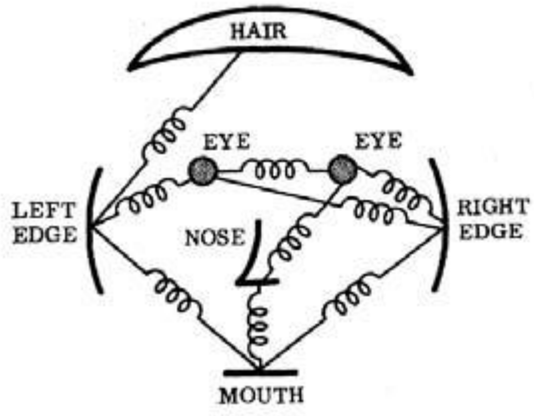
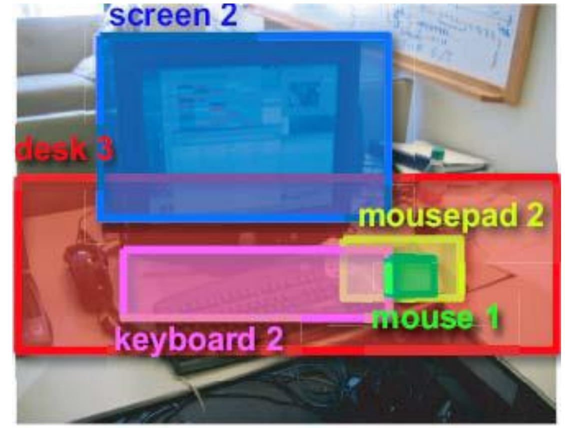


COMP 9517 Computer Vision

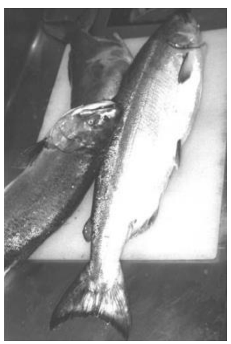
Recognition

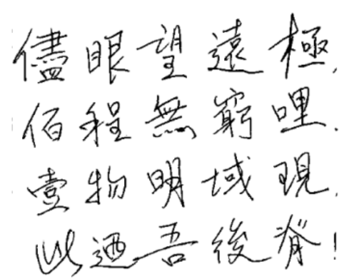
Introduction

- ***Pattern recognition*** is the scientific discipline whose goal is the classification of objects into a number of categories or classes
- Pattern recognition is used widely for object classification and recognition
- Objects can be images or any type of measurements that need to be classified, which are referred using the generic term ***pattern***

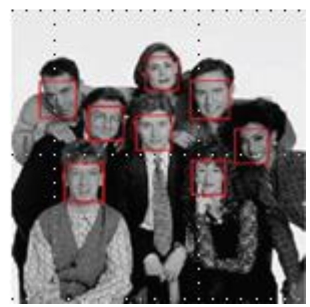
Applications

- Examples of pattern recognition in computer vision:
  - Machine vision
  - Character recognition
  - Face recognition
  - Human activity recognition





I looked as hard as I could see, beyond 100 plus infinity an object of bright intensity- it was the back of me!

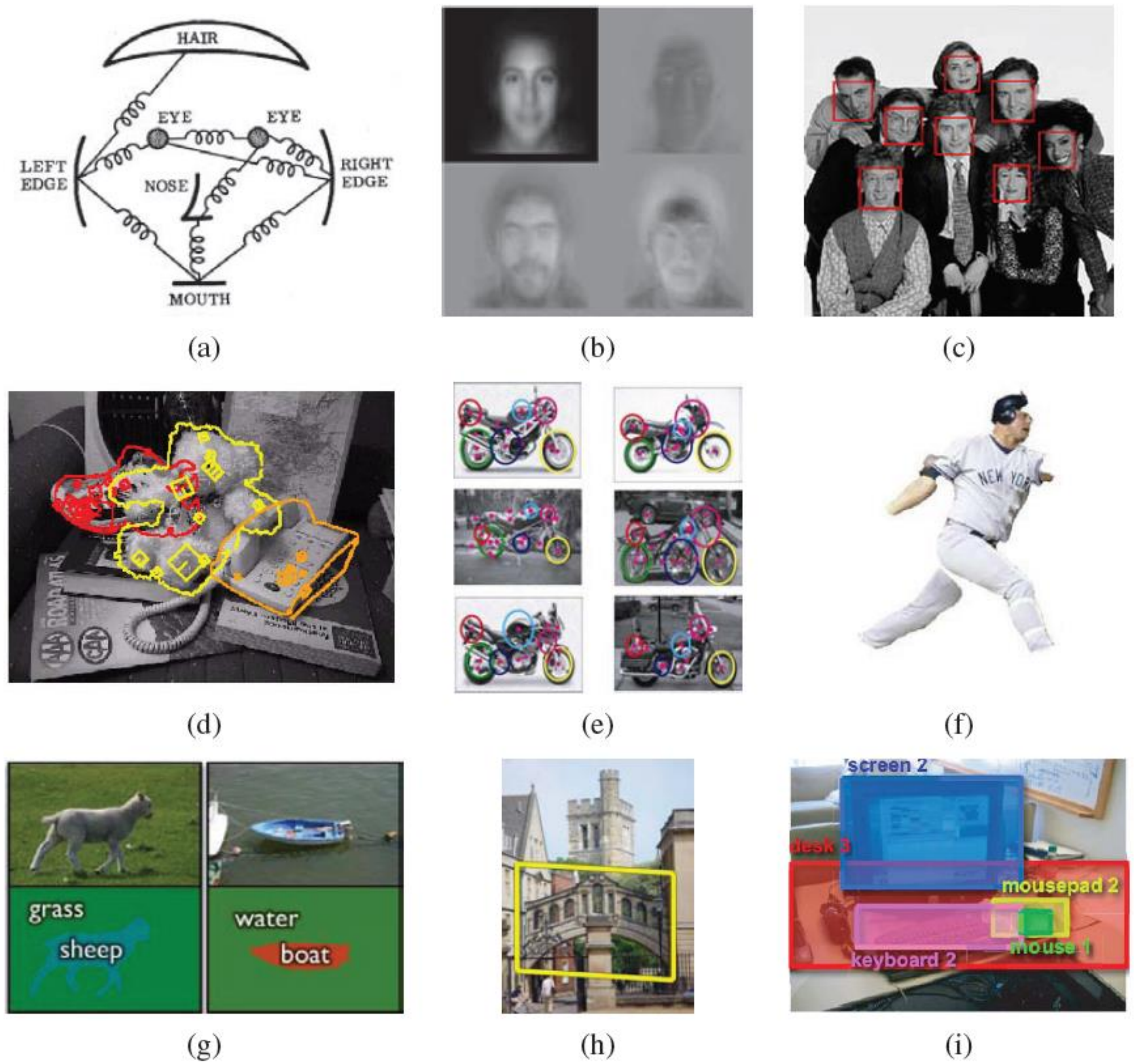


Applications

- Computer vision is an area in which pattern recognition is of importance
  - Making decisions about image content
  - Classifying objects in an image
  - Recognising activities

Images extracted from google.com

Recognition in Vision

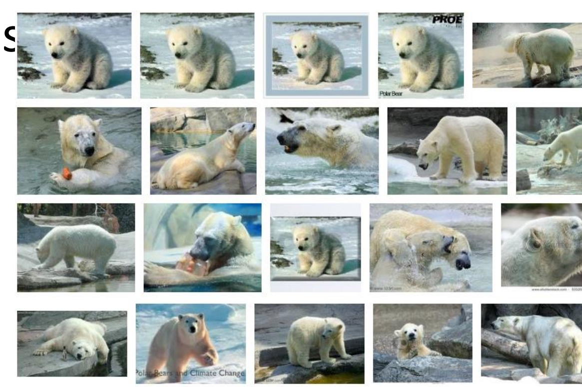


Recognition in Vision

- Recognition is hard
  - Analysing a scene and recognising all of the objects remains the most challenging of all the visual tasks
  - Computer cannot name all the objects and animals present in a picture even at the level of a two-year old child

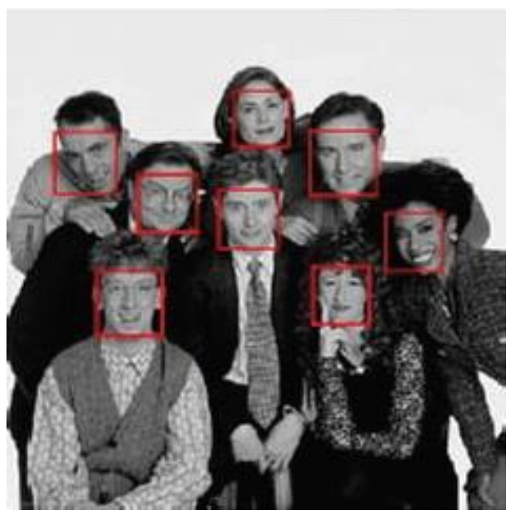
Recognition in Vision

- Why is it so hard?
  - real world made of a jumble of objects (called clutter) which all occlude one another and appear in different poses
  - Variability intrinsic within a class
    - Complex non-rigid articulation and extreme variations in shape and appearance
    - Simple exhaus e



Three Recognition Problems

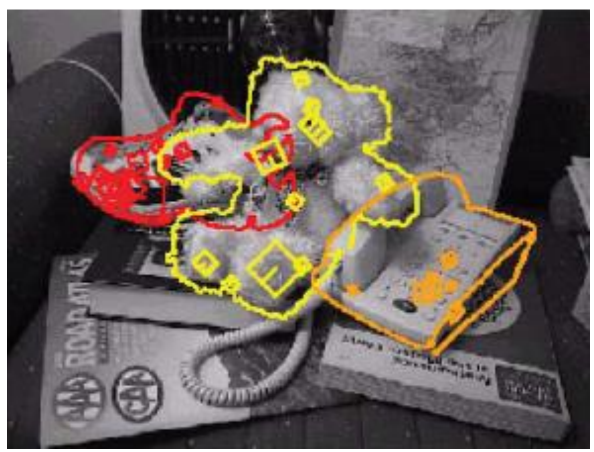
1. Object detection
   – know what we are looking for
   – quickly scan an image to determine where a match may occur



Three Recognition Problems

2. Instance recognition

- have a specific rigid object to recognise
- search for characteristic feature points and verify that they align in a geometrically plausible way



Three Recognition Problems

3. General category / class recognition

- recognising instances of extremely varied classes, such as animals or furniture
- Techniques rely on
  - presence of features (“bag of words” model)
  - their relative positions (part-based models)
  - segmenting the image into semantically meaningful regions

The Recognition Problem

- Recognition depends heavily on the context of surrounding objects and scene elements
- Woven into all of these techniques is the topic of *learning*
  - study and construction of algorithms that can learn from and make predictions on data

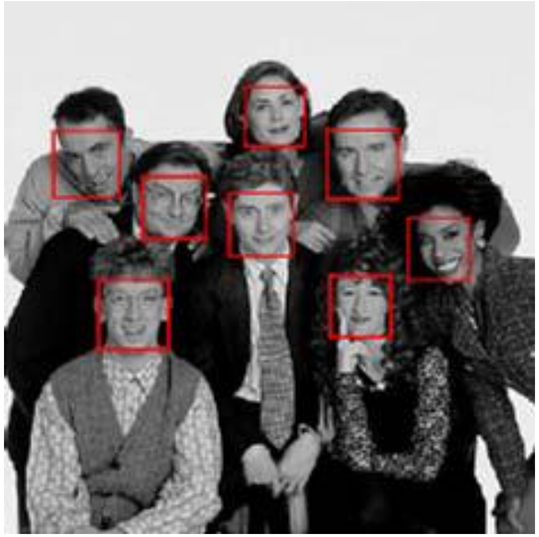
Object Detection

Approaches

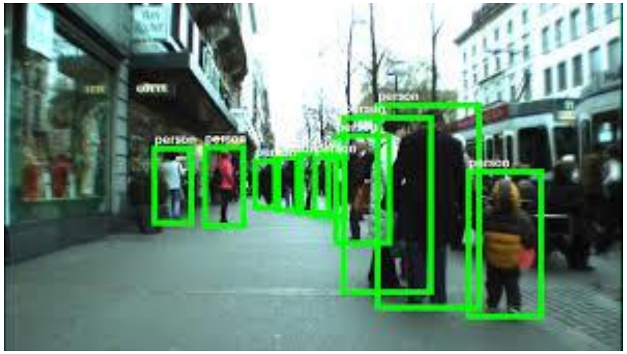
- apply a recognition algorithm to every possible sub-window in the image
  - slow
  - error-prone
- instead, construct special-purpose detectors
  - to rapidly find likely regions where particular objects might occur

Special-purpose Detectors

- Face detector
  - built into most of today's digital cameras to enhance auto-focus
  - built into video conferencing systems to control pan-tilt heads



- Pedestrian detectors
  - use more general methods for object detection
  - used to detect pedestrians and other cars from moving vehicles

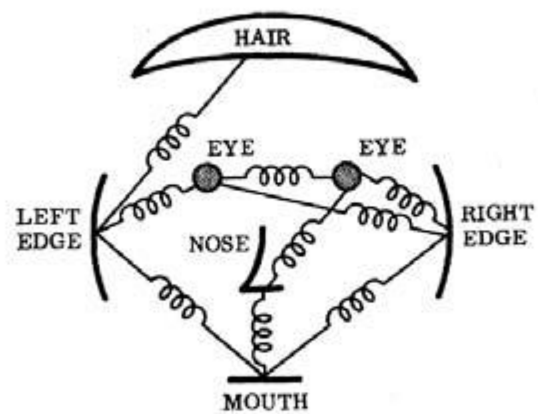


Face Detection

- Find locations and sizes of any faces in image first
  - apply a face recognition algorithm at every pixel and scale
    - Too slow
  - fast face detection algorithms
    - Feature-based
    - Template-based
    - Appearance-based

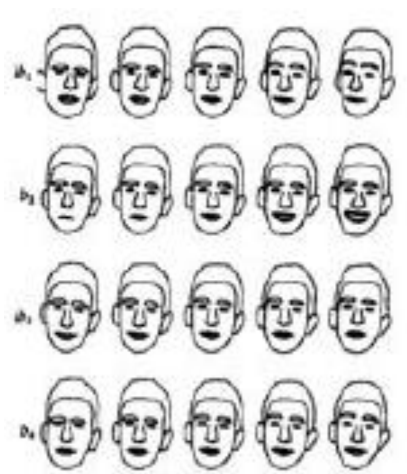
Face Detection

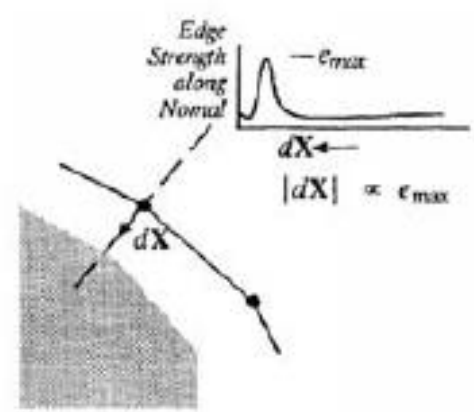
- Approaches
  - Feature based:
    - find the locations of distinctive image features, such as the eyes, nose, and mouth
    - verify whether these features are in a plausible geometrical arrangement



Face Detection

- Approaches
  - Template-based (eg AAM):
    - match shape and pose with model
    - can deal with a wide range of pose and expression variability
    - require good initialization near a real face
    - not suitable as fast face detectors



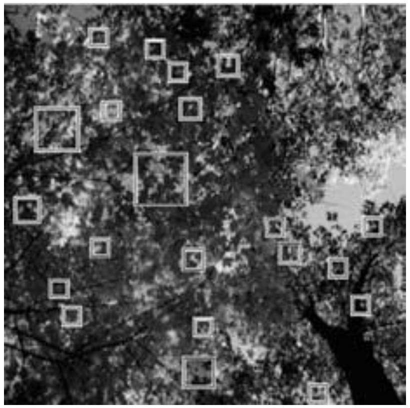


Appearance-based Face Detection

- scan over small overlapping rectangular patches of the image
- searching for likely face candidates
- Refine using cascade of more expensive but selective detection algorithms

Appearance-based Face Detection

– collect training images
  - a set of labelled face patches
  - a set of patches taken from images that are known not to contain faces
  - collected face images are augmented by artificially mirroring, rotating, scaling, and translating the images by small amounts

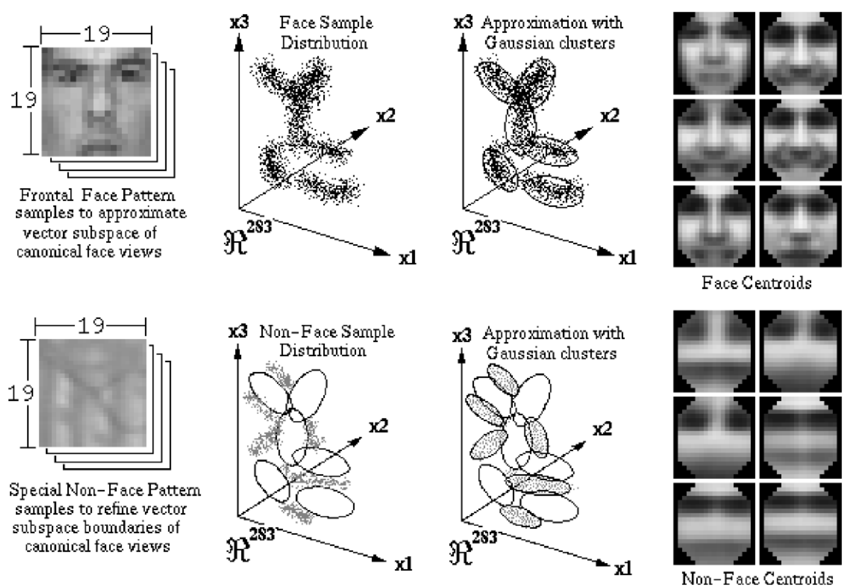


Appearance-based Face Detection

– Pre-processing
  - subtract an average gradient (linear function) from the image to compensate for global shading effects
  - using histogram equalization to compensate for varying camera contrast

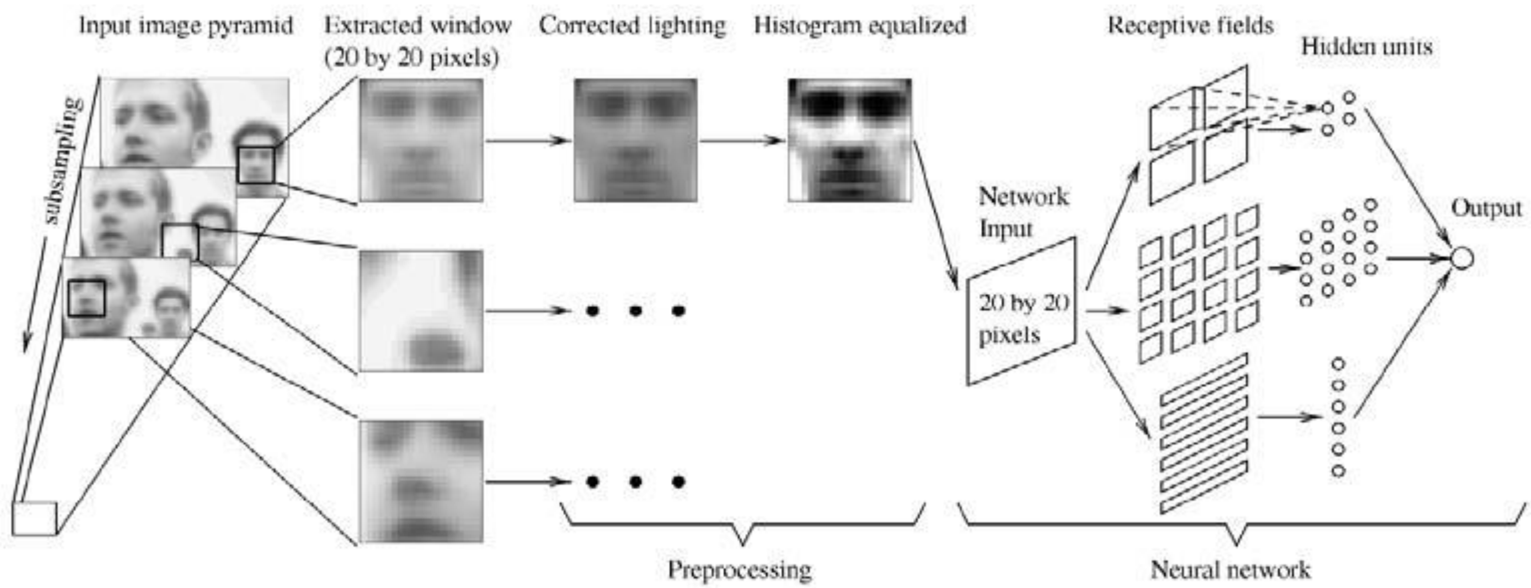
Appearance-based Face Detection

- Clustering and PCA
  - Cluster the datasets using k-means
  - fit PCA subspaces to each of the resulting clusters
  - Mahalanobis distance measurements of sample to each cluster input to a multi-layer perceptron to perform classification



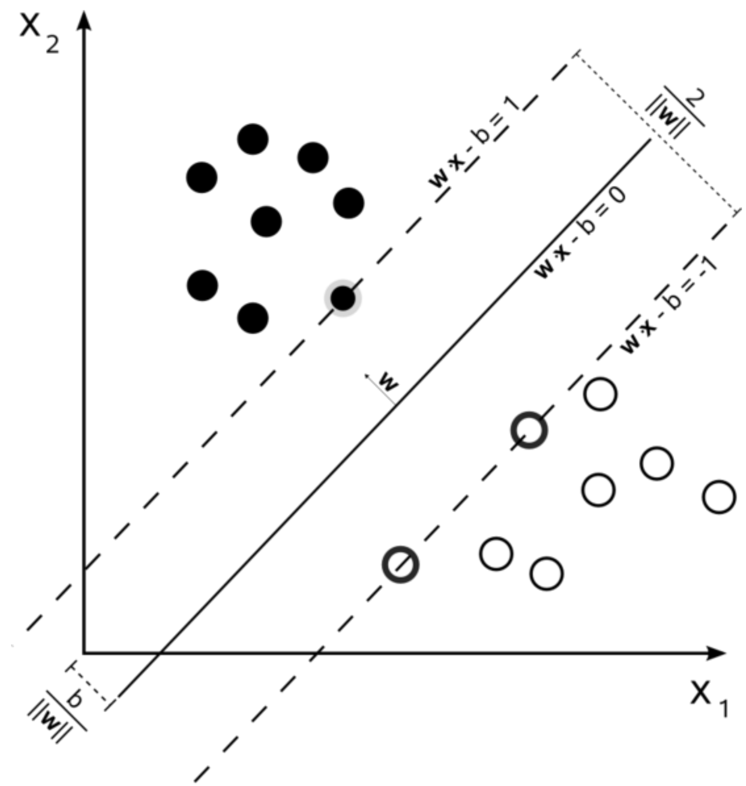
Appearance-based Face Detection

- Neural networks
  - apply multi-layer perceptron directly to patches of gray-level intensities to perform classification



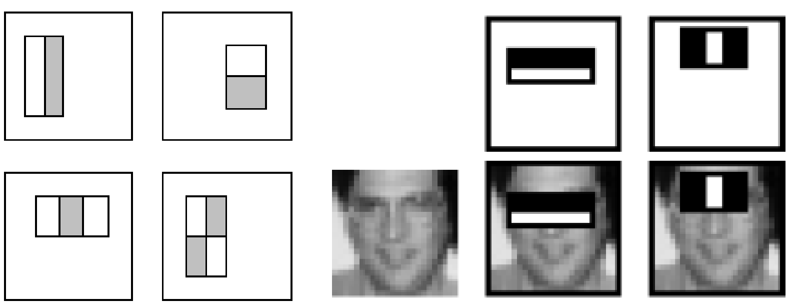
Appearance-based Face Detection

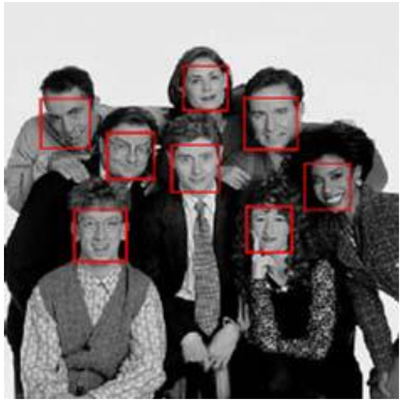
- Support vector machines
  - a support vector machine instead of neural network to classify pre-processed patches



Appearance-based Face Detection

- Boosting (Viola and Jones approach)
  - best known and most widely used approach of all face detectors currently in use
  - first to introduce the concept of boosting to computer vision community
    - train a series of increasingly discriminating simple classifiers and then blending their outputs – eg diff of rectangle features below



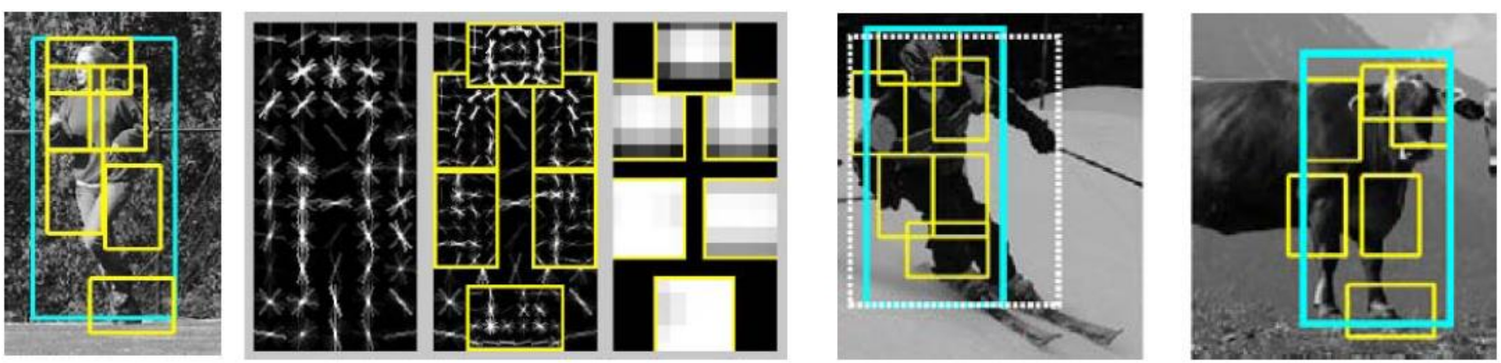


Pedestrian Detection using HOG

- due to Dalal and Triggs, 2005
  - Compute histogram of oriented gradients (HOG) descriptors
  - Each HOG has cells to accumulate magnitude-weighted votes for gradients at particular orientations- similar to SIFT, with some differences- evaluated on overlapping grid, at single scale and fixed orientation
  - Train SVM classifier using the descriptors as input
  - When video sequences are available, the additional information present in the optic flow and motion discontinuities can greatly aid in the detection task

Pedestrian Detection using Parts-based models

- extend HOG person detector to incorporate flexible parts models
- each part trained and detected on HOGs evaluated at two pyramid levels below the overall object model
- locations of the parts relative to the parent node are learned and used during recognition



Part-based Detection for people, bicycles, horses

Face Recognition

- Used in variety of applications: HCI, identity verification, login, patient monitoring
- Works best for full frontal images under uniform illumination
- Early approach found locations of facial features (eyes, nose, mouth), then measured distances between their locations

Face Recognition by Eigenfaces

- Compare gray-level images projected onto lower dimensional subspaces called eigenfaces
- Face image **x** can be compressed and reconstructed starting with mean image **m** and adding small number of scaled signed images **u** which are derived by PCA:

$$\tilde{\boldsymbol{x}} = \boldsymbol{m} + \sum_{i=0}^{M-1} a_i \boldsymbol{u}_i,$$

Face Recognition By Active Appearance Models

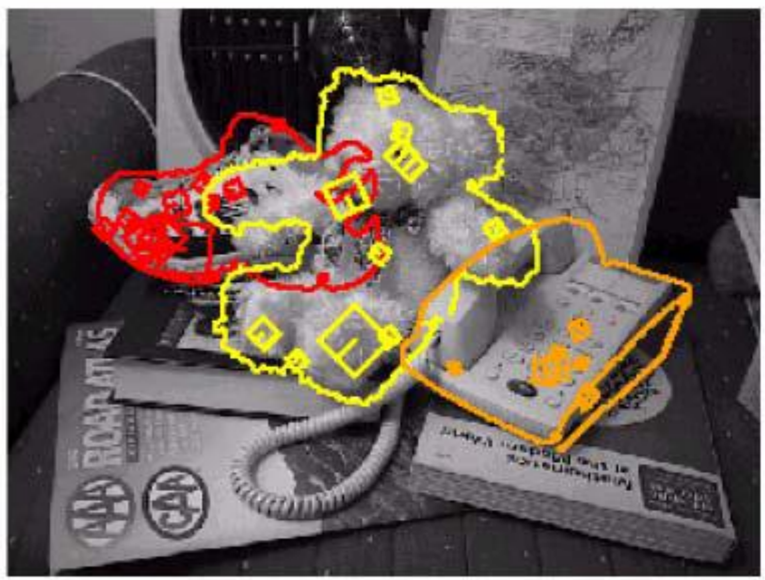
– Model the variation in the shape of an image **s**, which is normally encoded by the location of key feature points on the image, as well as variation in texture **t**
– Image normalized to a canonical shape before being analysed
– Shape and texture represented as deviations from mean shape s and texture t:

$$\begin{aligned} \boldsymbol{s} &= \bar{\boldsymbol{s}} + \boldsymbol{U}_s \boldsymbol{a} \\ \boldsymbol{t} &= \bar{\boldsymbol{t}} + \boldsymbol{U}_t \boldsymbol{a}, \end{aligned}$$

- fit an active appearance model to a novel image for recognition

Instance Recognition

- Instance recognition
  - Re-recognising a known 2D or 3D rigid object, potentially being viewed from a novel viewpoint, against a cluttered background, and with partial occlusions



Instance Recognition Approaches

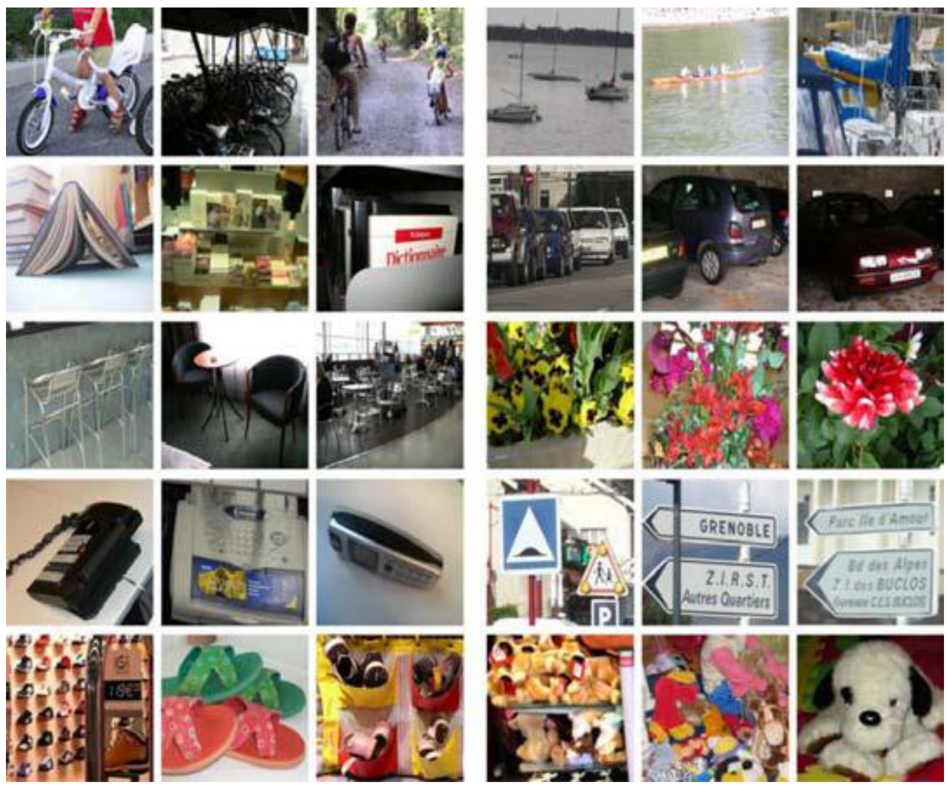
– Earlier approaches focused on extracting lines, contours or 3D surfaces from images and matching them to known 3D object models
– Acquire images from a large set of viewpoints and illuminations and represent them using an eigenspace decomposition
– Recent approaches tend to use viewpoint-invariant 2D features and match features of new image against those in the object database

Category Recognition

- Recognise instance of generic class, eg cat, car
- Generic category recognition is still a largely unsolved problem
- Computer cannot perform at the level of a two-year-old child
- Approaches
  - part-based representations and recognition
  - bag of features that represent objects and images as unordered collections of feature descriptors
  - simultaneously segmenting images while recognising objects

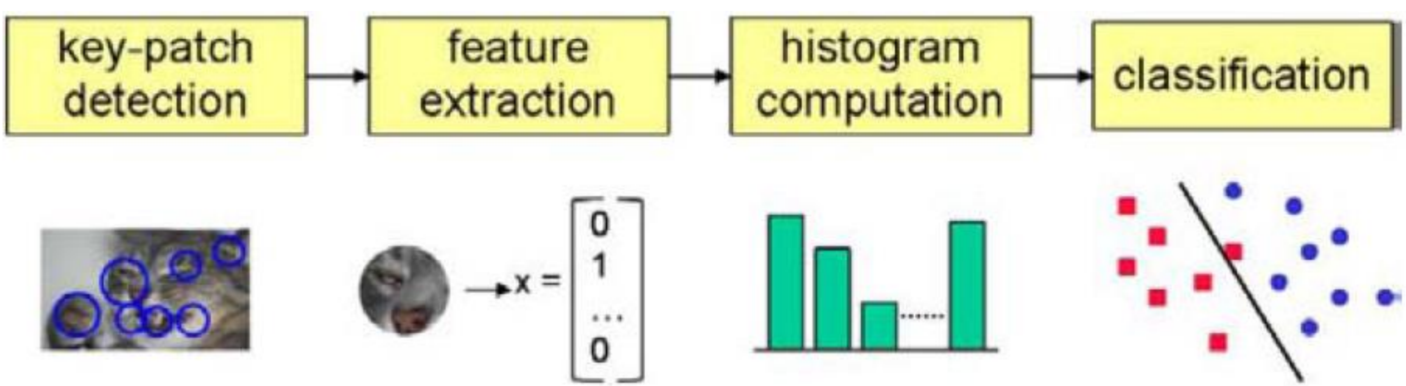
Category Recognition

- How to recognise each of these images?
  - visual category recognition is an extremely challenging problem
  - no system that approaches the performance level of a two year-old child



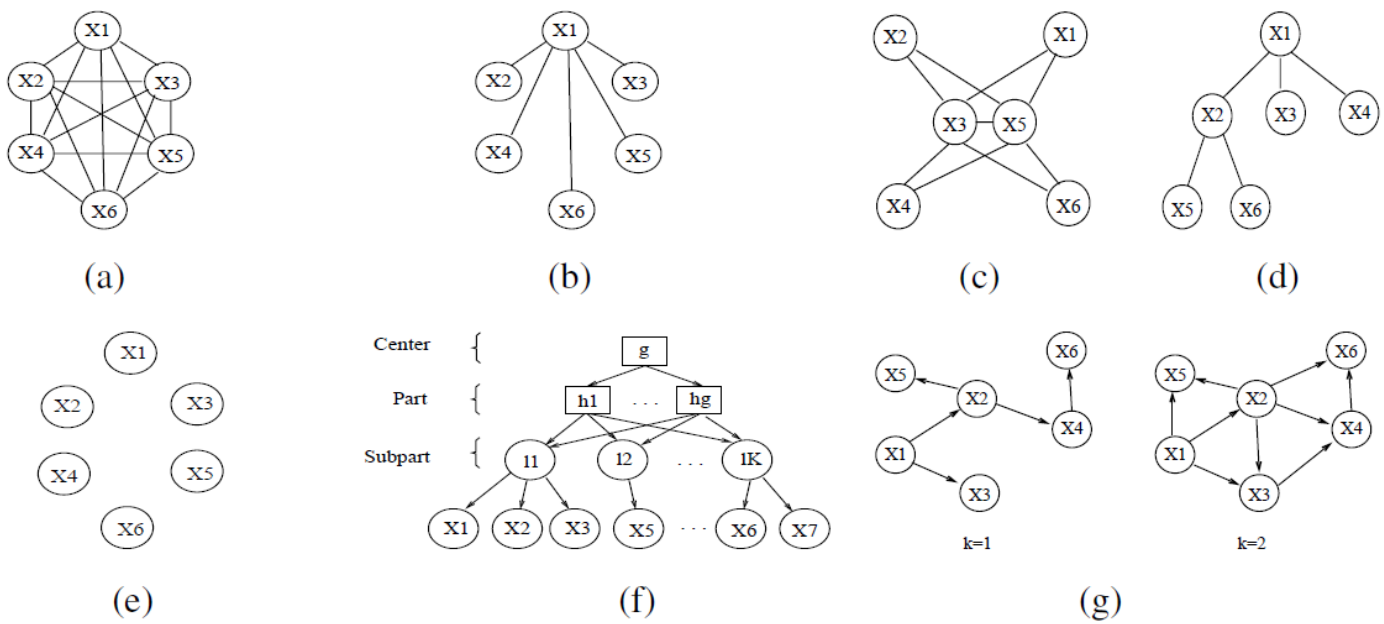
Category Recognition by Bag of words

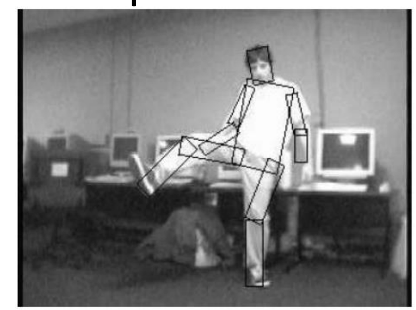
– simply computes distribution (histogram) of visual words found in the query image
  - Eg, affine covariant regions and SIFT descriptors
  - k-means visual vocabulary construction
– compares this distribution to those found in the training images
  - naive Bayesian classifier and support vector machines for classification

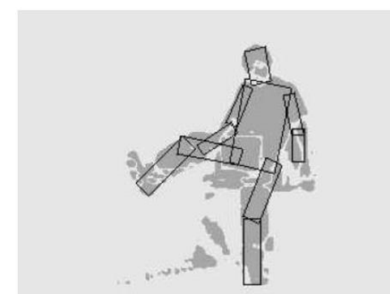


Category Recognition by Parts-based models

- Find constituent parts and measure their geometric relationships
- Requires representation of individual parts and geometric relationships, and algorithms for learning the descriptions and recognizing them
- Eg, Pictorial structures to representing geometric relationships

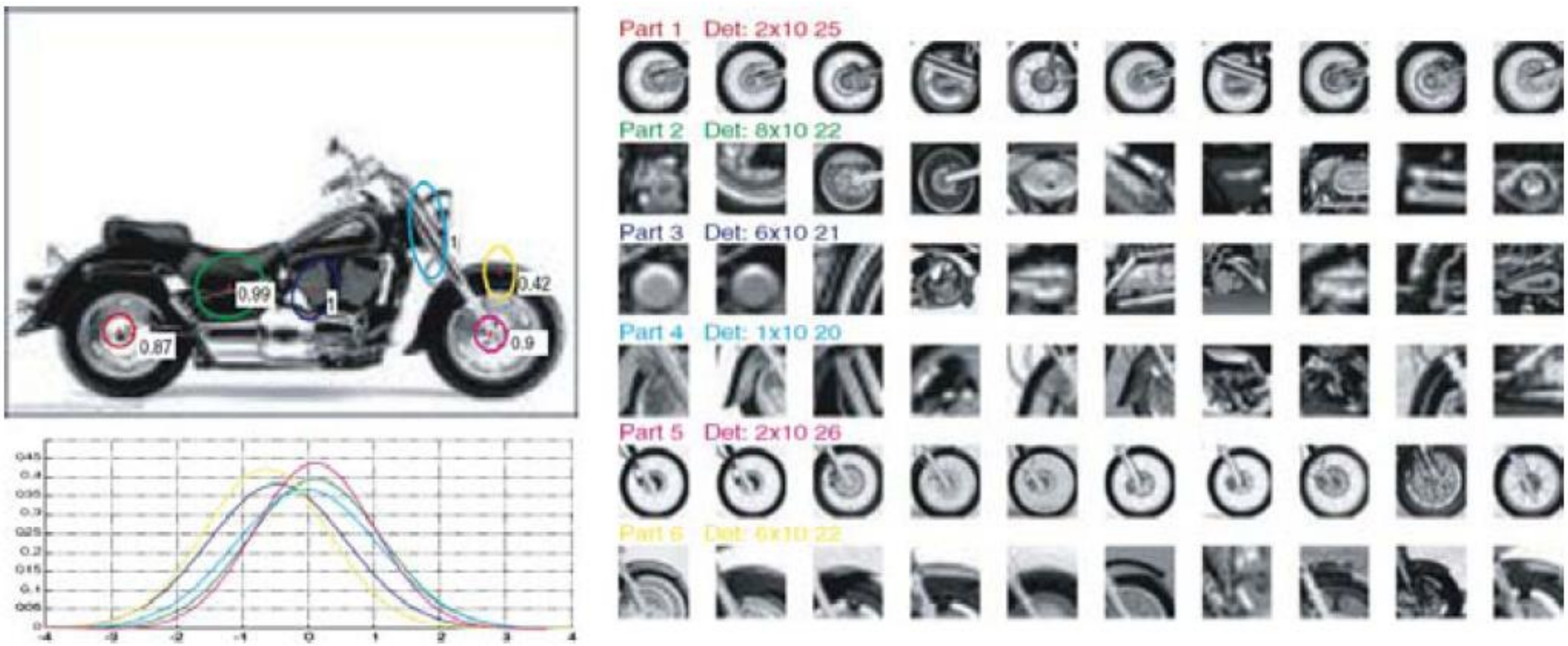






a)constellation; b)star; c)K-fan; d)tree; e)bag of words; f)hierarchy; g)sparse flexible model

Category Recognition by Parts-based models

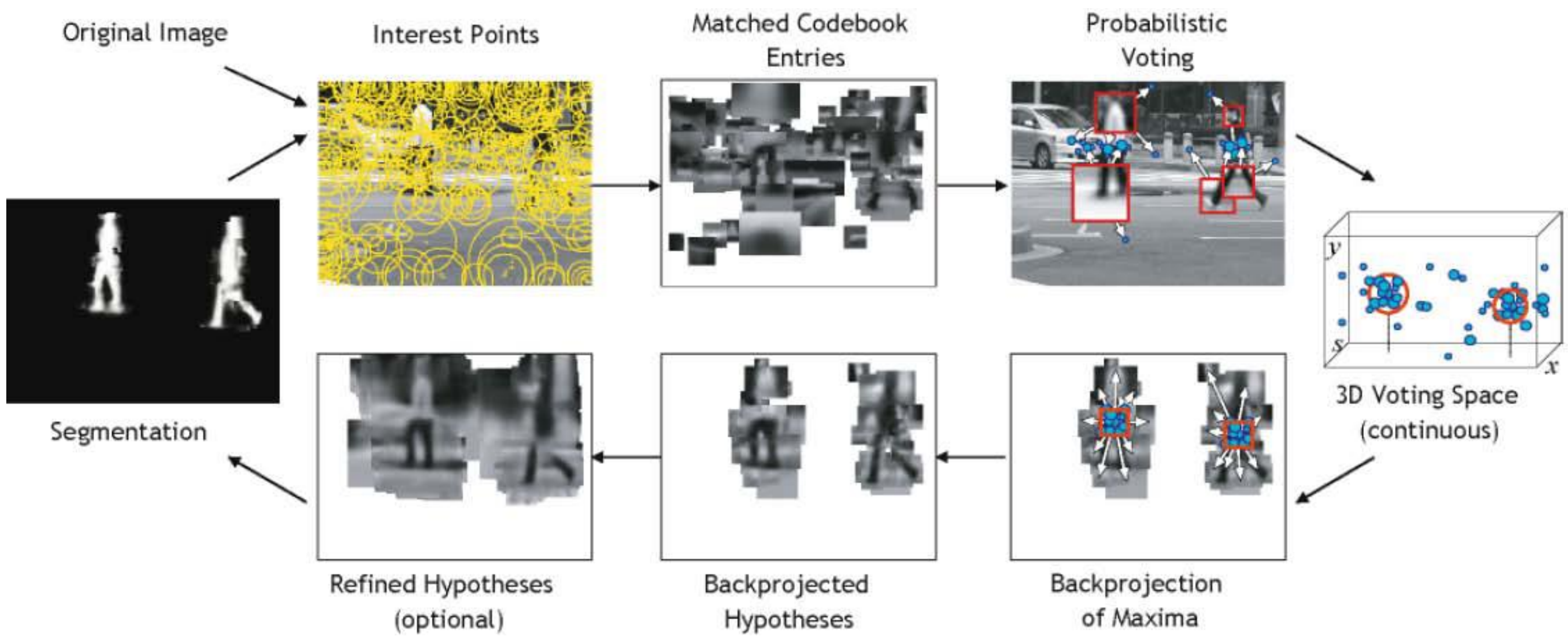


Category Recognition

- Part-based models

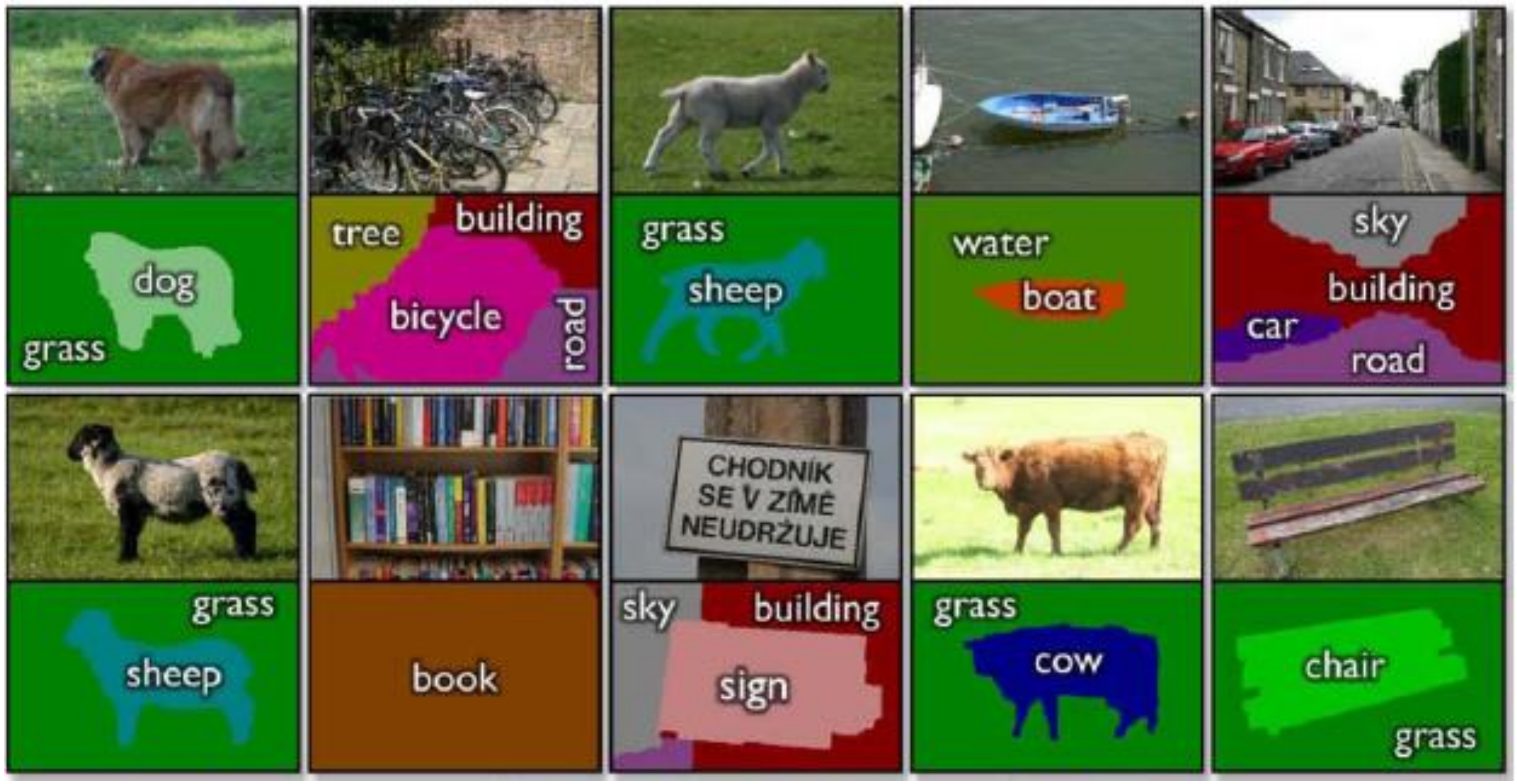
Category Recognition with Segmentation

- simultaneously perform recognition with accurate boundary segmentation
- One approach is detection of features corresponding to pre-clustered visual codebook entries



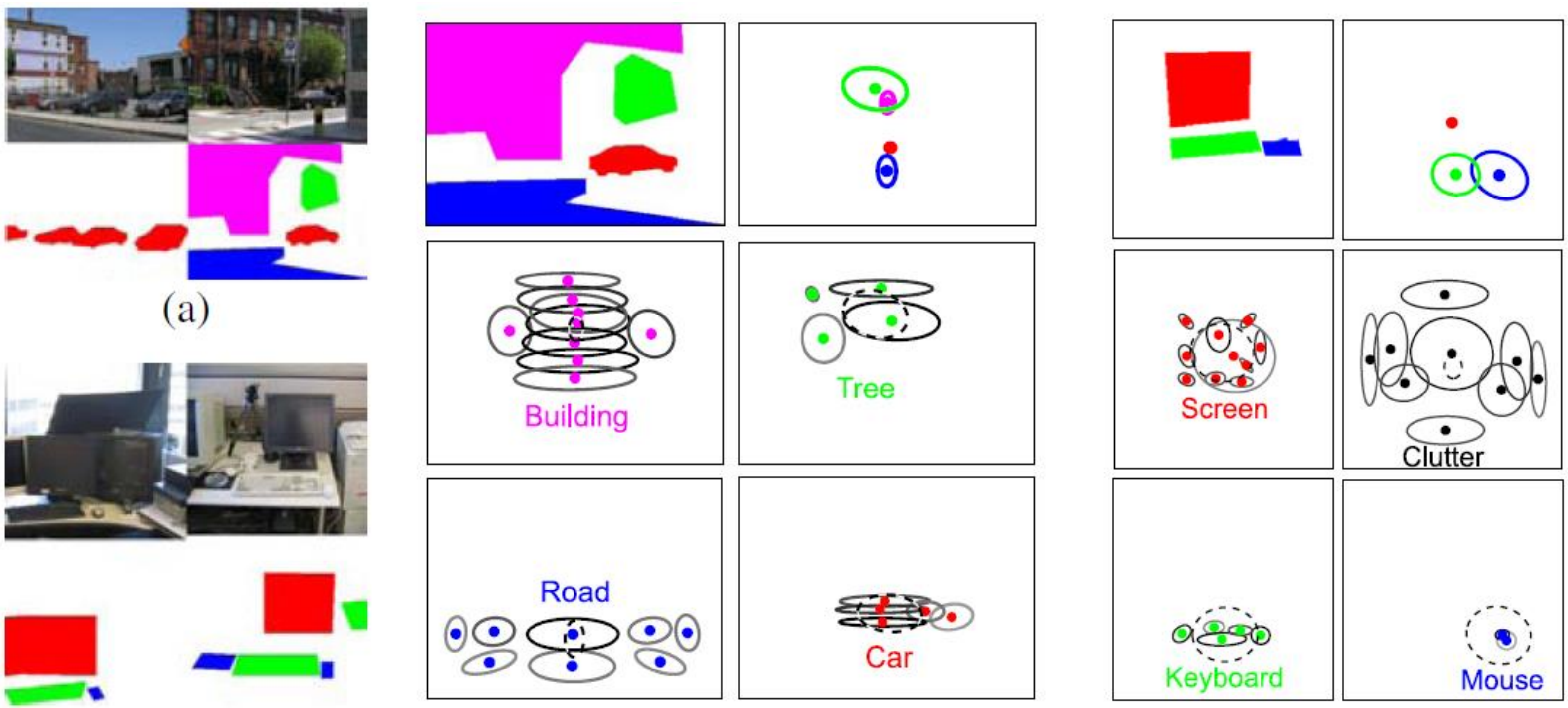
Category Recognition with segmentation

– Another approach: label every pixel in an image with its class membership



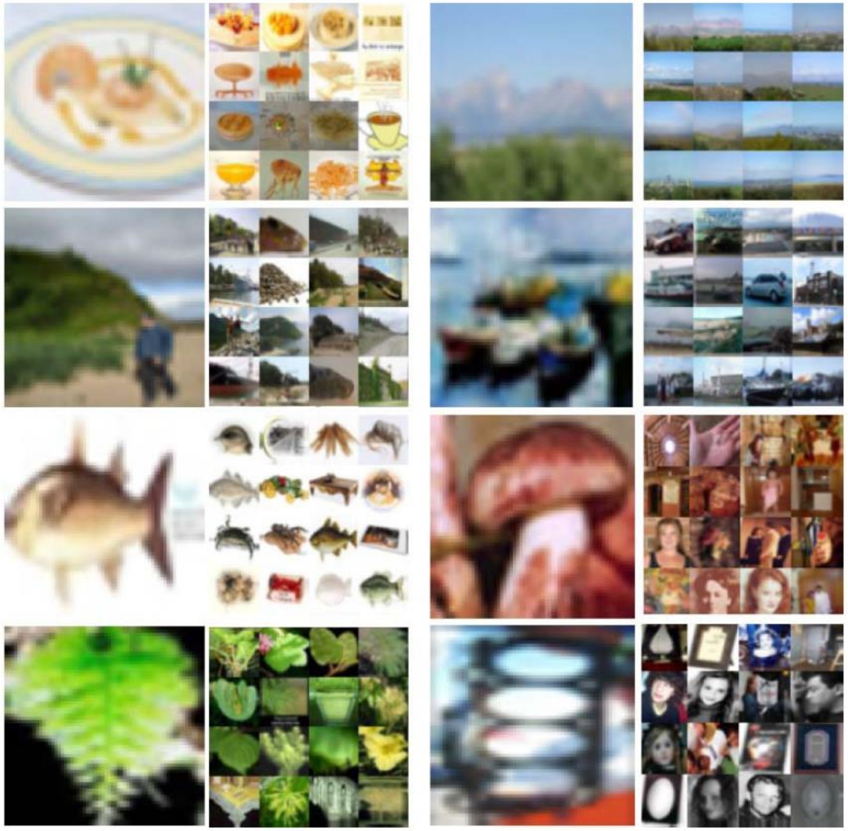
Context, Scene and Image Databases

- Context plays very important role in human object recognition
- Geometric model for describing their relative positions



Context, Scene and Image Databases

- Directly match complete images
  - matching directly against the training images rather than using them to learn the parameters of recognition algorithms



That takes us to: Imagenet!

- Deep learning based recognition, 2012!
- Covered to some extent in Deep Learning lecture

References and Acknowledgements

- Shapiro and Stockman, Chapter 4
- Duda, Hart and Stork, Chapter 1
- Richard Szeliski， Chapter 14
- Some diagrams are extracted from the above resources