COMP 9517 Computer Vision

Pattern Recognition (1)
Introduction

• **Pattern recognition** is the scientific discipline whose goal is the classification of objects into a number of categories or classes

• Pattern recognition used widely for object classification and recognition
  – To recognise a face
  – To read handwritten characters
  – To identify our car keys in our pocket by feel
  – To understand spoken words

• Objects can be images or any type of measurements that need to be classified, which are referred using the generic term **pattern**
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
Applications

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

• Other areas beside CV
  – Computer-aided diagnosis
  – Recommender systems
Pattern Recognition Systems

- Prototype of pattern recognition

- The basic stages involved in the design of a classification system

Diagram:

1. Preprocessing
   - Feature extraction
   - Classification
   - "salmon" 
   - "sea bass"

Flowchart:

- Patterns
  - Sensor
  - Feature Extraction
  - Feature Selection
  - Classifier Design
  - System Evaluation
Pattern Recognition Concepts

- **Object** - an object is a physical unit
- **Regions** - that correspond to objects are obtained, after segmentation of an image
- **Classes** - the set of objects can be divided into disjoint subsets that may have some common features- such sets are called classes
- **Object recognition/pattern recognition** - object recognition assigns classes to objects
- **Classifier** - the corresponding algorithm/method is called the classifier
- **Pattern** - the classifier bases its decision on object features, called the pattern
More Concepts

- **Features** - description of the objects
- **Model** - description of the classes
- **Pre-processing** - noise removal, segmentation
- **Feature Extraction** - reduce the data by measuring certain “features” or properties
- **Training samples** - experience, objects with known ground truth
- **Cost** - consequence of making incorrect decision
- **Decision boundary** - boundary between regions in feature space
Features and Descriptions

- **Features**
  - descriptions representing scalar properties of objects are called *features*
  - used to represent knowledge as part of more complex representation structure

- **Feature vector**
  - combines many features, e.g. size feature represents area property, compactness feature represents circularity

- Good representation is important to solve a problem
- Rich structured representation can simplify control strategies
Feature Vector Representation

- $X = [x_1, x_2, \ldots, x_n]$, each $x_j$ is a real number
  - $x_j$ may be an object measurement
  - $x_j$ may be count of object parts

- Example:
  - $[\#\text{holes}, \#\text{strokes}, \text{moments}, \ldots]$  
  - $[\text{length}, \text{colour}, \text{lightness}, \ldots]$
Feature Extraction

• Goal of feature extraction is to characterise object by measurements that are
  – similar for objects in the same class/category, and
  – different for objects in different classes
• Must find *distinguishing features* that are invariant to input transformations
• Design of features often based on prior experience or intuition
Feature Extraction

- Selecting features that are
  - translation, rotation and scale invariant in images
  - handling *occlusion*, projective distortion for 3-D objects in images
  - invariant to translations in time and changes in amplitude
  - handling *non-rigid deformations* common in 3-D vision
- Feature selection is problem- and domain-dependent
- But classification techniques can help to
  - make feature values less noise sensitive, and
  - to select valuable features out of a larger set
Classification

• Classifier performs object recognition by assigning an object to a class
  – using the object description in the form of features

• Perfect classification is often impossible
  – we determine probability for each possible category

• Variability in feature values for objects in the same class versus those in different classes causes the difficulty of the classification problem
  – Variability in feature values may arise due to complexity, but also due to *noise*
  – Noisy features and missing features are major issues
Bayesian Decision Theory

• A classifier's decision may or may not be correct, so setting should be probabilistic
• Probability distributions may be used to make classification decisions with least expected error rate
Bayesian Decision Theory

• **Bayesian classifier** classifies an object into the class to which it is most likely to belong, based on observed features

• Assume:
  – *a priori* probability $P(\omega_i)$ for each class $\omega_i$
  – unconditional distribution $P(x)$
  – class conditional distribution $P(x | \omega_i)$

• If all the classes are disjoint, by Bayes Rule, the *a posteriori* probabilities are given by:

\[
P(\omega_i | x) = \frac{P(x | \omega_i)P(\omega_i)}{\sum_j P(x | \omega_j)P(\omega_j)}
\]
Bayesian Decision Theory

• If we have an observation $x$ for which $P(\omega_1 | x)$ is greater than $P(\omega_2 | x)$, we would naturally prefer to decide that the true state of nature is $\omega_1$

• Whenever we observe a particular $x$, the probability of error is

$$P(\text{error} | x) = \begin{cases} P(\omega_1 | x), & \text{if we decide } \omega_2 \\ P(\omega_2 | x), & \text{if we decide } \omega_1 \end{cases}$$

• Clearly, for a given $x$ we can minimise the probability of error by deciding $\omega_1$ if $P(\omega_1 | x) > P(\omega_2 | x)$

• The **Bayes decision rule**

Decide $\omega_1$ if $P(\omega_1 | x) > P(\omega_2 | x)$; otherwise decide $\omega_2$. 
Parametric Models for Distributions

- To compute $P(x|\omega_i)$ and $P(\omega_i)$, we can use an empirical method based on given samples.
- Or if we know that the distribution of $x$ follows a parametric model, then we may estimate the parameters using the samples.
- **An Example**
  - Assume that the patterns in the $r_{th}$ class can be described by a normal distribution, whose dispersion matrix $\Sigma_r$ is known but the mean $\mu_r$ is unknown.
  - Then, an estimate of the mean may be the average of the labelled samples available in the training set:
    \[
    \mu = \bar{x}
    \]
References and Acknowledgements

• Shapiro and Stockman, Chapter 4
• Duda, Hart and Stork, Chapter 1
• More references
• Some content are extracted from the above resources

22/08/16