**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.

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

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.

\[
\text{sim}(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|}
\]

\[
\text{sim}(d_j, q) = \frac{\sum_{i=1}^{V} d_j(i) \cdot q(i)}{\sqrt{\sum_{i=1}^{V} d_j(i)^2 \cdot \sum_{i=1}^{V} q(i)^2}}
\]

for vocabulary of \(V\) words

Application: Large-Scale Retrieval

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

Spatial Verification: RANSAC
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% jump in exports to $750bn, compared with a 18% rise in imports to $660bn. The figures are likely to further 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.
Query expansion

New query  \rightarrow \text{Query image}  \rightarrow \text{Results}  \rightarrow \text{Spatial verification}  \rightarrow \text{New results}

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

Slide credit: Ondřej Chum

Recognition via local-feature based alignment

Pros:

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

Cons:

• Spatial verification as post-processing – not seamless, expensive for large-scale problems
• Not suited for generic category recognition

Making the Sky Searchable: Fast Geometric Hashing for Automated Astrometry

Sam Roweis, Dustin Lang & Keir Mierle
University of Toronto

David Hogg & Michael Blanton
New York University
Example

A shot of the Great Nebula, by jerry/photographs (c. 2006), from astropolix.com
http://astrometry.net/gallery.html

An amateur shot of M100, by Filippo Ciferri (c. 2007), from flickr.com
http://astrometry.net/gallery.html

A handheld image of Bode's nebula (c. 2007) by Peter Kunzler, from
http://astrometry.net/gallery.html
Today

• Generic object recognition

What does recognition involve?

Verification: is that a lamp?
Detection: are there people?

Identification: is that Potala Palace?

Object categorization

mountain  tree  building  banner  street lamp  vendor  people
Scene and context categorization

Instance-level recognition problem

Generic categorization problem
Object Categorization

• Task Description
  - "Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label."

• Which categories are feasible visually?

Visual Object Categories

• Basic Level Categories in human categorization
  [Rosch 76, Lakoff 87]
  - The highest level at which category members have similar perceived shape
  - The highest level at which a single mental image reflects the entire category
  - The level at which human subjects are usually fastest at identifying category members
  - The first level named and understood by children
  - The highest level at which a person uses similar motor actions for interaction with category members

Visual Object Categories

• Basic-level categories in humans seem to be defined predominantly visually.
• There is evidence that humans (usually) start with basic-level categorization before doing identification.
How many object categories are there?

~10,000 to 30,000

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Biederman 1987

Other Types of Categories

- Functional Categories
  - e.g. chairs = "something you can sit on"
Other Types of Categories

- Ad-hoc categories
  - e.g. “something you can find in an office environment”

Why recognition?

- Recognition a fundamental part of perception
  - e.g., robots, autonomous agents

- Organize and give access to visual content
  - Connect to information
  - Detect trends and themes

Posing visual queries

Kooaba, Bay & Quack et al.
Autonomous agents able to detect objects

Finding visually similar objects

Discovering visual patterns
Auto-annotation

Challenges: robustness
- Illumination
- Object pose
- Clutter
- Occlusions
- Intra-class appearance
- Viewpoint

Realistic scenes are crowded, cluttered, have overlapping objects.
Challenges: importance of context

Challenges: complexity

Almost 90% of web traffic is visual!
Challenges: complexity

- Thousands to millions of pixels in an image
- 30+ degrees of freedom in the pose of articulated objects (humans)
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

Challenges: learning with minimal supervision

What works today

- Reading license plates, zip codes, checks
What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition

Source: Lana Lazebnik

What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

Source: Lana Lazebnik

What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)

Source: Lana Lazebnik
What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)
- Recognition of generic categories(*)!
Progress charted by datasets

Evolution of methods

- Hand-crafted models
- 3D geometry
- Hypothesize and align

- Hand-crafted features
- Learned models
- Data-driven

“End-to-end” learning of features and models*,**

* Labeled data availability
** Architecture design decisions, parameters.
Supervised classification

- Given a collection of labeled examples, come up with a function that will predict the labels of new examples.

  - “four”
  - “nine”

  Training examples Novel input

- How good is the function we come up with to do the classification?
- Depends on
  - Mistakes made
  - Cost associated with the mistakes

Supervised classification

- Consider the two-class (binary) decision problem
  - \( L(4 \rightarrow 9) \): Loss of classifying a 4 as a 9
  - \( L(9 \rightarrow 4) \): Loss of classifying a 9 as a 4

- Risk of a classifier \( s \) is expected loss:

\[
R(s) = \Pr(4 \rightarrow 9 \mid s) L(4 \rightarrow 9) + \Pr(9 \rightarrow 4 \mid s) L(9 \rightarrow 4)
\]

- We want to choose a classifier so as to minimize this total risk

Supervised classification

- Optimal classifier will minimize total risk.
- At decision boundary, either choice of label yields same expected loss.

If we choose class “four” for point \( x \) at boundary, expected loss is:

\[
= \Pr(\text{class is } 9 \mid x) L(9 \rightarrow 4) + \Pr(\text{class is } 4 \mid x) L(4 \rightarrow 4)
\]

If we choose class “nine” for point \( x \) at boundary, expected loss is:

\[
= \Pr(\text{class is } 4 \mid x) L(4 \rightarrow 9)
\]
Supervised classification

Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

So, best decision boundary is at point $x$ where

$$P(\text{class is } 9 \mid x) \cdot L(9 \rightarrow 4) = P(\text{class is } 4 \mid x) \cdot L(4 \rightarrow 9)$$

To classify a new point, choose class with lowest expected loss; i.e., choose “four” if

$$P(4 \mid x) \cdot L(4 \rightarrow 9) > P(9 \mid x) \cdot L(9 \rightarrow 4)$$

Supervised classification

Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

So, best decision boundary is at point $x$ where

$$P(\text{class is } 9 \mid x) \cdot L(9 \rightarrow 4) = P(\text{class is } 4 \mid x) \cdot L(4 \rightarrow 9)$$

To classify a new point, choose class with lowest expected loss; i.e., choose “four” if

$$P(4 \mid x) \cdot L(4 \rightarrow 9) > P(9 \mid x) \cdot L(9 \rightarrow 4)$$

How to evaluate these probabilities?

Probability

Basic probability

- $X$ is a random variable
- $P(X)$ is the probability that $X$ achieves a certain value
- Called a PDF: probability distribution/density function

$$P(X)$$

- $0 \leq P(X) \leq 1$
- $\int_{\text{continuous } X} P(X) \, dX = 1$ or $\sum_{\text{discrete } X} P(X) = 1$

- Conditional probability: $P(X \mid Y)$
  - probability of $X$ given that we already know $Y$
Example: learning skin colors
• We can represent a class-conditional density using a histogram (a “non-parametric” distribution)

\[ P(x|\text{skin}) \]
\[ P(x|\text{not skin}) \]

Percentage of skin pixels in each bin
Feature \( x = \text{Hue} \)

Now we get a new image, and want to label each pixel as skin or non-skin. What’s the probability we care about to do skin detection?

Bayes rule
\[
P(\text{skin} | x) = \frac{P(x | \text{skin})P(\text{skin})}{P(x)}
\]

Where does the prior come from?
Why use a prior?
Example: classifying skin pixels

Now for every pixel in a new image, we can estimate probability that it is generated by skin.

Classify pixels based on these probabilities
- if \( p(\text{skin} | x) > \theta \), classify as skin
- if \( p(\text{skin} | x) < \theta \), classify as not skin

Brighter pixels \( \rightarrow \) higher probability of being skin

---

Example: classifying skin pixels

Gary Bradski, 1998

---

Example: classifying skin pixels

Figure 6: A video image and its flesh probability image

Figure 7: Orientation of the flesh probability distribution marked on the source video image

Gary Bradski, 1998

---

Example: classifying skin pixels

Figure 1: CAMSHIFT-based face tracker used to overlay 3D graphic's model of Hawaii

Using skin color-based face detection and pose estimation as a video-based interface

Gary Bradski, 1998
Supervised classification

- Want to minimize the expected misclassification
- Two general strategies
  - Use the training data to build representative probability model; separately model class-conditional densities and priors (generative)
  - Directly construct a good decision boundary, model the posterior (discriminative)

Coming up

Face detection
Categorization with local features and part-based models
Deep convolutional neural networks

Questions?

See you Tuesday!