



Category vs. instance recognition

Category:

- Find all the people
- Find all the buildings
- Often within a single image
- Often 'sliding window'

Instance:

- Is this face James?
- Find this specific famous building
- Often within a database of images





Scene recognition dataset



Instance or category?

Recognition



Recognition Issues

How to summarize the content of an entire image? How to gauge overall similarity?

How large should the vocabulary be? How to perform quantization efficiently?

How to score the retrieval results?

How might we add more spatial verification?

Recognition Issues

How to summarize the content of an entire image? How to gauge overall similarity?

How large should the vocabulary be? How to perform quantization efficiently?

How to score the retrieval results?

How might we add more spatial verification?



Visual words



Bag of visual words histograms





Comparing bags of words

Compute cosine similarity (normalized scalar (dot) product) between their occurrence counts, then rank and pick smallest. *Nearest neighbor* search for similar images.



Comparing bags of words

Why might we use cosine similarity here? What 'intuitive' effect does this provide?



Kristen Grauman

Spatial pyramid representation Extension of a bag of features

- Locally orderless representation at several levels of resolution



Lazebnik, Schmid & Ponce (CVPR 2006)

How can we quickly find images in a large database that match a given image region?



Instance recognition

Simple idea

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





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).



Kristen Grauman

Indexing local features

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



Visual words

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



- Quantize via clustering; cluster centers are the visual "words"
- Assign word to each 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

Sampling strategies



Sparse, at interest points



Multiple interest operators



Dense, uniformly



Randomly

- 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]

Fast lookup: inverted index

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 Isi) - 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 Maclay 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 'l'; 165 Big Cypress; 155,158 Big Foot Monster; 105 Billie Swamp Safari; 160 Blackwater River SP; 117 Blue Angels

Butterfly Center, McGuire; 134 CAA (see AAA) CCC. The: 111.113.115.135.142 Ca d'Zan; 147 Caloosahatchee River; 152 Name; 150 Canaveral Natni Seashore: 173 Cannon Creek Airpark; 130 Canopy Read; 106,169 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, Bingling: 147 Citrus: 88.97.130.136.140.180 CityPlace, W Palm Beach: 180 City Maps, Ft Lauderdale Expwys; 194-195 Jacksonville; 163 Kissimmee Expwys: 192-193 Miami Expressways; 194-195 Orlando Expressways; 192-193 Pensacola: 26 Tallahassee; 191 Tampa-St. Petersburg: 63 St. Augsutine; 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 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 Caliboxes; 83 Epiphyles; 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.

Build Inverted Index from Database



Query Inverted Index



Query Inverted Index



3. Compare/sort word counts

Kristen Grauman

Inverted index

Key requirement: sparsity.

If most images contain most words, then we're not better off than exhaustive search.

 Exhaustive search would mean comparing the visual word distribution of a query versus every page.

Recognition 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?

How to score the retrieval results?

How might we add more spatial verification?

Following slides by David Nister (CVPR 2006)



































Vocabulary tree built recursively



Each leaf has inverted index











Vocabulary size



Recognition with 6347 images



Nister & Stewenius, CVPR 2006

Influence on performance, sparsity

Higher branch factor works better (but slower)



(2006) 110,000,000 images in 5.8 Seconds





David Nister





David Nister



David Nister

Visual words/bags of words

- + flexible to geometry / deformations / viewpoint
- + compact summary of image content
- + provides fixed dimensional vector representation for sets
- + very good results in practice
- background and foreground mixed when bag covers whole image -> is it really instance recognition?
- optimal vocabulary formation remains unclear
- basic model ignores geometry must verify afterwards, or encode via features

Recognition 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?

How to score the retrieval results?

How might we add more spatial verification?

Precision and Recall

True positive (tp) – correct attribution True negative (tn) – correct rejection

False positive (fp) – incorrect attribution False negative (fn) – incorrect rejection

$$\begin{aligned} \text{Precision} &= \frac{tp}{tp + fp} \\ \text{Precision} &= \texttt{#relevant} / \texttt{#returned} \\ \text{Recall} &= \frac{tp}{tp + fn} \\ \text{Recall} &= \texttt{#relevant} / \texttt{#total relevant} \end{aligned}$$

By Walber - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=36926283



Scoring retrieval quality



Query

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

precision = #relevant / #returned
recall = #relevant / #total relevant



Results (ordered):















Slide credit: Ondrej Chum

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 Isi) - I-95 Access; 86 AAA (and CAA); 83 AAA National Office; 88 Abbreviations. Colored 25 mile Maps; cover Exit Services; 196 Traveloque: 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 Maclay 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 'l'; 165 Big Cypress; 155,158 Big Foot Monster; 105

Butterfly Center, McGuire; 134 CAA (see AAA) CCC, The; 111,113,115,135,142 Ca d'Zan; 147 Caloosahatchee River; 152 Name; 150 Canaveral Natni Seashore; 173 Cannon Creek Airpark; 130 Canopy Read; 106,169 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 Expwys; 194-195 Jacksonville: 163 Kissimmee Expwys: 192-193 Miami Expressways; 194-195 Orlando Expressways; 192-193 Pensacola: 26 Tallahassee: 191 Tampa-St. Petersburg; 63 St. Augsutine; 191 Civil War; 100,108,127,138,141 Clearwater Manine 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, Mai, Francis: 139-140.161 Dania Beach Hurricane: 184

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 Caliboxes; 83 Epiphyles; 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 Tumpike (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 Sour 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 \$750bn, a predicted 30% compared w China, trade, \$660bn. 7// annoy th vsurplus, commerce, China's exports, imports, US, deliber ^{agrees} yuan, bank, domestic, yuan is foreign, increase, governo trade, value also need demand so country. China е yuan against the dom. nd permitted it to trade within a narrow but the US wants the yuan to be allowed. freely. However, Beijing has made it ch it will take its time and tread carefully be allowing the yuan to rise further in value.

tf-idf weighting

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



Query expansion

Query: golf green

Results:

- How can the grass on the *greens* at a *golf* course be so perfect?
- For example, a skilled *golf*er 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

Slide credit: Ondrej Chum

Query expansion

Results



Spatial verification





Query image

New results



New query

Chum, Philbin, Sivic, Isard, Zisserman: Total Recall..., ICCV 2007 Ondrej Chum

Recognition 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?

How to score the retrieval results?

How might we add more spatial verification?

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

Multi-view matching

VS



Matching two given views for depth

Search for a matching view for recognition

Kristen Grauman

Spatial Verification



Both image pairs have many visual words in common.

Slide credit: Ondrej Chum

Spatial Verification



Only some of the matches are mutually consistent with real-world geometry imaged by a camera.

Ondrej Chum

Spatial Verification: two basic strategies

- RANSAC
 - Typically sort by BoW similarity as initial filter
 - Verify by checking support (inliers) for possible transformations
 - e.g., "success" if find a transformation with > N inlier correspondences
- Generalized Hough Transform
 - Let each matched feature cast a vote on location, scale, orientation of the model object
 - Verify parameters with enough votes

No verification





RANSAC verification



Fails to meet threshold on # inliers! Good!





Recognition via alignment

- Pros:
- Effective for reliable features within clutter
- Great for matching specific instances

Cons:

- Expensive post-process (how long for proj3?!)
- Not suited for category recognition

Summary

- **Bag of words**: quantize feature space into discrete visual words
 - Summarize image by distribution of words
- Inverted index: visual word index for faster query time
- Evaluation:
- Additional spatial verification alignment:
 - Robust fitting : RANSAC, Generalized Hough Transform
 - We will do this in detail later on in the course

Lessons from a decade later

For *Category* recognition (project 3)

- Bag of Feature models remained the state of the art until Deep Learning.
- Spatial layout either isn't that important or its too difficult to encode.
- Quantization error is, in fact, the bigger problem.
 Advanced feature encoding methods address this.
- Bag of feature models are nearly obsolete.
 At best they seem to be inspiring tweaks to deep models e.g., NetVLAD.

Lessons from a decade later

For *instance* retrieval (this lecture):

- deep learning is taking over.
- learn better local features (replace SIFT)
 e.g., MatchNet 2015
- learn better image embeddings (replace visual word histograms)
 e.g., Vo and Hays 2016.
- learn spatial verification
 e.g., DeTone, Malisiewicz, and Rabinovich 2016.
- learn a monolithic deep network to recognition all locations
 e.g., Google's PlaNet 2016.