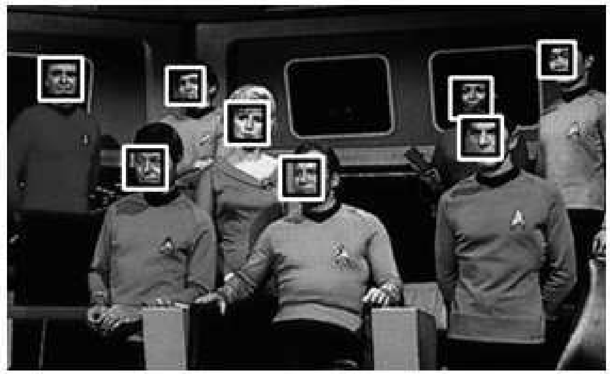
Object Detection and Recognition

Object detection and recognition are two important computer vision tasks.

- Object detection determines the presence of an object and/or its scope, and locations in the image.
- Object recognition identifies the object class in the training database, to which the object belongs to.
- Object detection typically precedes object recognition. Object detection can be treated as a binary classification problem, where one class represents the object class and another class represents non-object class.
- Object recognition can be formulated as a multi-class classification, one class for each object category.

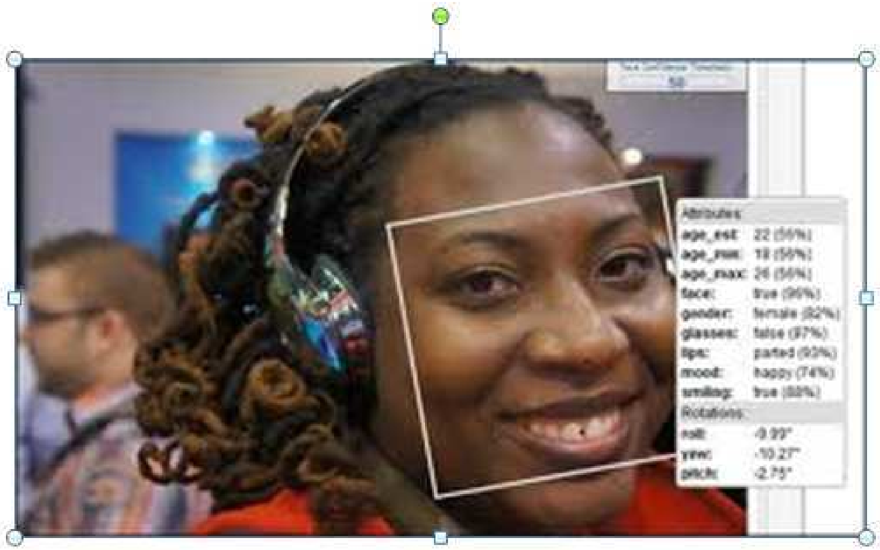
Face Detection



Face detection from image

Face detection from video

Face Recognition



Object Detection

Object detection can be further divided into soft detection, which only detects the presence of an object, and hard detection, which detects both the presence and location of the object.

Soft Detection: Stripes detection

Object Detection

- Object detection starts with constructing an object model using training data. It is then carried out by searching each part of an image to localize parts, whose photometric or geometric properties match those of the target object in the training database.
- This can be accomplished by scanning the object model across an image at different locations, scales, and rotations, and a detection is declared if the similarity between the object model and the image is sufficiently high.
- The similarity between the object model and an image region can be measured by their correlation (SSD).

Object Detection (cont'd)

Two key issues in object detection:

- Object model

  The object model can be geometric, appearance or hybrid. It can also be deterministic or probabilistic.

- Object search

  The search can be brute force, optimization based search, or limited to more likely locations. Current research focuses on identifying likely image regions through region proposals generated by an attention mechanism.

Detection Performance Metrics

- detection accuracy measures

  true positive detection rate

  false positive detection rate

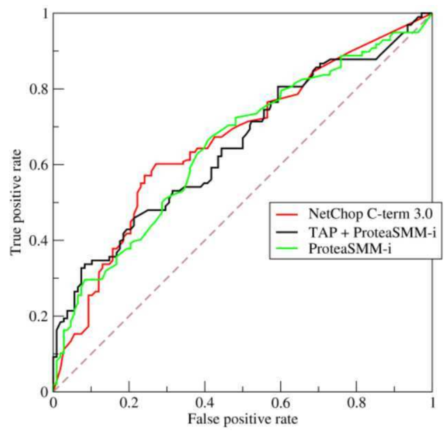
  false negative detection rate

  precision (true positive)-ratio of correct detection to total detection

  recall (sensitivity)-ratio of correct detection to actual number of target objects in the image

Detection Performance Metrics (cont'd)

- F-score=$\frac{Precision*Rcall}{Precision+Recall}$
- Receiver operating characteristic (ROC): area under ROC



- Robustness-performance under different operating and imaging conditions
- Detection speed

Object Recognition

- 2D object recognition
- 3D object recognition
- Object category recognition
- Fine-grained object recognition
- Open-set/open world object recognition

Object Category Recognition

The task is to recognize object that belongs to a particular object category such as mammal recognition (different kinds of mammals) or vehicle recognition (all kinds of vehicles) .

Fine-grained Recognition

The task is to recognize sub-category of an object such as different models/brands of a vehicle.

Open-set Recognition

The task is to not only recognize known object classes but also detect unknown object class.

Object Recognition Methods

Like object detection, object recognition starts with constructing an object model for each object class, and then compute the similarity between the object class model and an image to find the best match class.

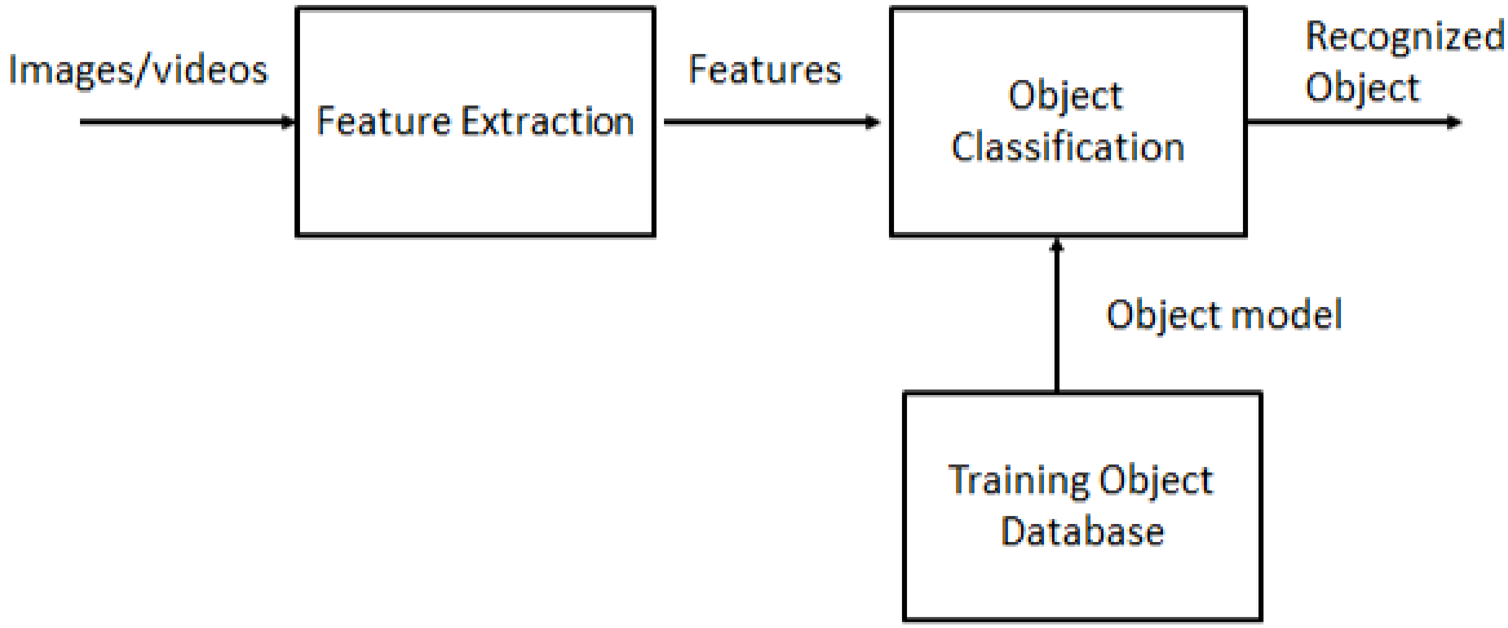
Object Recognition Methods

Object recognition can be classified into

- Learning based classification
- Model-based recognition

Learning-based Recognition

Object recognition is carried out through a multi-class classification



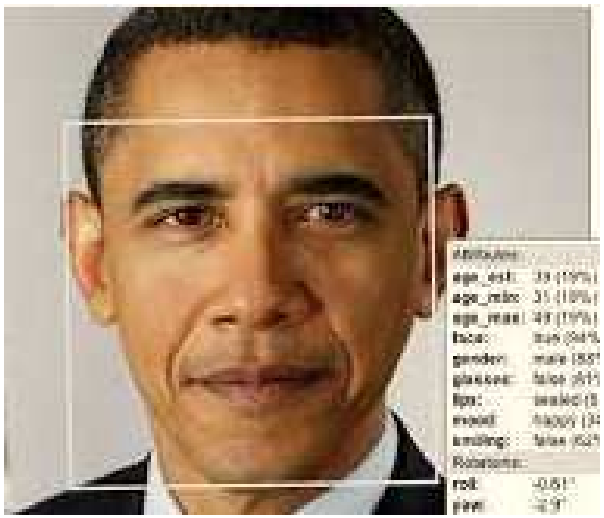
Hand-crafted Object Features

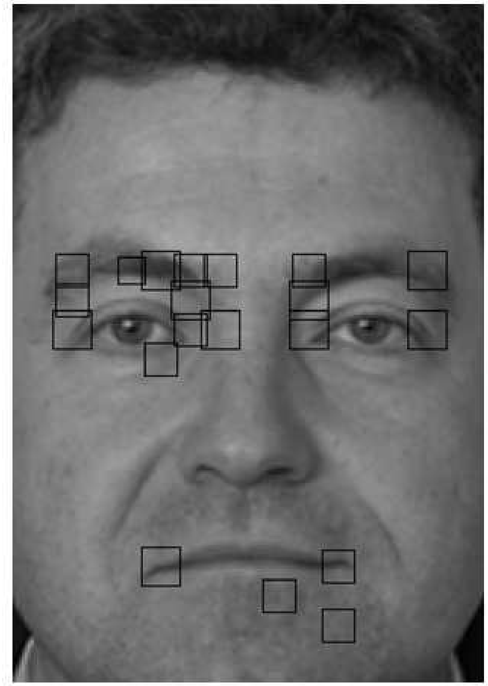
- Appearance Features
- Shape Features

Appearance Features

The appearance features characterize an object based on its photometric properties, i.e., intensity or color or their derivatives.

- Global features: extract features from all pixels in the image such as SIFT features
- Local features: extract features from pixels in some sub-regions of the images

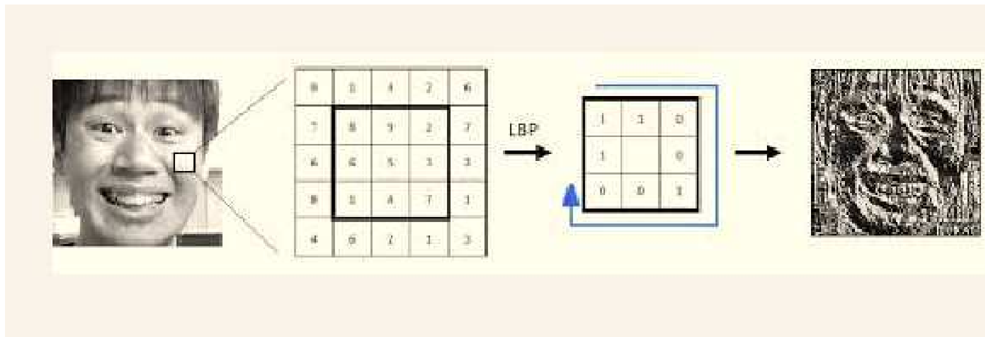


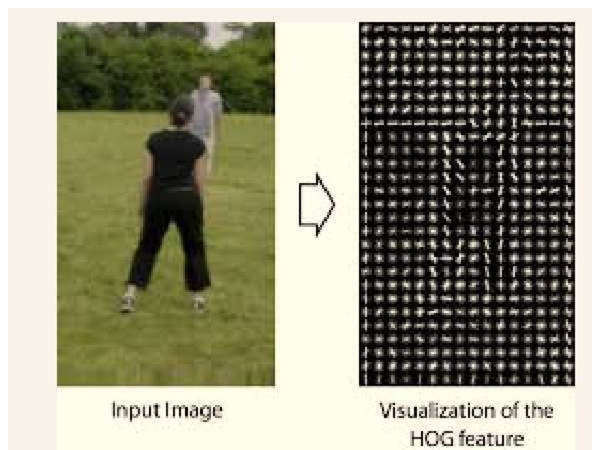


Global features are more complete but are computationally more expensive and less robust to shape and illumination variation.

Appearance Features: Spatial domain

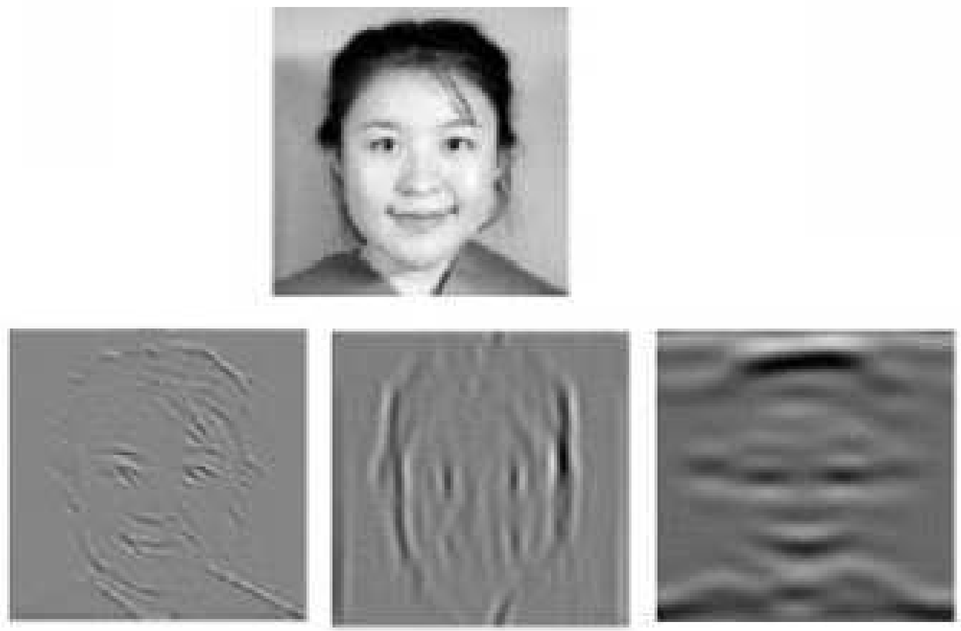
Spatial domain features: intensity, SIFT, Local Binary Patterns (LBP), HoG, LoG, DoG, etc..





Appearance Features: Transform Domain

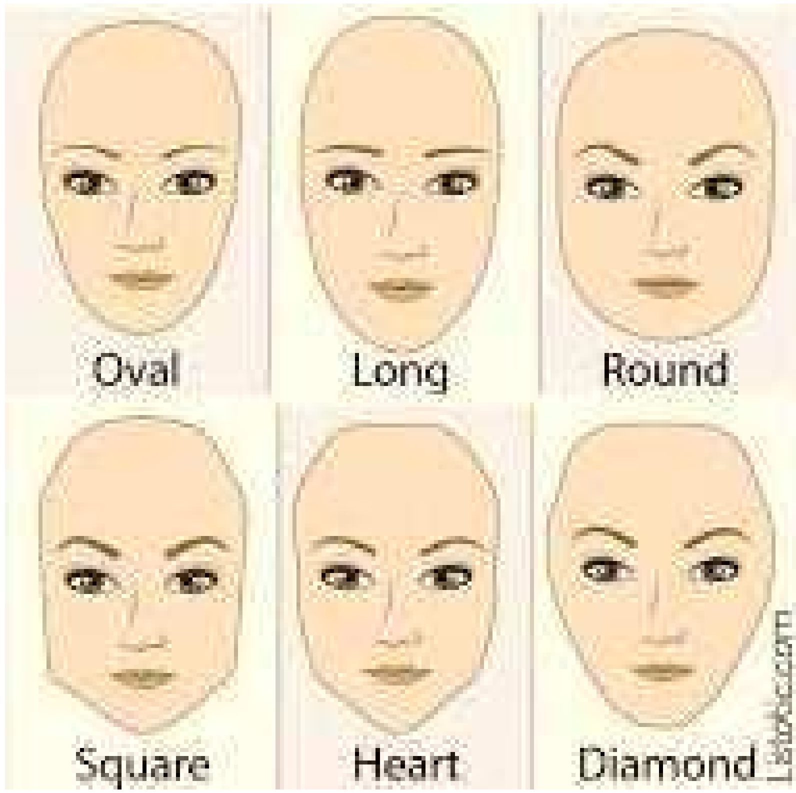
Transform-domain features: PCA, Gabor Wavlet, Harr Wavelet, Fourier transform



Geometric Features

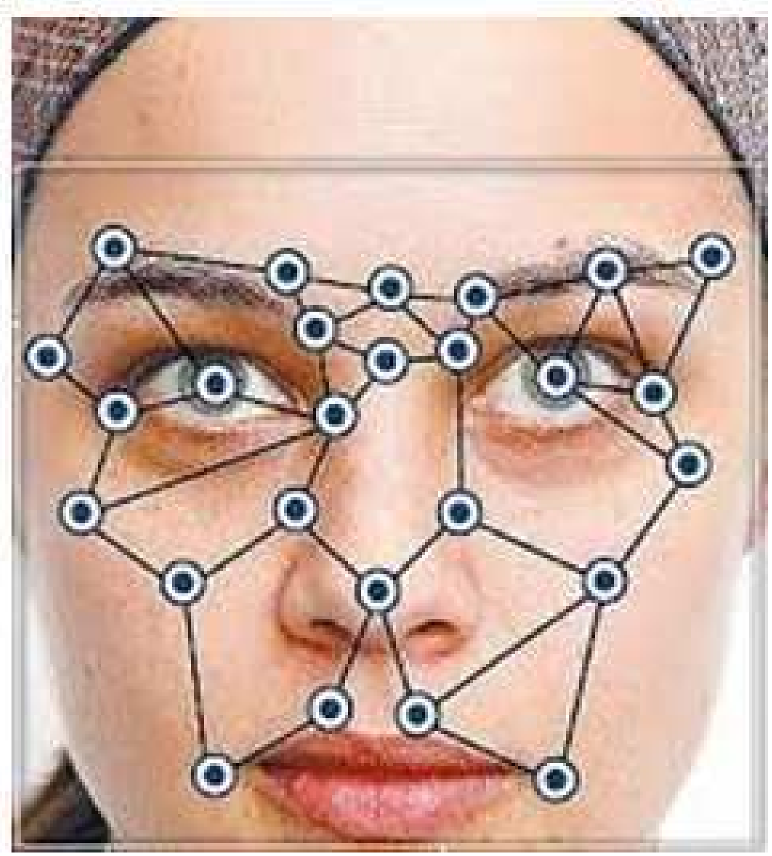
Geometric Features characterize an object based on its geometric properties including

- size and area
- curvature
- moments (centroid, central moments, Hu's moments)


Oval
Long
Round
Square
Heart
Diamond
Listotic.com

Geometric Features: Relation

Spatial relations between geometric entities including distances, orientations

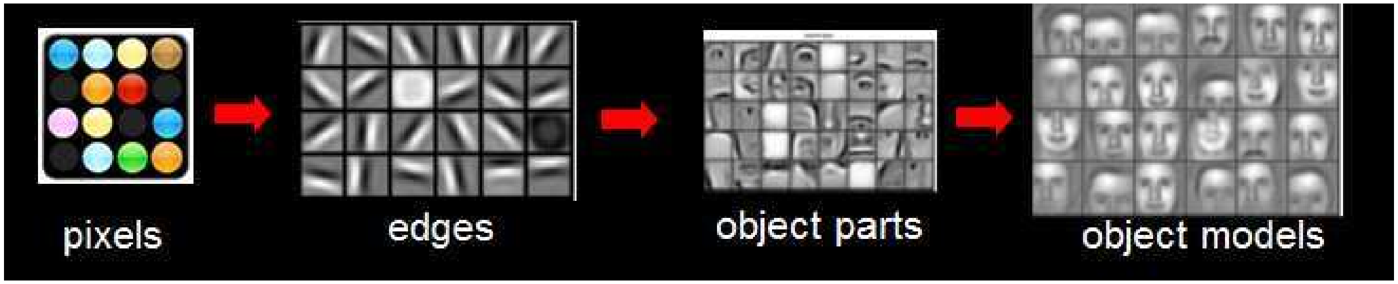


Feature Invariance

- Feature invariance to affine transformation, projective projection, view, and illumination change.
- SIFT feature is an important type of invariant appearance features. Cross-ratio is a projective invariant geometric feature.

Deep Learning Feature Learning

Instead of extracting manually defined features from image, recent research focuses on learning features automatically from the data using the deep learning models such as CNNs and DNNS to learn photometric features, GNNs to learn geometric features, and VAEs to learn feature in an unsupervised manner.



The learnt features have proven to be significantly better than the hand-crafted features.

Classifiers

Given features, classifiers can be trained to perform object recognition

- Parametric Classifiers

  Support Vector Machine (SVM)

  Neural Networks

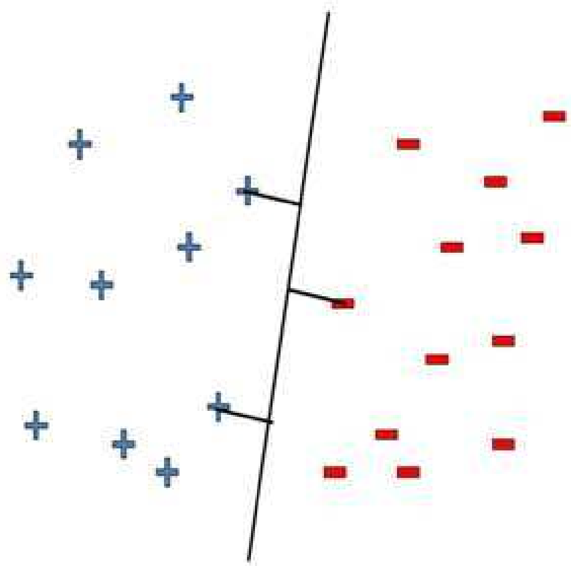
  Bayes Classifier

  Logistic regression

- Non-parameter classifiers

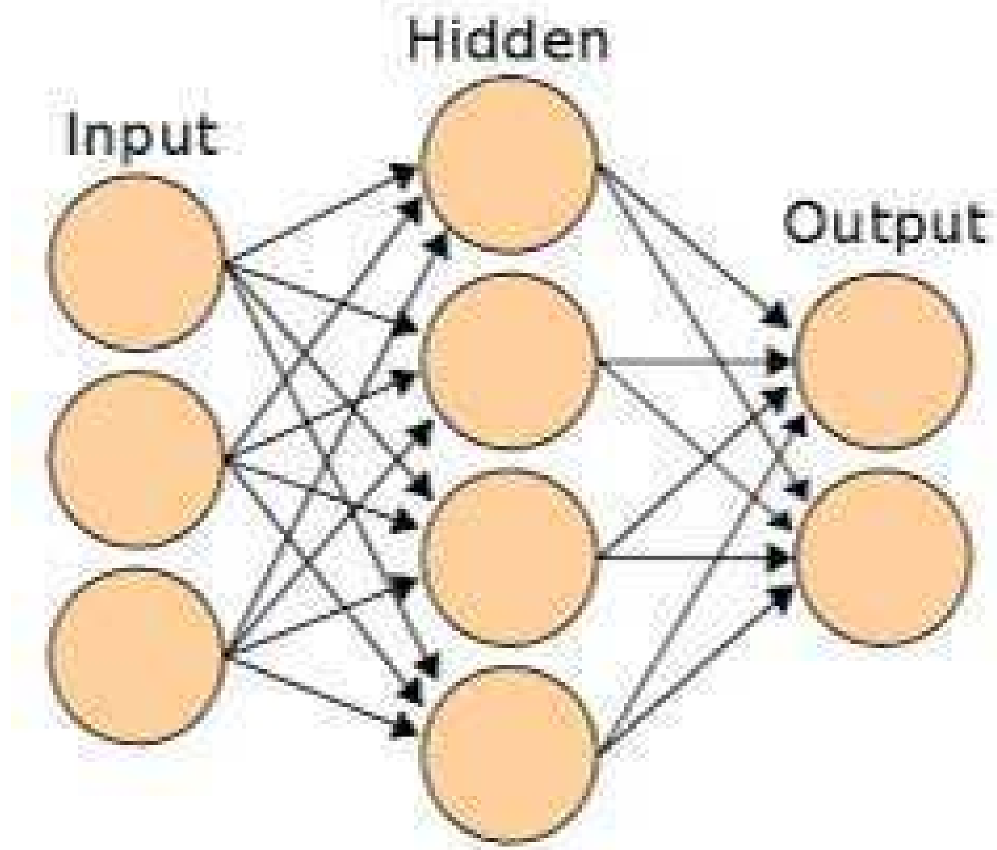
  k-nearest neighbor classifier (KNN)

Classifiers: SVM

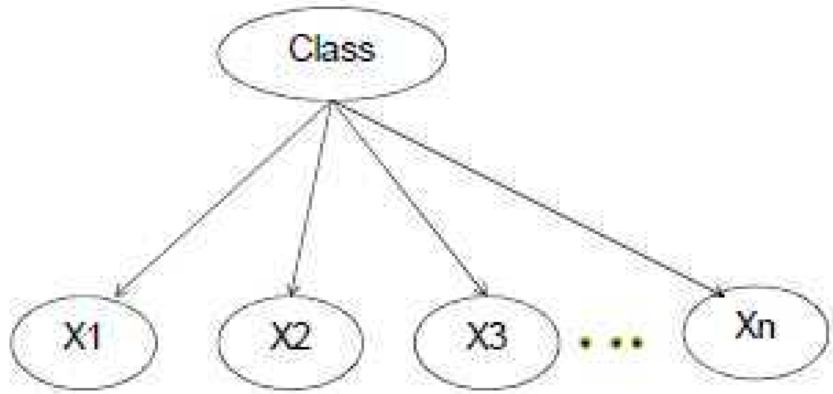


- Pick the decision boundary with the largest margin!
- Linear hyperplane defined by "support" vectors
- Moving other points does not affect the decision boundary
- Only need to store the support vectors to predict labels of new points

Classifiers: Neural Networks



Classifiers: Navie Bayesian Classifier



X1, X2, X3, and Xn are conditionally independent of each other given Class

$$P(Class, X1, X2, X3..., Xn) = P(Class)\prod_i P(X_i \mid Class)$$

54

Classifiers: Logic Regression

$$p(y|x) = \frac{1}{1 + \exp^{-(a_0 + a_1 x_1, a_2 x_2, \ldots, a_n x_n)}}$$

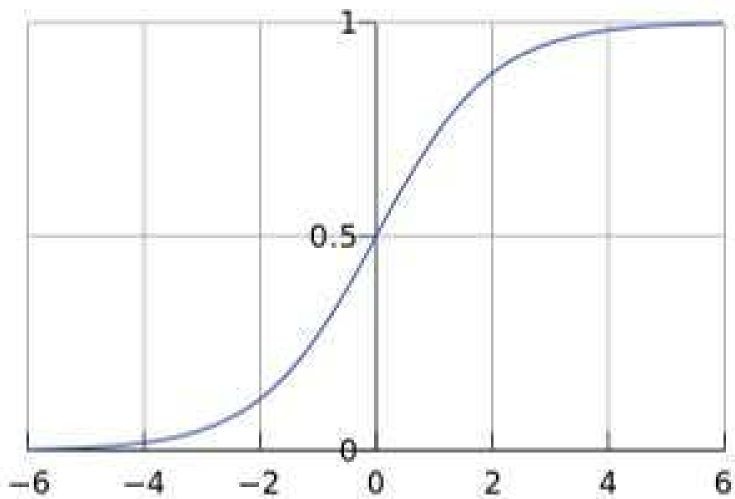
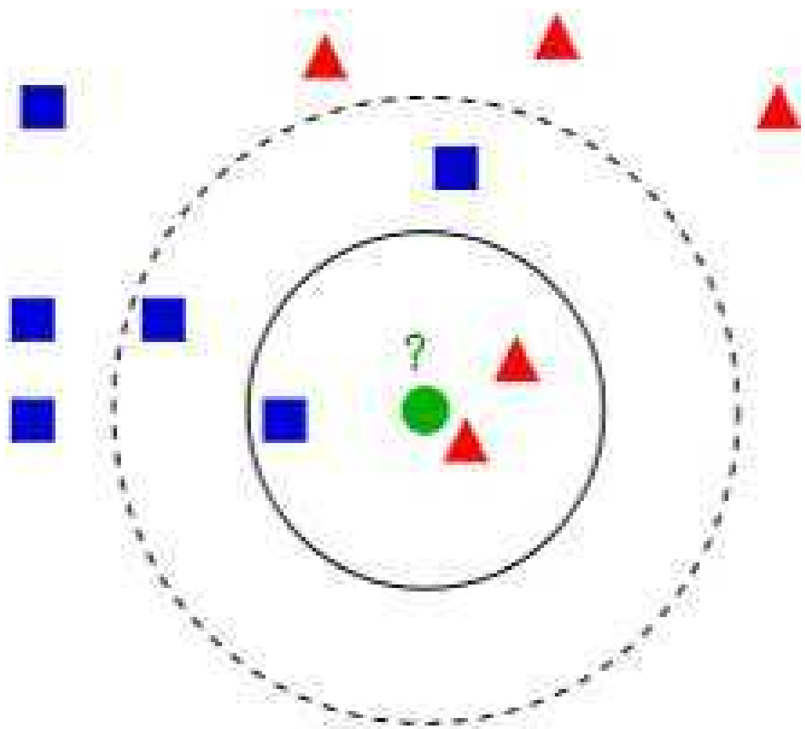


Figure 1: Logistic or sigmoid function

Classifiers: KNN



Feature Matching based Recognition

Object recognition is accomplished by measuring the distance between the target features and the image features. One common of such measure is the cosine distance between two feature vectors.

Model-based Object Recognition

- Mathematical model based recognition
- Geometric model based recognition

Mathematical Model based Object Recognition

- Construct a mathematical (probabilistic) (e.g. HMM) model for each class and given the testing data, evaluate the likelihood of each model. The testing data is classified into the class, whose corresponding model has the highest likelihood
- Model can better capture an object's latent states and their relationships

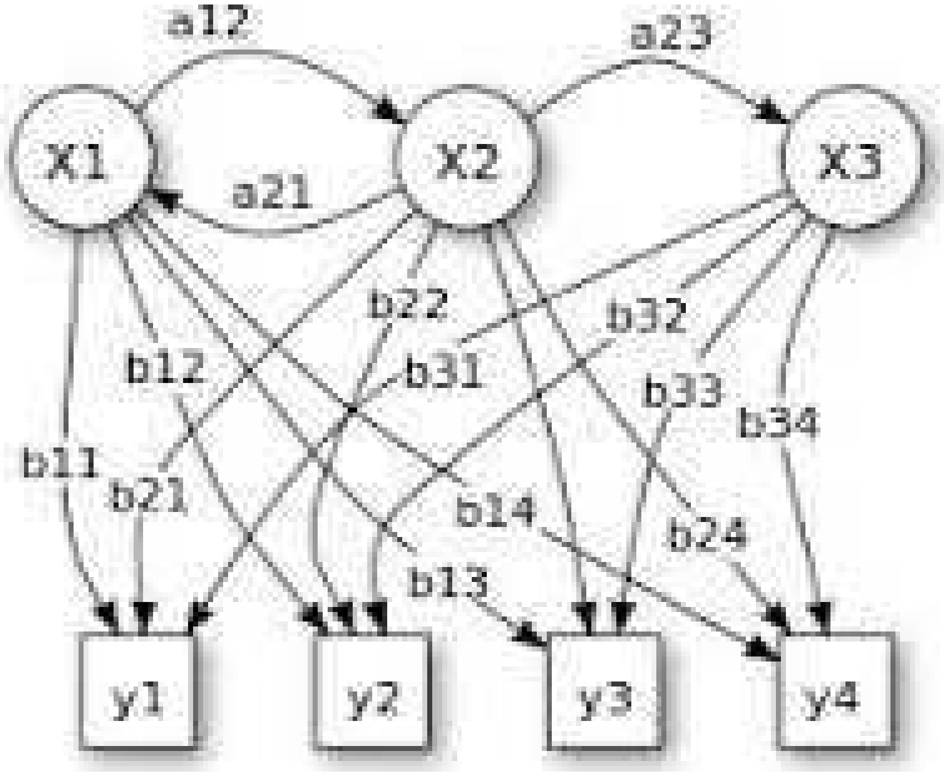


Figure 2: Hidden Markov Model

Geometric Model based Object Recognition

Given a database of 3D geometric models (e.g CAD models) and an input image, model-based object recognition identifies the model in the database that matches the image as well as determine the 3D pose of the object that produces the observed object image.

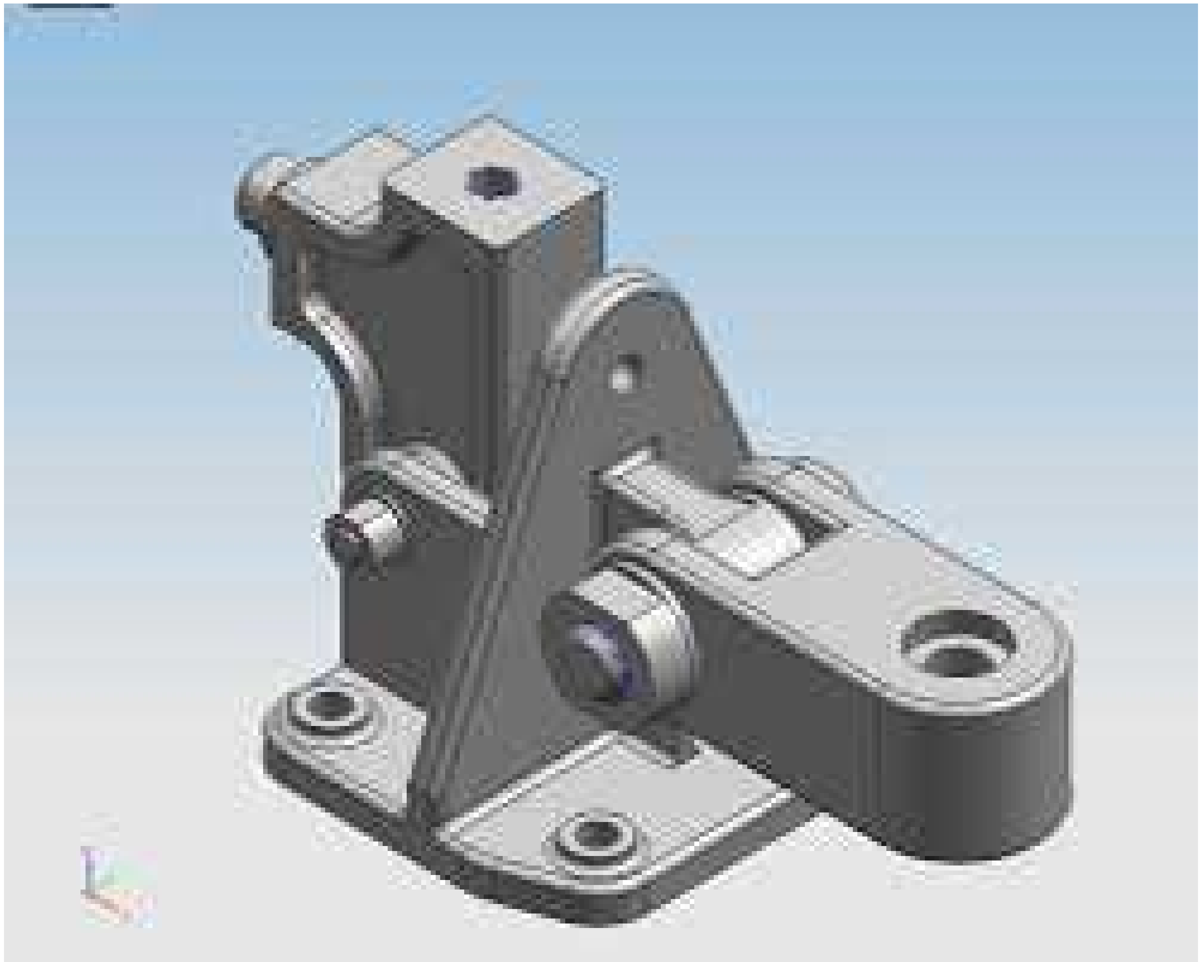


Figure 3: CAD Model

Object Detection and Recognition Challenges

- view point variation
- illumination variation
- scale variation
- occlusion
- background clutters
- shape change due to non-rigid deformation
- poor image quality-low resolution and low SNR
- Open set detection and recognition

References