Plan for Today

• Matching features
• Indexing features
  – Visual words
• Application to image retrieval
Matching local features

- To generate **candidate matches**, find patches that have the most similar appearance (e.g., lowest feature Euclidean distance)
- Simplest approach: compare query to all other features, take the closest (or closest k, or within a thresholded distance) as matches
Robust matching

• At what Euclidean distance value do we have a good match?
• To add robustness to matching, can consider ratio: distance of query to best match / distance to second best match
  • If low, first match looks good
  • If high, could be ambiguous match

Image 1

Image 2
Matching SIFT descriptors

- Nearest neighbor (Euclidean distance)
- Threshold ratio of 1\textsuperscript{st} nearest to 2\textsuperscript{nd} nearest descriptor

![Graph showing PDF for correct and incorrect matches with ratio of distances (closest/next closest)]

Lowe IJCV 2004
Efficient matching

- So far we discussed matching features across just two images
- What if you wanted to match a query feature from one image, to features from all frames in a video?
- Or an image to other images in a giant database?
- 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?
Indexing local features: Setup

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

- When we see close points in feature space, we have similar descriptors, which indicates similar local content
Indexing local features:
Inverted file index

- For text documents, an efficient way to find all \textit{pages} on which a \textit{word} occurs is to use an index…

- We want to find all \textit{images} in which a \textit{feature} occurs.

- To use this idea, we’ll need to map our features to “visual words”.

\begin{itemize}
  \item \textbf{Index}
  \begin{quote}
  "Along I-75," From Detroit to Florida; \textit{inside back cover}
  "Drive I-75," From Boston to Florida; \textit{inside back cover}
  1929 Spanish Trail Roadway; \textit{101-102,104}
  511 Traffic Information; \textit{83}
  A1A (Barrier Isl) - I-95 Access; \textit{86}
  AAA (and CAA); \textit{83}
  AAA National Office; \textit{88}
  Abbreviations, Colored 25 mile Maps; cover
  Exit Services; \textit{196}
  Travelogue; \textit{85}
  Africa; \textit{177}
  Agricultural Inspection Stns; \textit{126}
  Ah-Tah-Ti-Ki Museum; \textit{160}
  Air Conditioning; \textit{First; 112}
  Alabama; \textit{124}
  Alachua; \textit{131}
  Alafia River; \textit{143}
  Alapaha, Name; \textit{126}
  Alfred B Maclay Gardens; \textit{106}
  Alligator Alley; \textit{154-155}
  Alligator Farm, St Augustine; \textit{169}
  Alligator Hole (definition); \textit{157}
  Alligator, Buddy; \textit{155}
  Alligator; \textit{100,135,138,147,156}
  Anastasia Island; \textit{170}
  Anhaica; \textit{108-109,146}
  Apalachicola River; \textit{112}
  Apalachon Mus of Art; \textit{135}
  Aquifer; \textit{102}
  Arabian Nights; \textit{94}
  Art Museum; \textit{Ringling; 147}
  Aruba Beach Cafe; \textit{183}
  Aucliff River Project; \textit{106}
  Balbofoot Web WMA; \textit{151}
  Bahia Mar Marina; \textit{184}
  Baker County; \textit{99}
  Barefoot Mallmen; \textit{182}
  Barge Canal; \textit{137}
  Bee Line Expwy; \textit{80}
  Belz Outlet Mall; \textit{89}
  Bernard Castro; \textit{136}
  Big "I"; \textit{165}
  Big Cypress; \textit{155,158}
  Big Foot Monster; \textit{105}
  Billie Swamp Safari; \textit{160}
  Blackwater River River; \textit{179}
  Blue Angels
  Butterfly Center, McGuire; \textit{134}
  CAA (see AAA)
  CDC, The; \textit{111,113,115,135,142}
  Ca d'Zan; \textit{147}
  Caloosahatchee River; \textit{152}
  Name; \textit{150}
  Canaveral Natl Seashore; \textit{173}
  Cannon Creek Airpark; \textit{130}
  Canopy Road; \textit{106,169}
  Cape Canaveral; \textit{174}
  Castillo San Marcos; \textit{169}
  Cave Diving; \textit{131}
  Cayo Costa, Name; \textit{150}
  Celebration; \textit{93}
  Charlotte County; \textit{149}
  Charlotte Harbor, \textit{150}
  Chautauqua; \textit{116}
  Chipley; \textit{114}
  Name; \textit{115}
  Choctawhatchee, Name; \textit{115}
  Circus Museum, Ringling; \textit{147}
  Citrus; \textit{88,87,130,136,140,180}
  CityPlace, W Palm Beach; \textit{180}
  City Maps,
  FL Lauderdale Expwy; \textit{194-196}
  Jacksonville; \textit{163}
  Kissimmee Expwy; \textit{192-193}
  Miami Expressways; \textit{194-195}
  Orlando Expressways; \textit{192-193}
  Pensacola; \textit{26}
  Tallahassee; \textit{191}
  Tampa-St Petersburg; \textit{63}
  St Augustine; \textit{191}
  Civil War; \textit{100,108,127,138,141}
  Clearwater Marine Aquarium; \textit{187}
  Collier County; \textit{154}
  Collier, Barron; \textit{152}
  Colonial Spanish Quarters; \textit{168}
  Columbia County; \textit{101,128}
  Cocoa Building Material; \textit{165}
  Corkscrew Swamp, Name; \textit{154}
  Cowboys; \textit{195}
  Crab Trap II; \textit{144}
  Cracker, Florida; \textit{88,95,132}
  Cross town Expwy; \textit{11,35,98,143}
  Cuban Bread; \textit{184}
  Dade Battlefield; \textit{140}
  Dade, Maj. Francis; \textit{139-140,161}
  Dania Beach Hurricane; \textit{184}
  Daniel Boone, Florida Walk; \textit{177}
  Daytona Beach; \textit{172-173}
  De Land; \textit{87}
  Driving Lanes; \textit{85}
  Duval County; \textit{163}
  Eau Gallie; \textit{175}
  Edison, Thomas; \textit{152}
  Eglinton AFB; \textit{116-118}
  Eight Reale; \textit{176}
  Ellenton; \textit{144-145}
  Emanuel Point Wreck; \textit{120}
  Emergency Callboxes; \textit{83}
  Epiphyses; \textit{142,148,157,159}
  Escambia Bay; \textit{119}
  Bridge (I-10); \textit{119}
  County; \textit{120}
  Estero; \textit{153}
  Everglades,90,95,139,140,154-160
  Draining of; \textit{156,181}
  Wildlife MA; \textit{160}
  Wonder Gardens; \textit{154}
  Falling Waters SP; \textit{115}
  Fantasy of Flight; \textit{95}
  Fayer Dykes SP; \textit{171}
  Fires, Forest; \textit{166}
  Fires, Prescribed; \textit{148}
  Fisherman’s Village; \textit{151}
  Flagler County; \textit{171}
  Flagler, Henry; \textit{97,165,167,171}
  Florida Aquarium; \textit{186}
  Florida; \textit{12,000 years ago; 187}
  Cavern SP; \textit{114}
  Map of all Expressways; 2-3
  Mus of Natural History; \textit{134}
  National Carerey ; \textit{141}
  Part of Africa; \textit{177}
  Platform; \textit{187}
  Sheriffs Boys Camp; \textit{126}
  Sports Hall of Fame; \textit{130}
  Sun ‘n Fun Museum; \textit{97}
  Supreme Court; \textit{107}
  Florida’s Turnpike (FTP); \textit{178,189}
  25 mile Strip Maps; 66
  Administration; \textit{189}
  Coin System; \textit{190}
  Exit Services; \textit{188}
  HEFT; \textit{76,161,190}
  History; \textit{189}
  Names; \textit{189}
  Service Plazas; \textit{190}
  Spur SR91; \textit{76}
  Ticket System; \textit{190}
  Toi Plazas; \textit{190}
  Ford, Henry; \textit{182}
  \end{quote}
\end{itemize}

K. Grauman
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

D. Nister, CVPR 2006
Visual words: Main idea

D. Nister, CVPR 2006
Each point is a local descriptor, e.g. SIFT feature vector.
“Quantize” the space by grouping (clustering) the features. Note: For now, we’ll treat clustering as a black box.
Visual words

- Patches on the right = regions used to compute SIFT
- Each group of patches belongs to the same "visual word"

Figure from Sivic & Zisserman, ICCV 2003

Adapted from K. Grauman
Visual words for indexing

- Map high-dimensional descriptors to tokens/words by quantizing the feature space.
  - Each cluster has a center.
  - Determine which word to assign to each new image region by finding the closest cluster center.
  - To compare features: Only compare query feature to others in same cluster (speed up).
  - To compare images: see next slide.

Adapted from K. Grauman.
Inverted file index

- Database images are loaded into the index, by mapping words to image numbers
Inverted file index

When will this indexing process give us a gain in efficiency?

For a new query image, find which database images share a word with it, and retrieve those images as matches (or inspect only those further).

We call this retrieval process *instance recognition*

Adapted from K. Grauman
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 the brain from our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the cerebral cortex; the cerebral cortex was a movie screen, so to speak, upon which the image in the eye was projected. Through the discoveries of Hubel and Wiesel we now know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path to the various cell layers of the optical cortex, Hubel and Wiesel have been able to demonstrate that the message about an image falling on the retina undergoes stepwise analysis in 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.

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 imports of $660bn. This is likely to annoy the US, which has long argued that China's exports are unfairly helped by a deliberately undervalued yuan. Beijing agrees the surplus is too high, but says the yuan is only one factor. Bank of China governor Zhou Xiaochuan said the country also needed to do more to boost domestic demand so more goods stayed within the country. China increased the value of the yuan against the dollar by 2.1% in July 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.
Describing images w/ visual words

• Summarize entire image based on its distribution (histogram) of word occurrences

• Analogous to bag of words representation commonly used for documents

Feature patches:

Adapted from K. Grauman
Describing images w/ visual words

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

Feature patches:
Comparing bags of words

- Rank images by normalized scalar product between their occurrence counts---*nearest neighbor* search for similar images.

\[
sim(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|} = \frac{\sum_{i=1}^{V} d_j(i) \times q(i)}{\sqrt{\sum_{i=1}^{V} d_j(i)^2} \times \sqrt{\sum_{i=1}^{V} q(i)^2}}
\]

for vocabulary of \( V \) words
Bags of words: pros and cons

+ flexible to geometry / deformations / viewpoint
+ compact summary of image content
+ good results in practice

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

Adapted from K. Grauman
Summary: Inverted file index and bags of words similarity

Offline:
- Extract features in database images, cluster them to find words = cluster centers, make index

Online (during search):
1. Extract words in query (extract features and map each to closest cluster center)
2. Use inverted file index to find database images relevant to query
3. Rank database images by comparing word counts of query and database image

Adapted from K. Grauman
One more trick: *tf-idf* weighting

- **Term frequency** – **inverse document frequency**
- Describe image by frequency of each word within it, but 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 in which word \( i \) occurs

Normalized bag-of-words

Adapted from K. Grauman
Bags of words for content-based image retrieval

Visually defined query

“Find this clock”

“Groundhog Day” [Rammis, 1993]

“Find this place”
Example

retrieved shots

Slide from Andrew Zisserman
Sivic & Zisserman, ICCV 2003
Video Google System

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

Sivic & Zisserman, ICCV 2003

- Demo online at:
  http://www.robots.ox.ac.uk/~vgg/research/vgoogle/index.html
Both image pairs have many visual words in common
Only some of the matches are mutually consistent
Example Applications

Mobile tourist guide
- Object/building recognition
- Self-localization
- Photo/video augmentation

[Quack, Leibe, Van Gool, CIVR’08]
Scoring retrieval quality

Database size: 10 images

Relevant (total): 5 images
(e.g. images of Golden Gate)

**precision** = # returned relevant / # returned

**recall** = # returned relevant / # total relevant

Results (ordered):

Ondrej Chum
Indexing and retrieval: Summary

• **Bag of words** representation: quantize feature space to make discrete set of visual words
  – Index individual words
  – Summarize image by distribution of words

• **Inverted index**: pre-compute index to enable faster search at query time

• **Recognition of instances**: match local features
  – Optionally, perform spatial verification

Adapted from K. Grauman