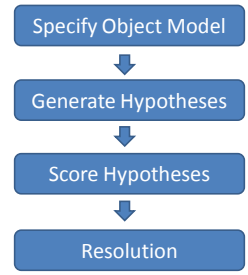
EEE422/6082 Computational Vision

Keypoint-based Recognition

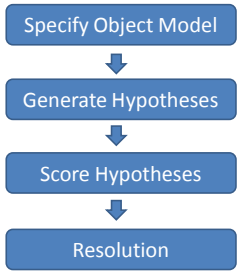
Ling Shao

Many slides from Derek Hoiem

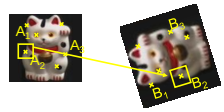
General Process of Object Recognition



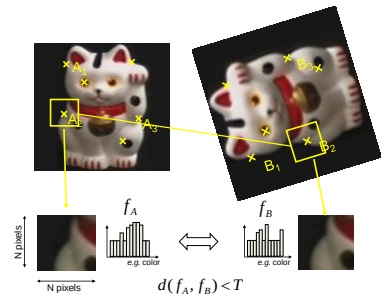
General Process of Object Recognition



Example: Keypoint-based Instance Recognition



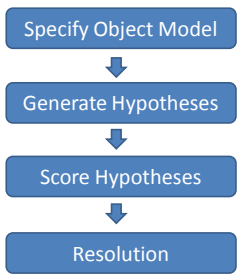
Overview of Keypoint Matching



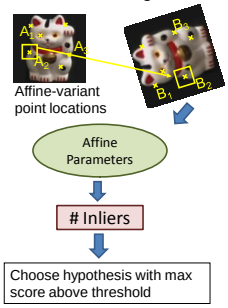
1. **Find a set of distinctive key-points**
2. **Define a region around each keypoint**
3. **Extract and normalize the region content**
4. **Compute a local descriptor from the normalized region**
5. **Match local descriptors**

K. Grauman, B. Leibe

General Process of Object Recognition



Example: Keypoint-based Instance Recognition

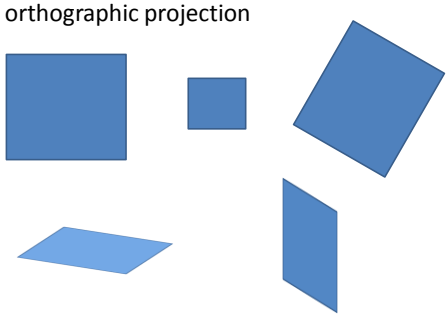


Matching Keypoints

- Want to match keypoints between:
  1. Training images containing the object
  2. Database images
- Given descriptor x_0, find two nearest neighbors x_1, x_2 with distances d_1, d_2
- x_1 matches x_0 if $d_1/d_2 < 0.8$
  - This gets rid of 90% false matches, 5% of true matches in Lowe's study

Affine Object Model

- Accounts for 3D rotation of a surface under orthographic projection



Affine Object Model

- Accounts for 3D rotation of a surface under orthographic projection

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Scaling/skew Translation

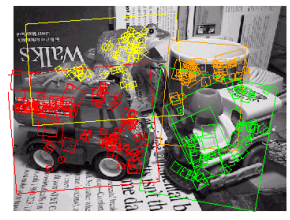
$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u \\ v \\ \vdots \end{bmatrix}$$

How many matched points do we need?

Finding the objects

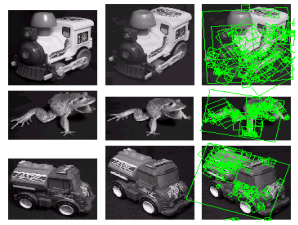
1. Get matched points in database image
2. Get location/scale/orientation
3. Geometric verification
4. Report object if > T inliers (T is typically 3, can be computed by probabilistic method)

Matched objects



View interpolation

- Training
  - Given images of different viewpoints
  - Cluster similar viewpoints using feature matches
  - Link features in adjacent views
- Recognition
  - Feature matches may be spread over several training viewpoints
  - ⇒ Use the known links to "transfer votes" to other viewpoints



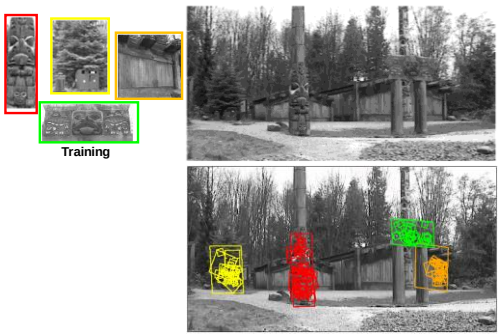
[Lowe01]

Slide credit: David Lowe

Applications

- Sony Aibo (Evolution Robotics)
- SIFT usage
  - Recognize docking station
  - Communicate with visual cards
- Other uses
  - Place recognition
  - Loop closure in SLAM

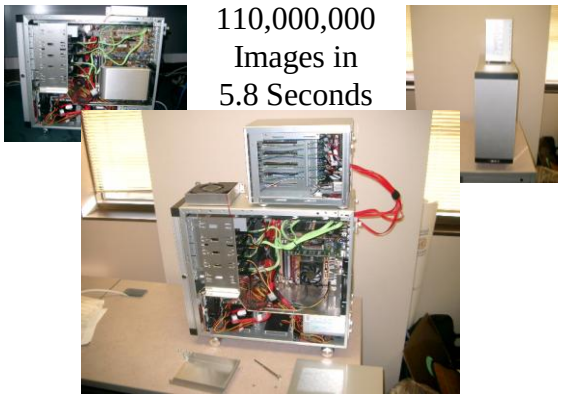
Location Recognition



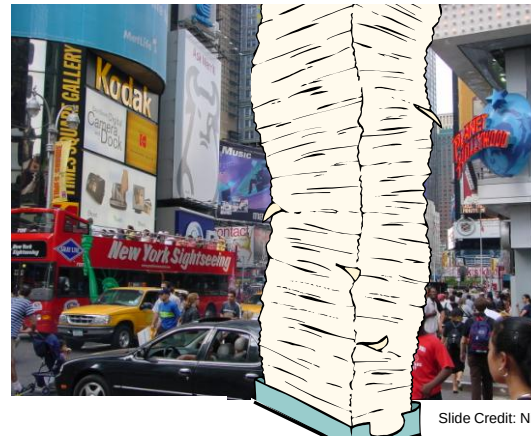
[Lowe04]
Slide credit: David Lowe

Fast visual search

"Video Google", Sivic and Zisserman, ICCV 2003
"Scalable Recognition with a Vocabulary Tree", Nister and Stewenius, CVPR 2006.

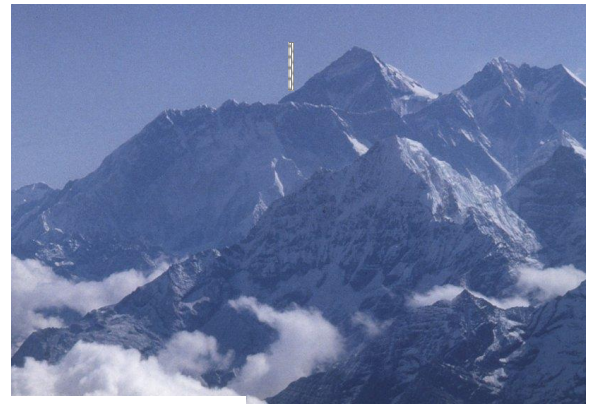


Slide Credit: Nister



Slide Credit: Nister

Slide Credit: Nister



Slide Credit: Nister

Key Ideas

- Visual Words
  - Cluster descriptors (e.g., K-means)
- Inverse document file
  - Quick lookup of files given keypoints

tf-idf: Term Frequency – Inverse Document Frequency

times word appears in document

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

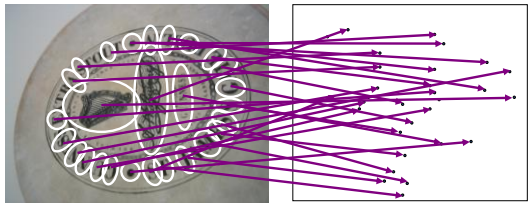
documents

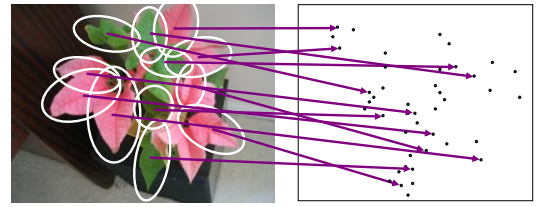
documents that contain the word

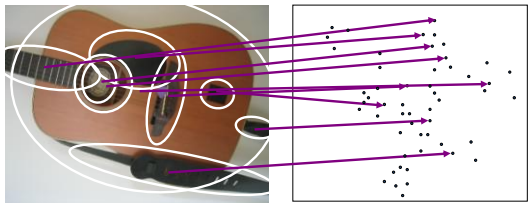
words in document

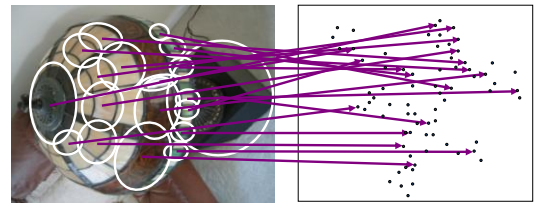
Recognition with K-tree

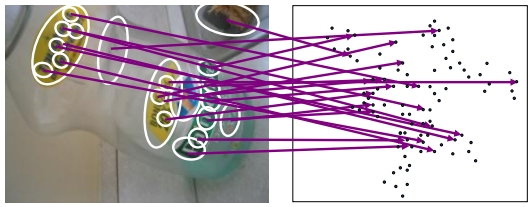
Following slides by David Nister (CVPR 2006)

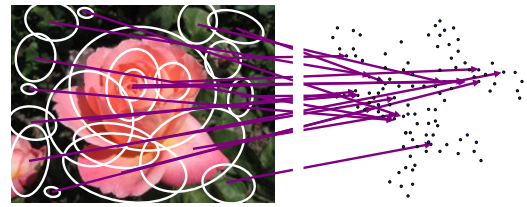


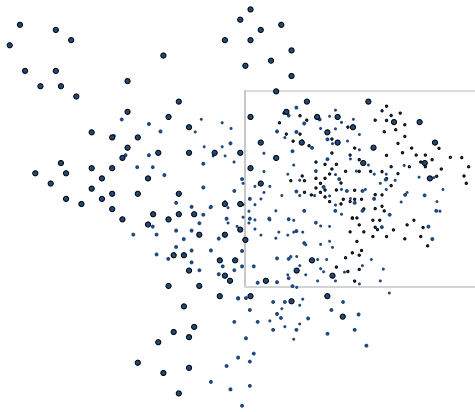


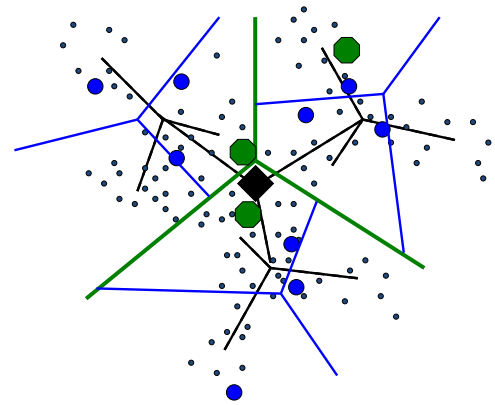


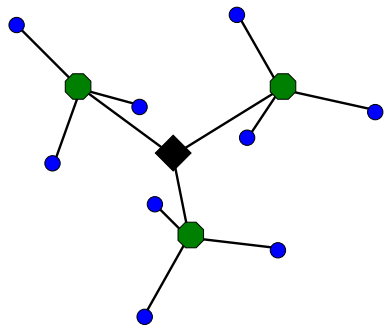


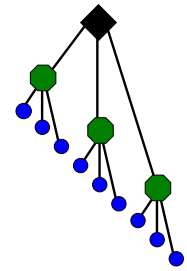


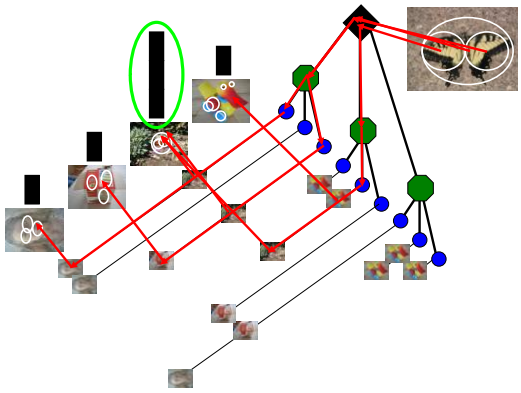




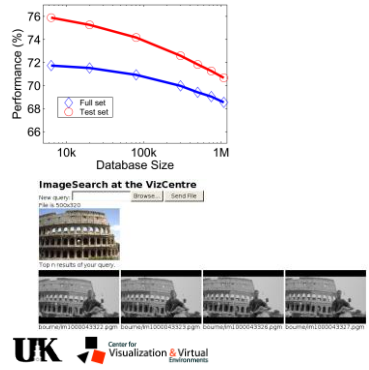


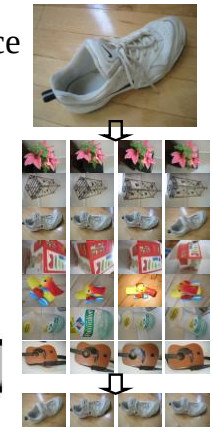






Performance



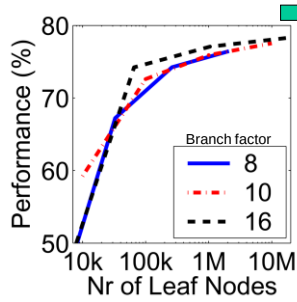


UK Center for Visualization & Virtual Environments

More words is better

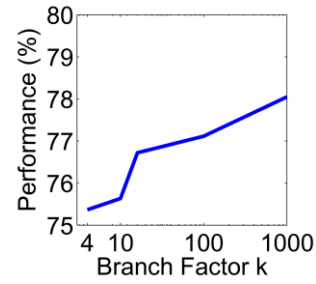
Improves Retrieval

Improves Speed



UK Center for Visualization & Virtual Environments

Higher branch factor works better (but slower)



UK Center for Visualization & Virtual Environments

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

36 K. Grauman, B. Leibe

Application: Large-Scale Retrieval

Query

Results on 5K (demo available for 100K)

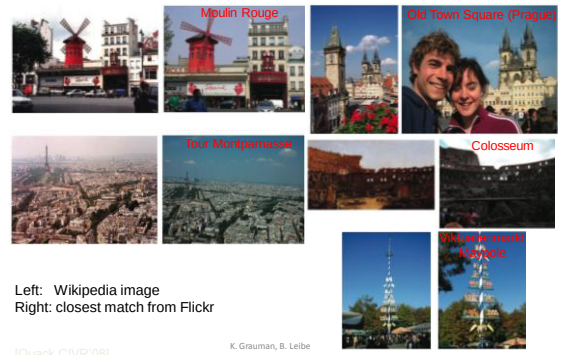
[Philbin CVPR'07]

38 K. Grauman, B. Leibe

Example Applications

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

Application: Image Auto-Annotation



Left: Wikipedia image
Right: closest match from Flickr

[Quack CIVR'08]

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