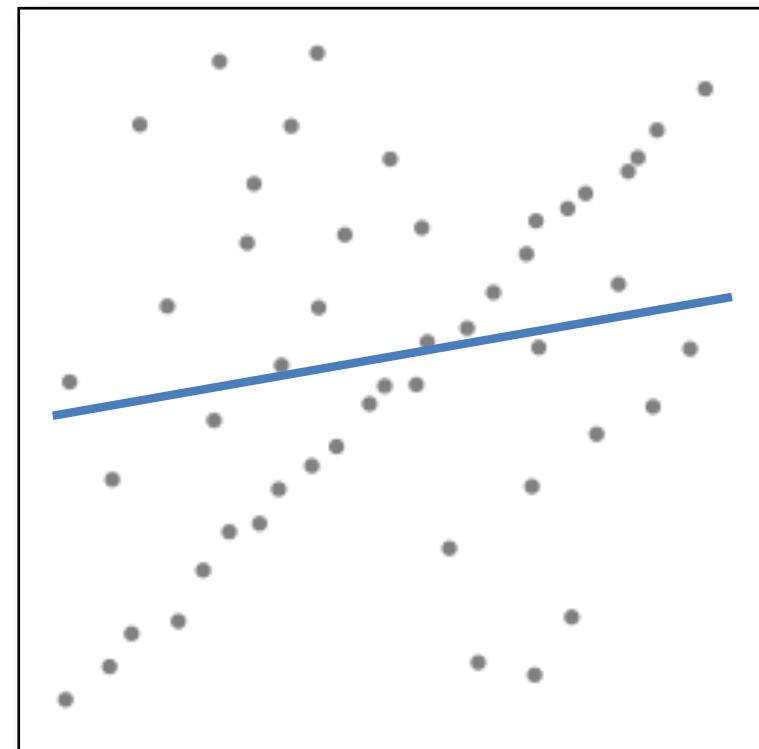
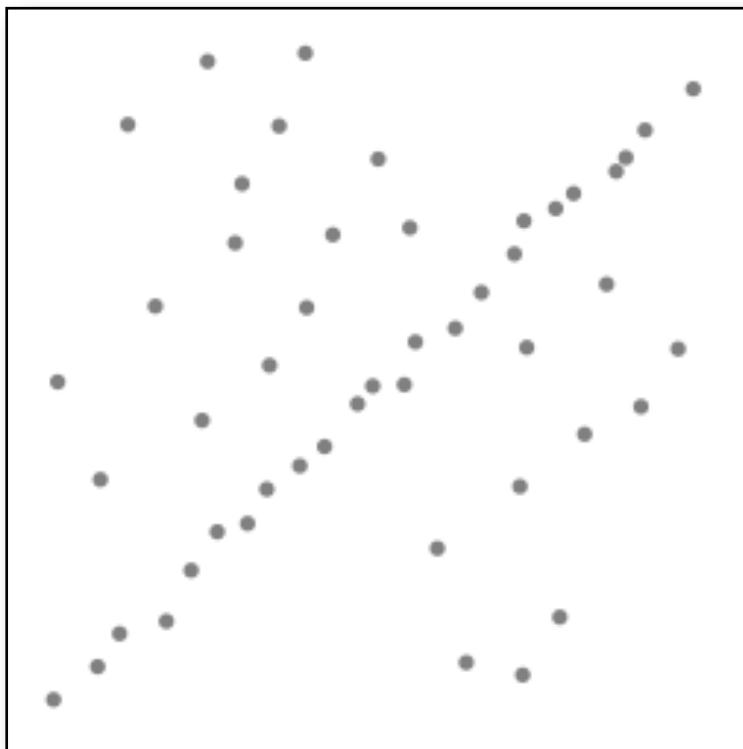


outliers

inliers

Robustness

- A simpler example: linear regression



Problem: Fit a line to these datapoints

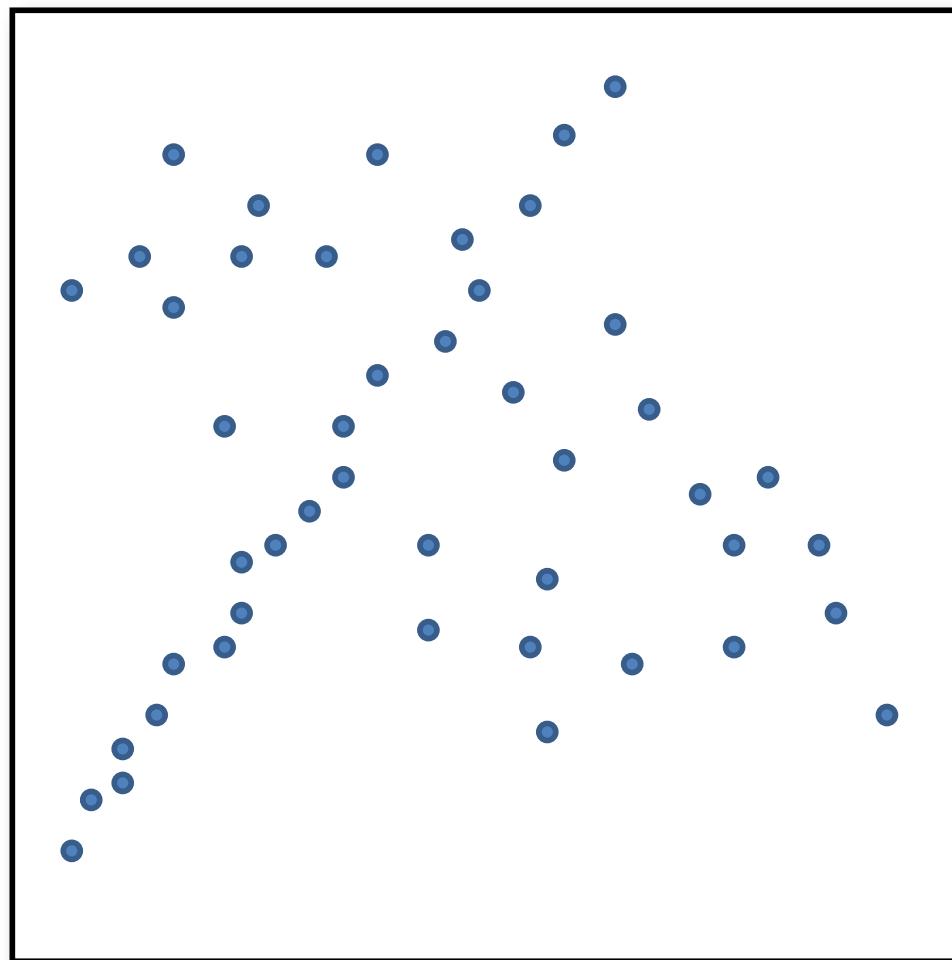
Least squares fit

- How can we fix this?

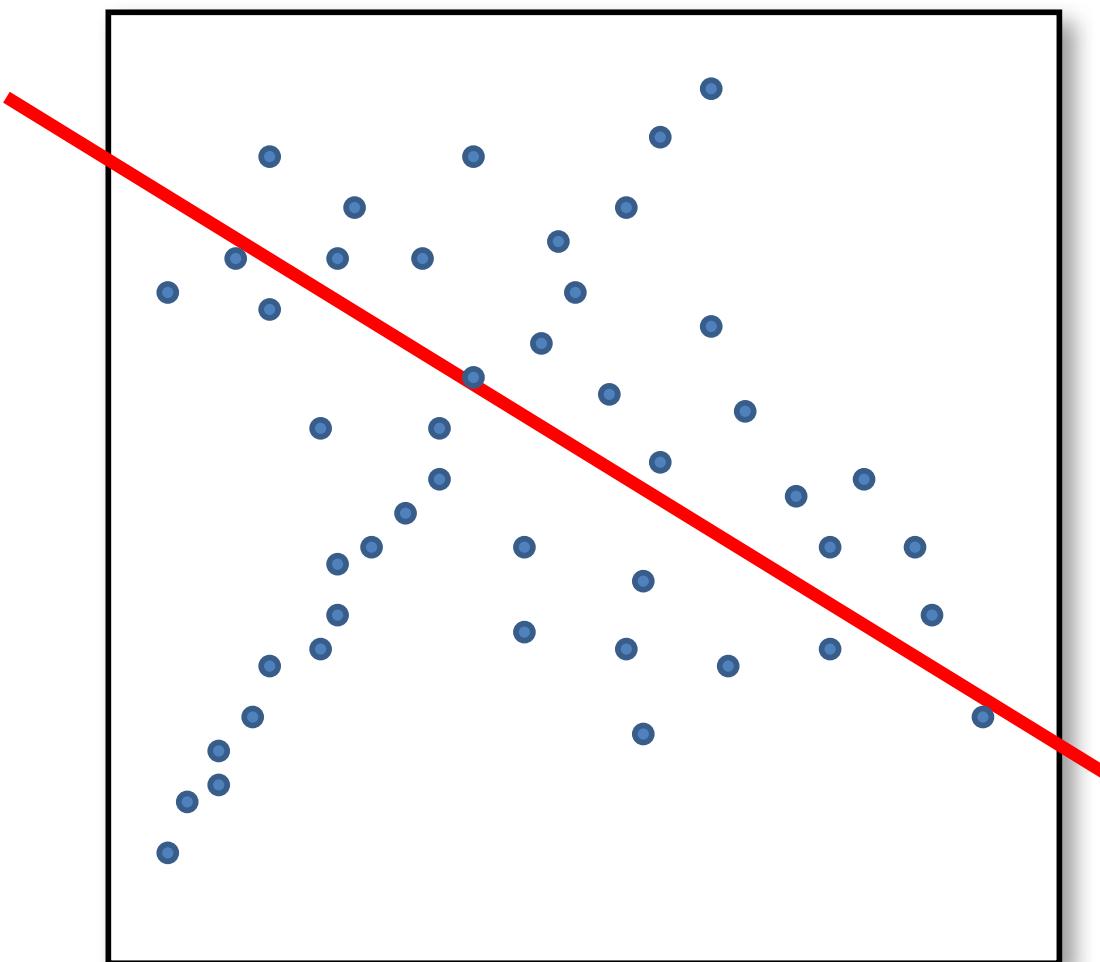
Idea

- Given a hypothesized line
- Count the number of points that “agree” with the line
 - “Agree” = within a small distance of the line
 - I.e., the **inliers** to that line
- For all possible lines, select the one with the largest number of inliers

Counting inliers

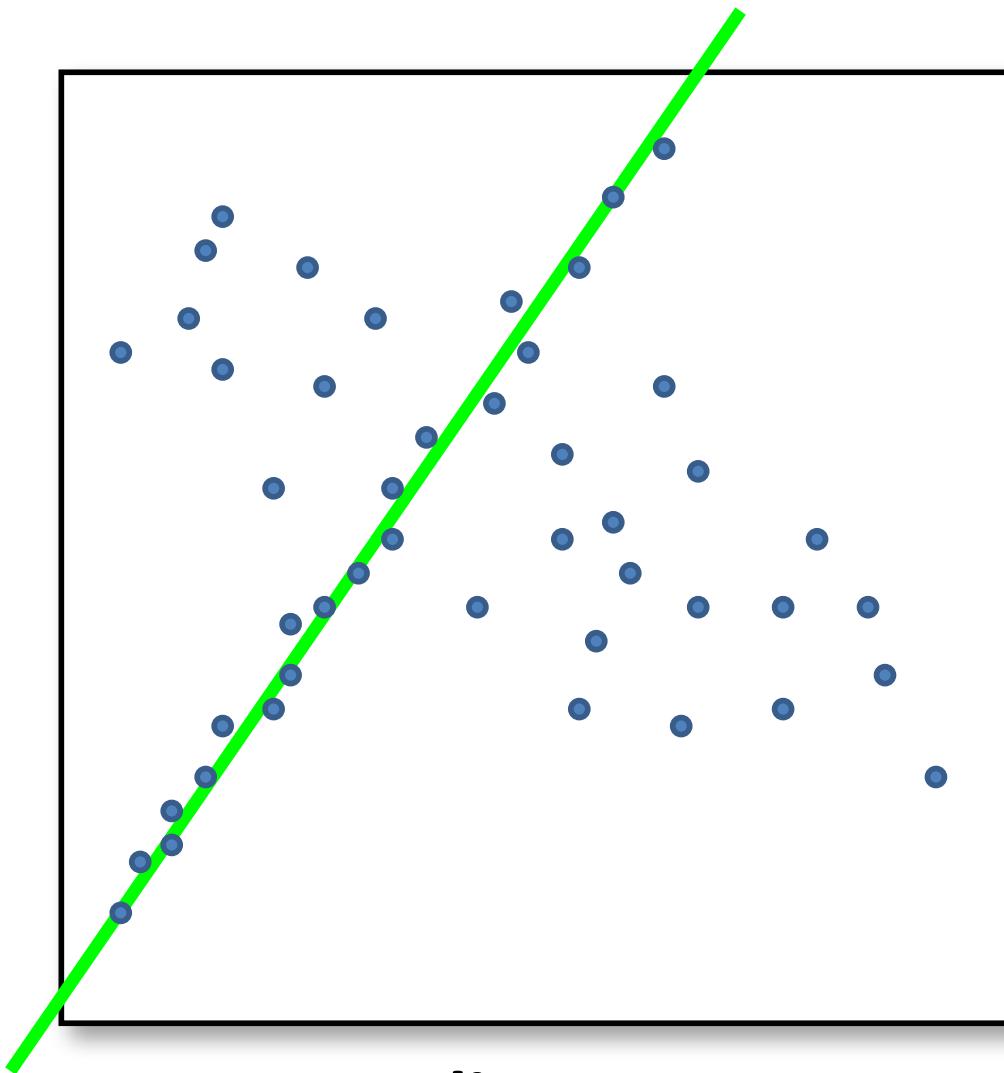


Counting inliers



Inliers: 3

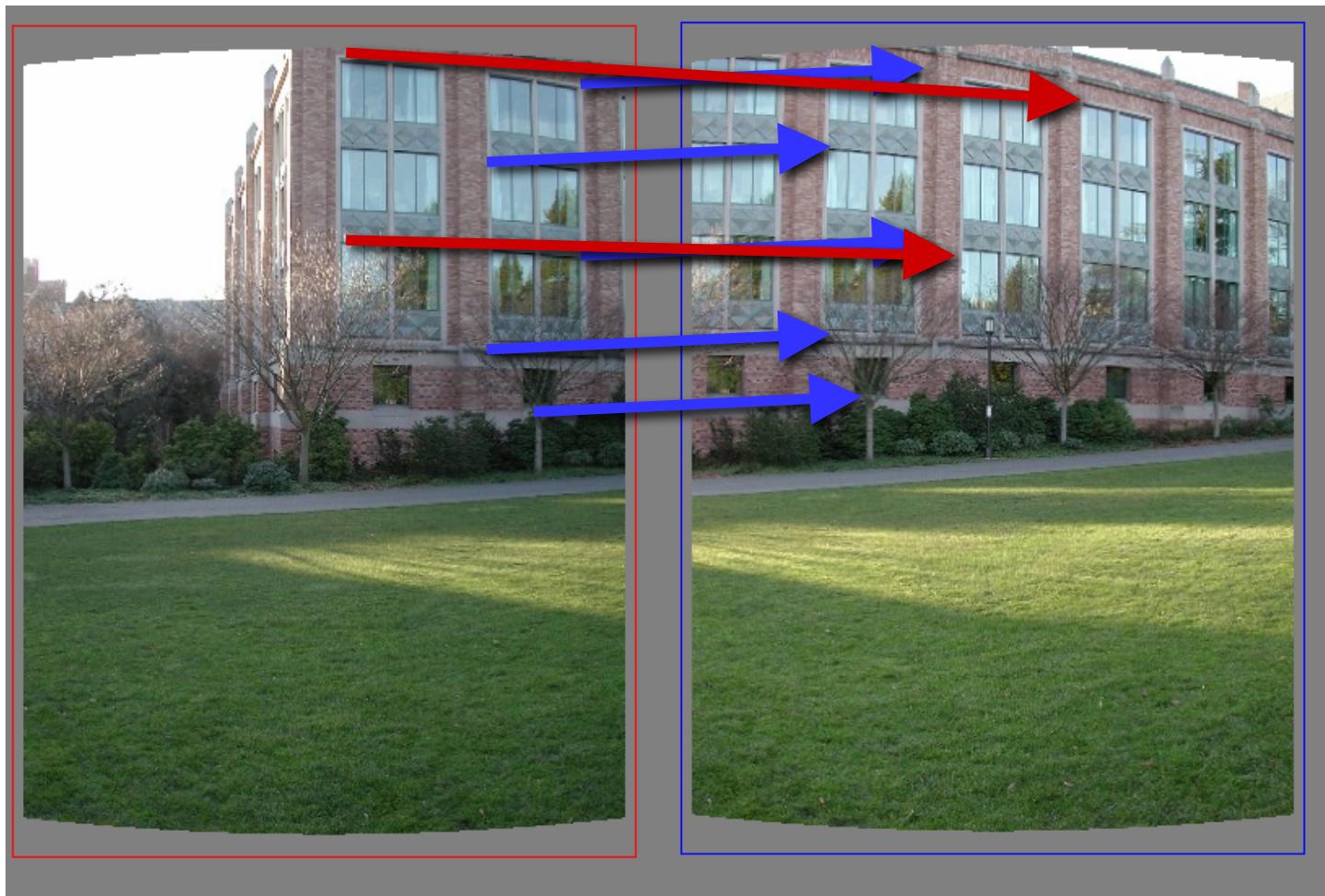
Counting inliers



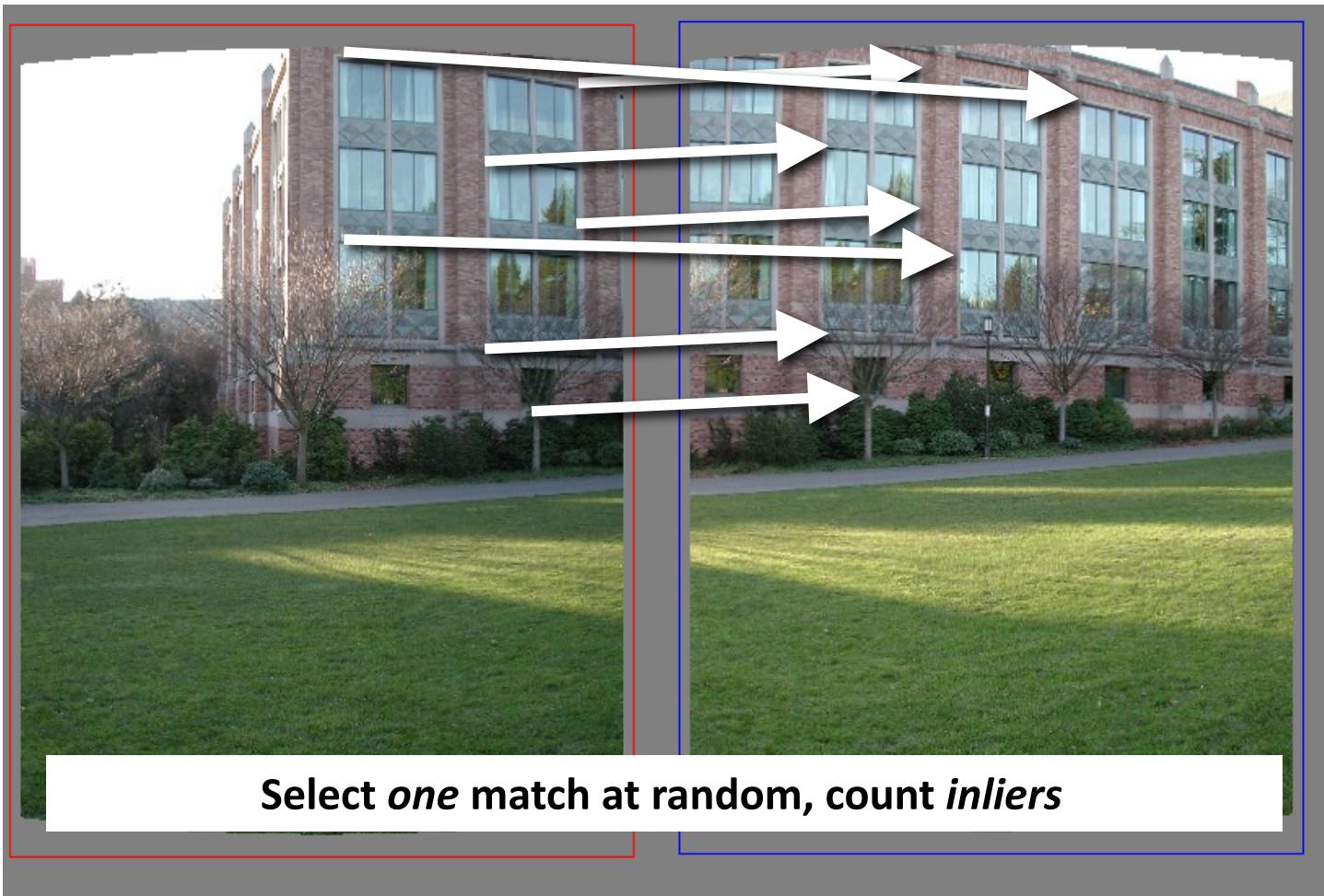
How do we find the best line?

- Unlike least-squares, no simple closed-form solution
- Hypothesize-and-test
 - Try out many lines, keep the best one
 - Which lines?

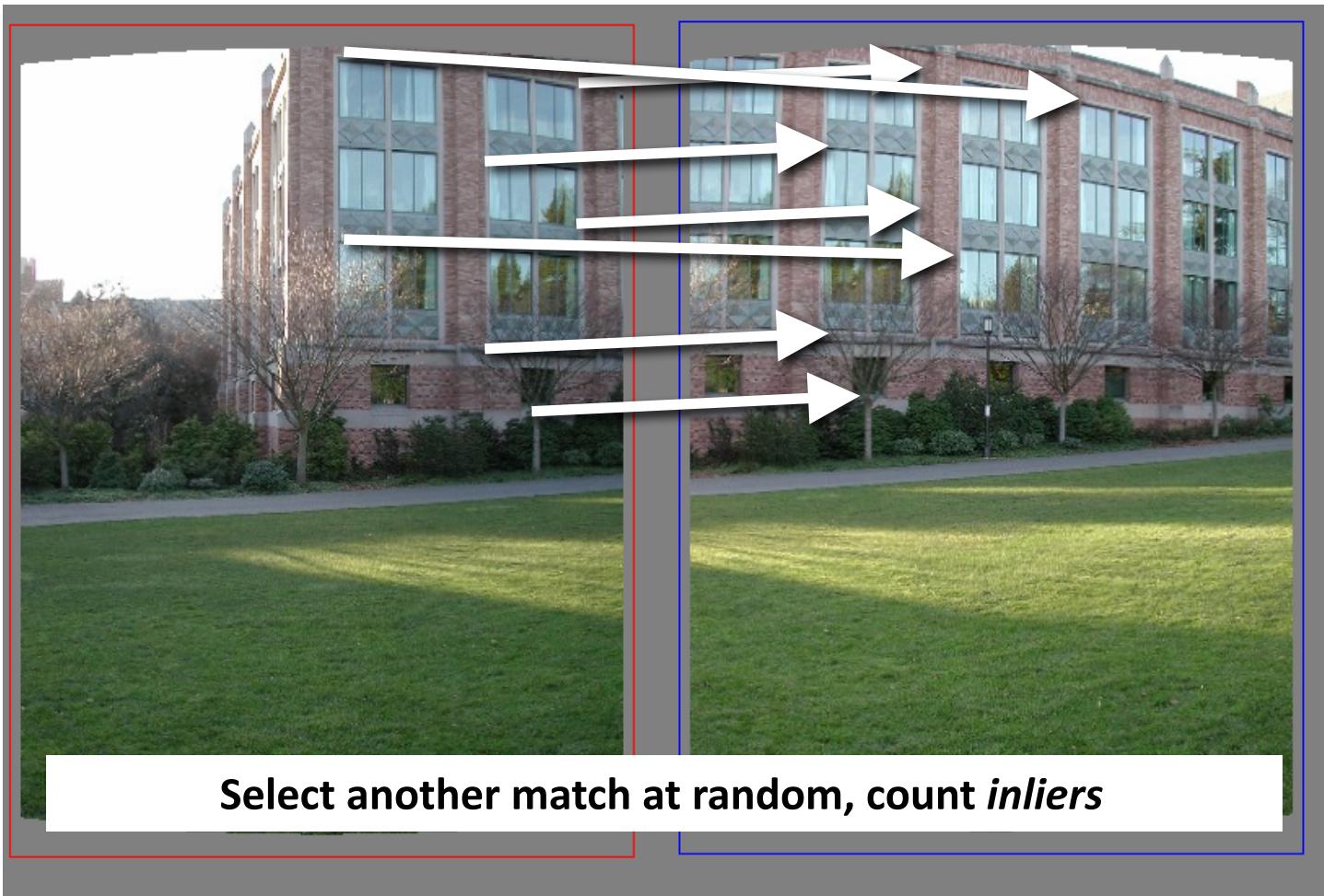
Translations



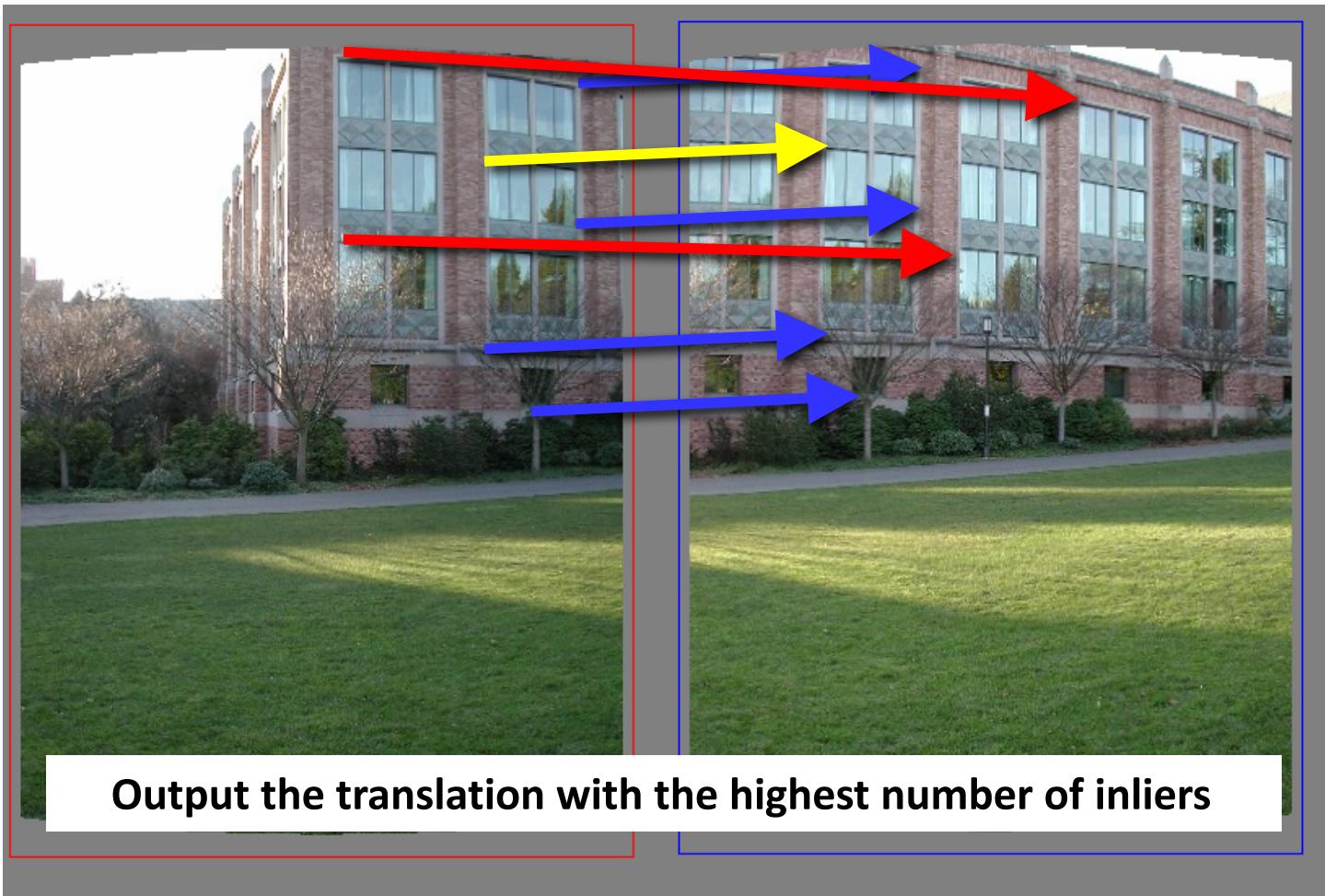
Random Sample Consensus



Random Sample Consensus



Random Sample Consensus



RANSAC

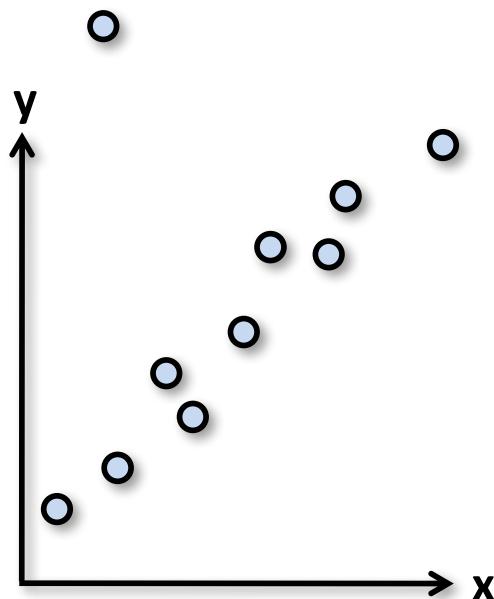
- Idea:
 - All the inliers will agree with each other on the translation vector; the (hopefully small) number of outliers will (hopefully) disagree with each other
- RANSAC only has guarantees if there are < 50% outliers

RANSAC

- **Inlier threshold** related to the amount of noise we expect in inliers
 - Often model noise as Gaussian with some standard deviation (e.g., 3 pixels)
- **Number of rounds** related to the percentage of outliers we expect, and the probability of success we would like to guarantee

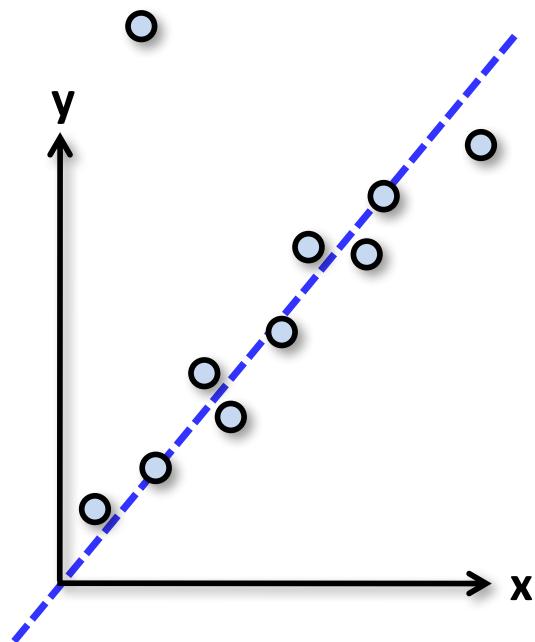
RANSAC

- Back to linear regression
- How do we generate a hypothesis?



RANSAC

- Back to linear regression
- How do we generate a hypothesis?



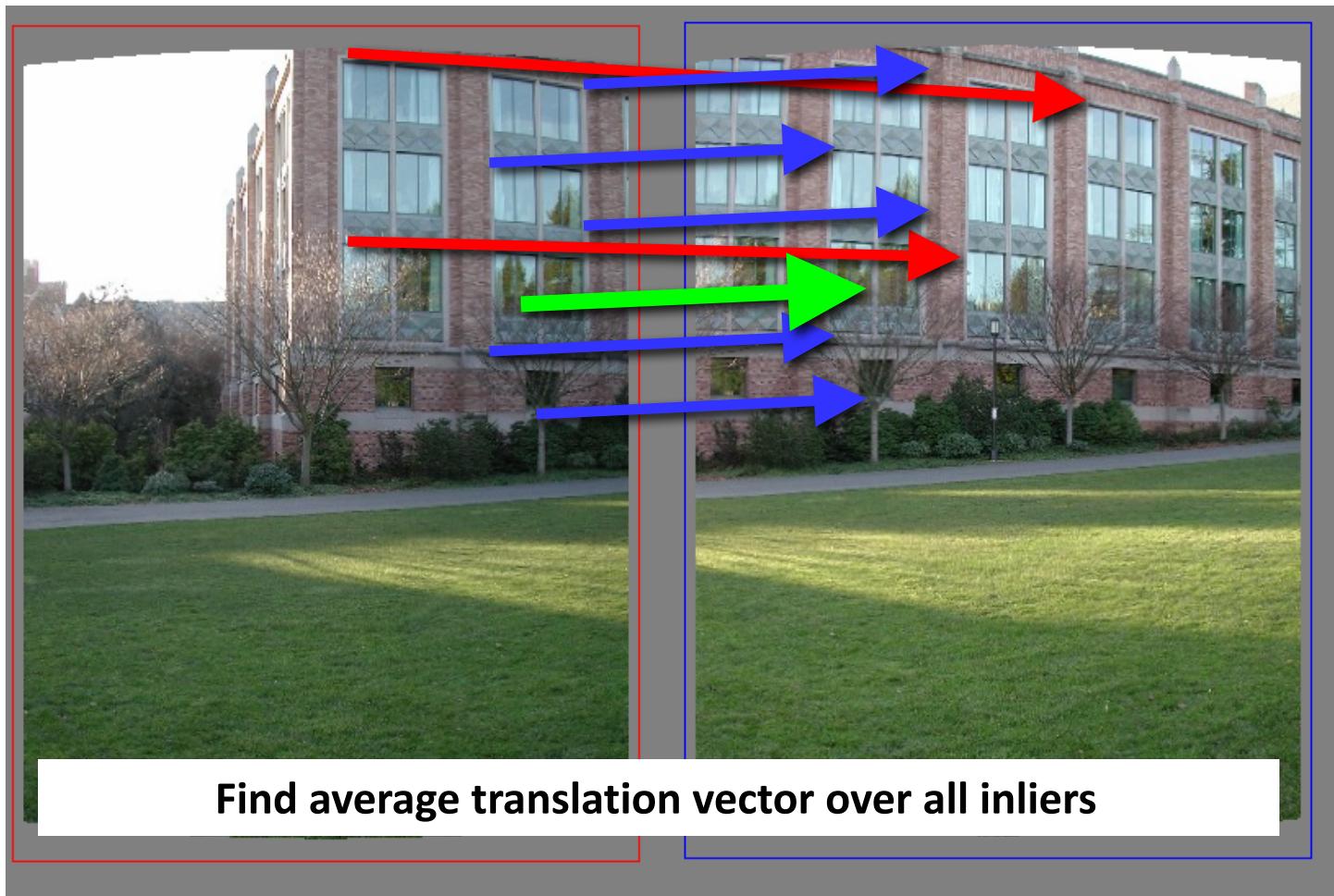
RANSAC

- General version:
 1. Randomly choose s samples
 - Typically $s = \text{minimum sample size that lets you fit a model}$
 2. Fit a model (e.g., line) to those samples
 3. Count the number of inliers that approximately fit the model
 4. Repeat N times
 5. Choose the model that has the largest set of inliers

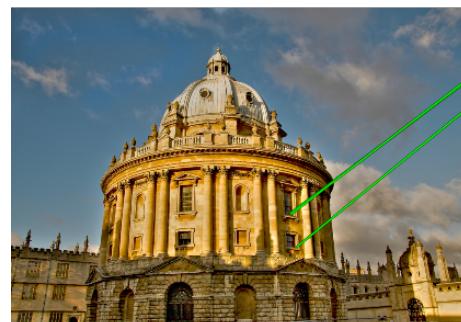
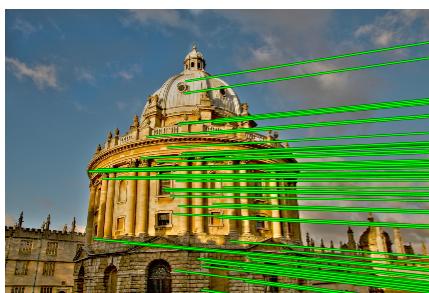
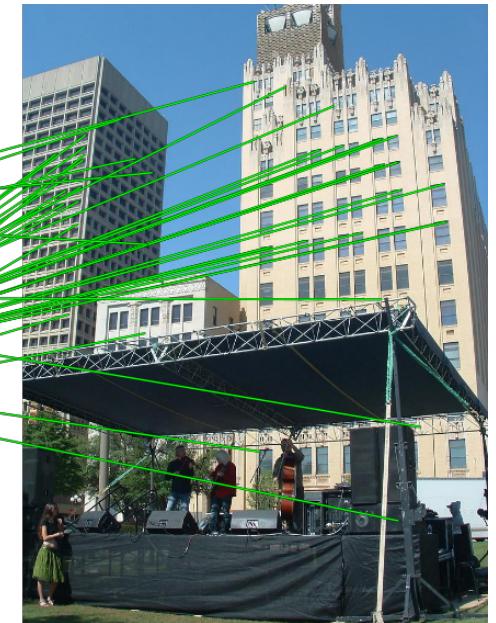
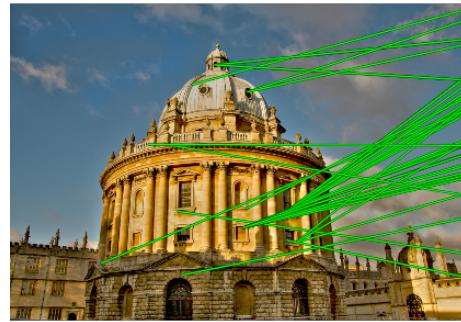
RANSAC pros and cons

- Pros
 - Simple and general
 - Applicable to many different problems
 - Often works well in practice
- Cons
 - Parameters to tune
 - Sometimes too many iterations are required
 - Can fail for extremely low inlier ratios
 - We can often do better than brute-force sampling

Final step: least squares fit



RANSAC verification



Recognition via alignment

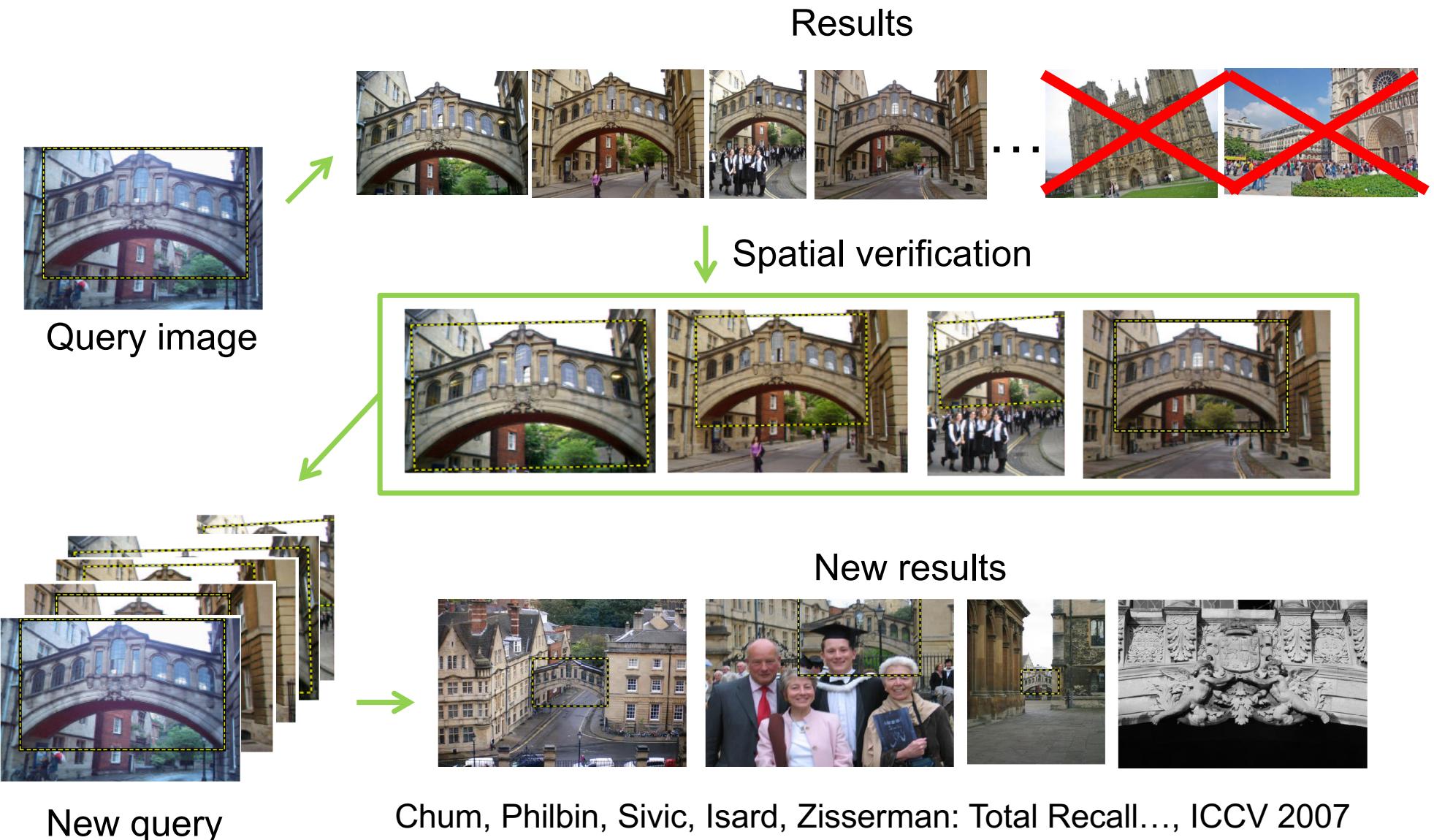
Pros:

- Effective when we are able to find reliable features within clutter
- Great results for matching specific instances

Cons:

- Scaling with number of models
- Spatial verification as post-processing – not seamless, expensive for large-scale problems
- Not suited for category recognition.

Query Expansion



Slide credit: Ondrej Chum