

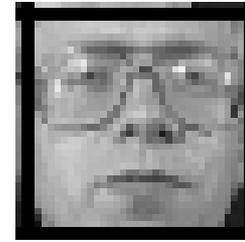
# Brief Review of Recognition + Context

Computer Vision  
CS 543 / ECE 549  
University of Illinois

Derek Hoiem

# Object Instance Recognition

- Want to recognize the same or equivalent object instance, which may vary
  - Slight deformations
  - Change in lighting
  - Occlusion
  - Rotation, rescaling, translation, perspective



=



=

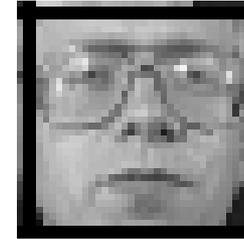


=



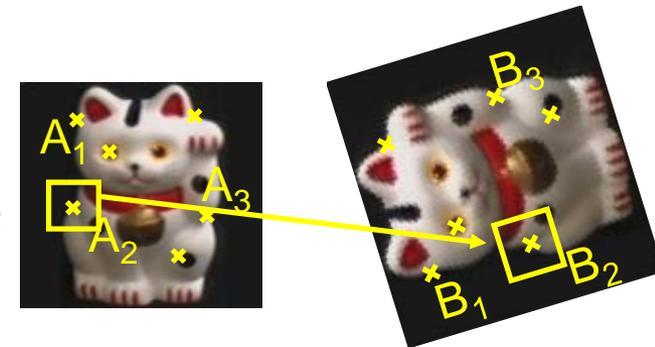
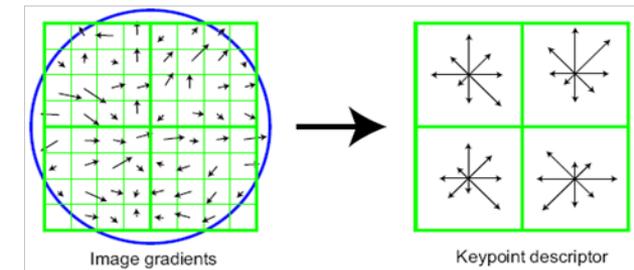
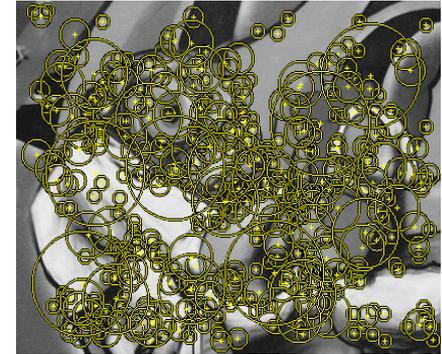
# Object Instance Recognition

- Template matching: faces
  - Recognize by directly computing pixel distance of aligned faces
  - Principal component analysis gives a subspace that preserves variance
  - Linear Discriminant Analysis (LDA) or Fisher Linear Discriminants (FLD) gives a subspace that maximizes discrimination
- This could work for other kinds of aligned objects



# Object Instance Recognition

- If object is not aligned, we need to perform **geometric matching**
  1. Find **distinctive** and **repeatable** keypoints
    - E.g., Difference of Gaussian, Harris corners, or MSER regions
  2. Represent the appearance at these points (e.g., SIFT)
  3. Match pairs of keypoints
  4. Estimate transformation (e.g., rotation, scale, translation) from matched keypoints
    - Hough voting
    - Geometric refinement
- Clustering (visual words) and inverse document frequency enable fast search in large datasets



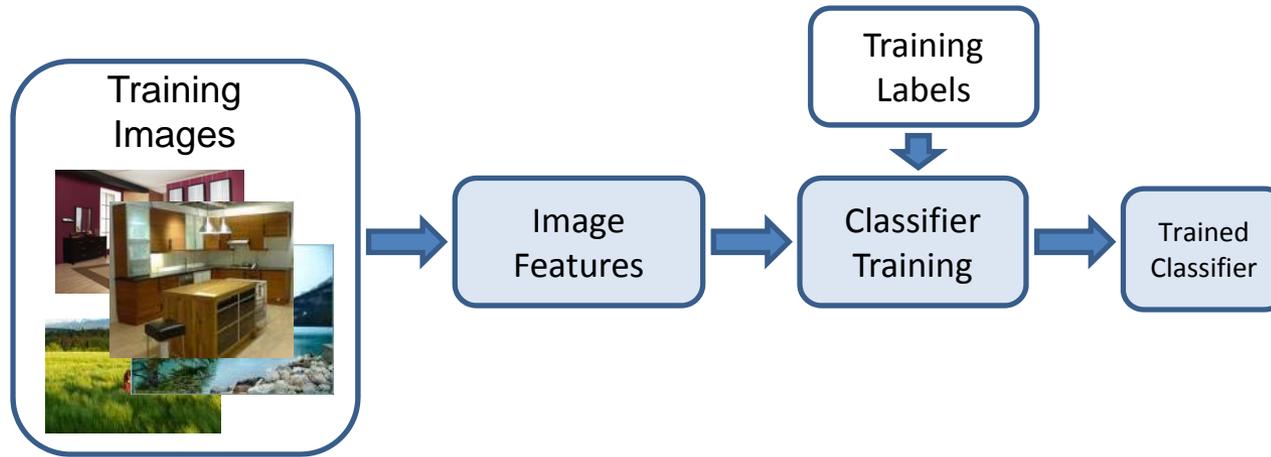
# Category recognition

- Instances across categories tend to vary in more challenging ways than a single instance across images



# Image Categorization

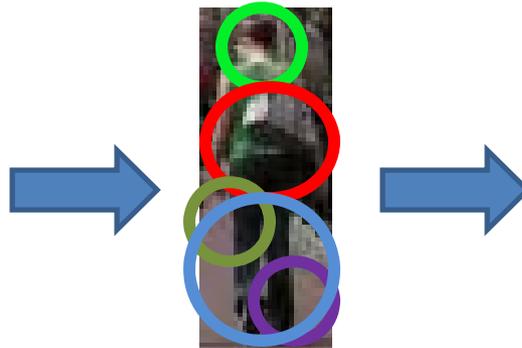
- In training, a classifier is trained for a particular feature representation using labeled examples



- The features should generally capture local patterns but with loose spatial encoding
- For scene categorization, a reasonable choice is often
  1. Compute visual words (detect interest points, represent them with SIFT, and cluster)
  2. Compute a spatial pyramid of these visual words, composed of histograms at different spatial resolutions
  3. Train a linear SVM classifier or one with a Chi-squared kernel

# Object Category Detection

- One difficulty of object category detection is that objects could appear at many scales or translations, and keypoint matching will be unreliable
- A simple way around this is to treat category detection as a series of image categorization tasks, breaking up the image into thousands of windows and applying a binary classifier to each
- Often, the object is classified using edge-based features whose positions are defined at fixed position in the sliding window



Object or  
Background?

# Object Category Detection

- Sliding windows might work well for rigid objects



- But some objects may be better thought of as spatial arrangements of parts



# Object Category Detection

- Part-based models have three key components
  - Part definition and **appearance** model
  - Model of **geometry** or layout of parts
  - Algorithm for efficient **search**
- ISM Model
  - Parts are clustered detected keypoints
  - Position of each part wrt object center/size is recorded
  - Search is done through Hough voting / Mean-shift clustering combination
- Pictorial structures model
  - Parts are rectangles detected in silhouette
  - Layout is articulated model with tree-shaped graph
  - Search through dynamic programming or probabilistic sampling

# Region-based recognition

- Sometimes, we want to label image pixels or regions
- Basic approach:
  - Segment the image into blocks, superpixels, or regions
  - Represent each region with histograms of keypoints, color, texture, and position
  - Classify each region (variety of classifiers used)



# Context in Recognition

- Objects usually are surrounded by a scene that can provide context in the form of nearby objects, surfaces, scene category, geometry, etc.



# Context provides clues for function

- What is this?



# Context provides clues for function

- What is this?

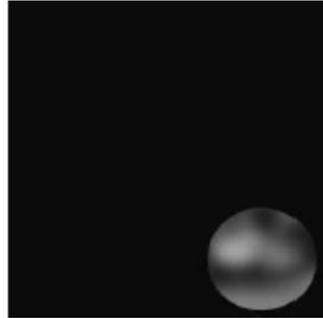


- Now can you tell?



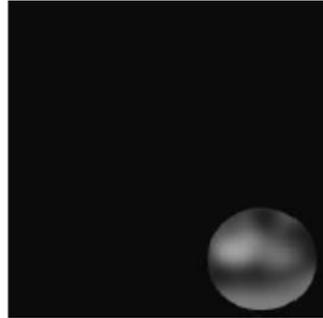
Sometimes context is *the* major component of recognition

- What is this?



Sometimes context is *the* major component of recognition

- What is this?

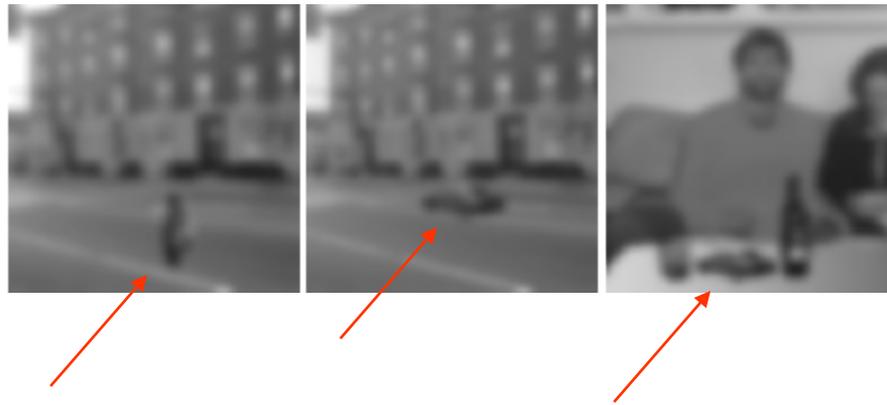


- Now can you tell?



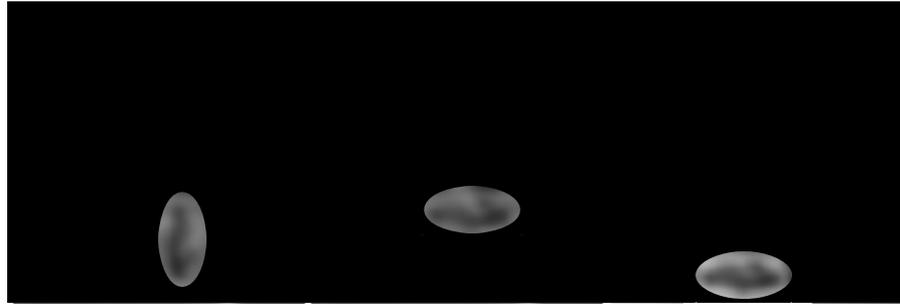
# More Low-Res

- What are these blobs?



# More Low-Res

- The same pixels! (a car)



We will see more on context later...