RAF103F Mynsturgreining

Pattern Recognition

Instructor: Yuliya Tarabalka, Tel. 844 73 06,

E-mail: yuliya.tarabalka@hyperinet.eu

Supervisory teacher: Jón Atli Benediktsson

Course Outline:

1. Introduction
2. Mathematical Preliminaries } 1 week
3. Bayesian Decision Theory 1 week
4. Maximum Likelihood Estimation 1 week
5. Non-Parametric Classification 1 week
6. Linear Discriminant Functions 1 week
7. Feature Extraction for Representation and Classification
8. Unsupervised Analysis } 1 week

Course Outline:

1. Introduction
2. Mathematical Preliminaries — 1 week
3. Bayesian Decision Theory — 1 week
4. Maximum Likelihood Estimation — 1 week

Mid-term exam

5. Non-Parametric Classification — 1 week
6. Linear Discriminant Functions — 1 week
7. Feature Extraction for Representation and Classification
8. Unsupervised Analysis — 1 week

Final exam

Generalities

Hours: Tuesday 13:20-14:50 (2h),
Thursday 13:20-15:40 (3h),

? (2h)

Location: O-104

Requirements:

- **Home works (10% of the final grade)**
- **Mid-term exam (10%)**
- **Final project (30%)**
- **Final exam (50%)**

Final Project

- Design, implementation and testing of a pattern recognition system
- Aim: to apply one or several pattern classification techniques to real world data

- Evaluation: report of 4 to 10 pages, including:
    - The problem you will try to solve
    - The method you are going to apply to solve it
    - Results and their discussion
    - Conclusions

Reading Material

1. R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., , John Wiley & Sons, 2000.
2. K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press 1990.
3. C. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.

1. Introduction

Introduction

- Examples of Patterns
- Examples of Applications
- Pattern Recognition Systems
- The Design Cycle
- Conclusion

What is a pattern?

- ***Pattern*: a description (sort) of an object**

- **"Each pattern is a three-part rule, which expresses a relation between a certain *context*, a *problem*, and a *solution*." (C. Alexander. "A Pattern Language". Oxford University Press, New York, 1977)**

Examples of Patterns

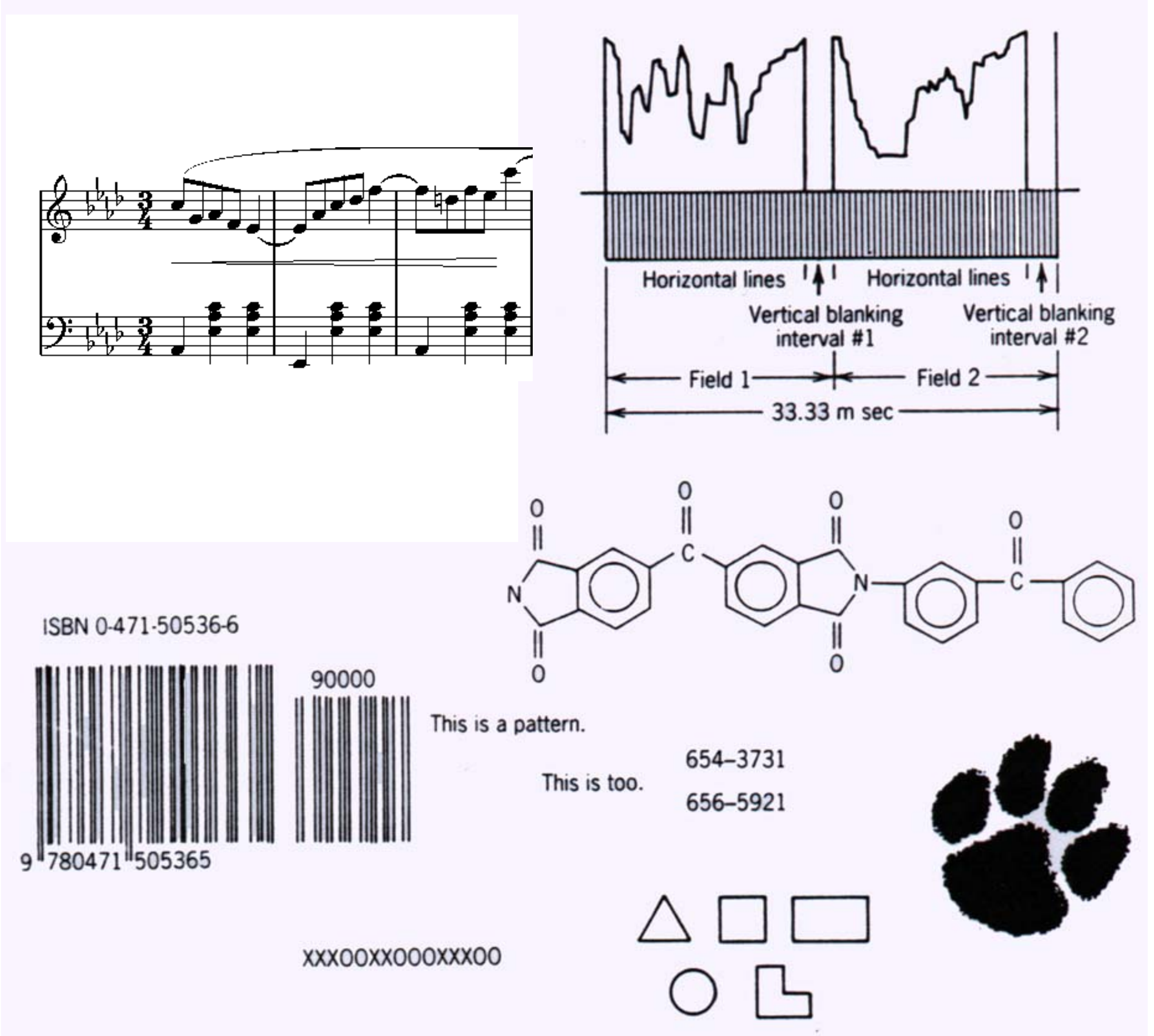


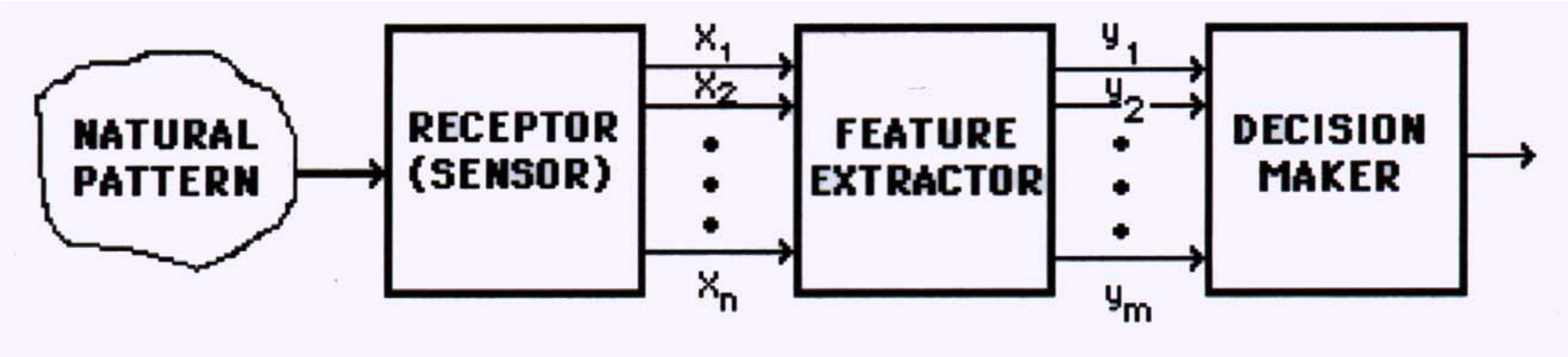
Figure 3: Examples of (visual) patterns/measurements.

Examples of Patterns

Figure 1.1. Variety of different images all representing the same character *A* (from Hofstadter's *Metamagical Themes: Questing for the Essence of Mind and Pattern* [HOF1985]).

Pattern Recognition

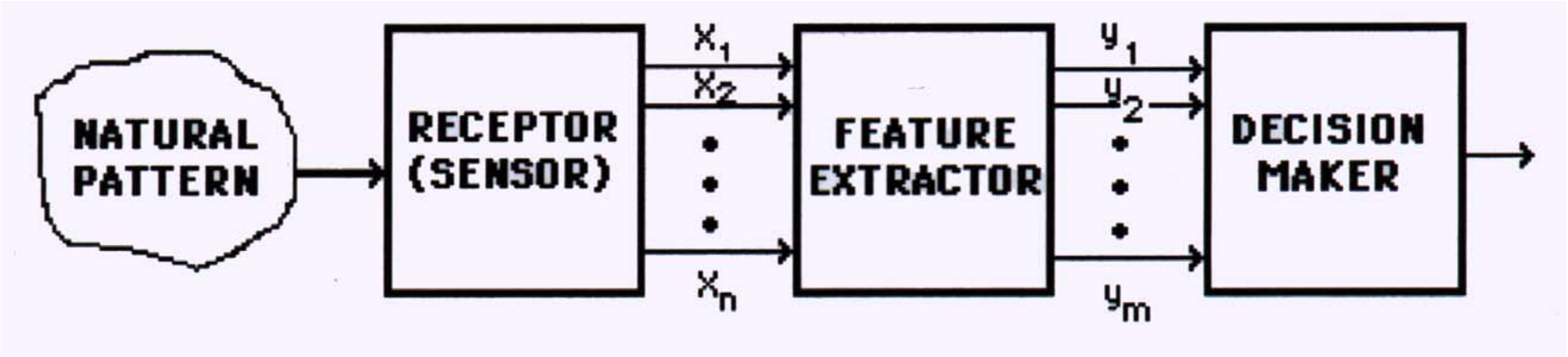
Pattern recognition **is the scientific discipline whose goal is the classification of** ***objects*** **(*****patterns*****) into a number of categories or** ***classes****



* *http://profsite.um.ac.ir/~patternrec/about.htm*

Pattern Recognition

***Pattern recognition* is the science of making *inferences* from *perceptual data*, using tools from statistics, probability, computational geometry, machine learning, signal processing, and algorithm design**



Pattern Recognition Approaches

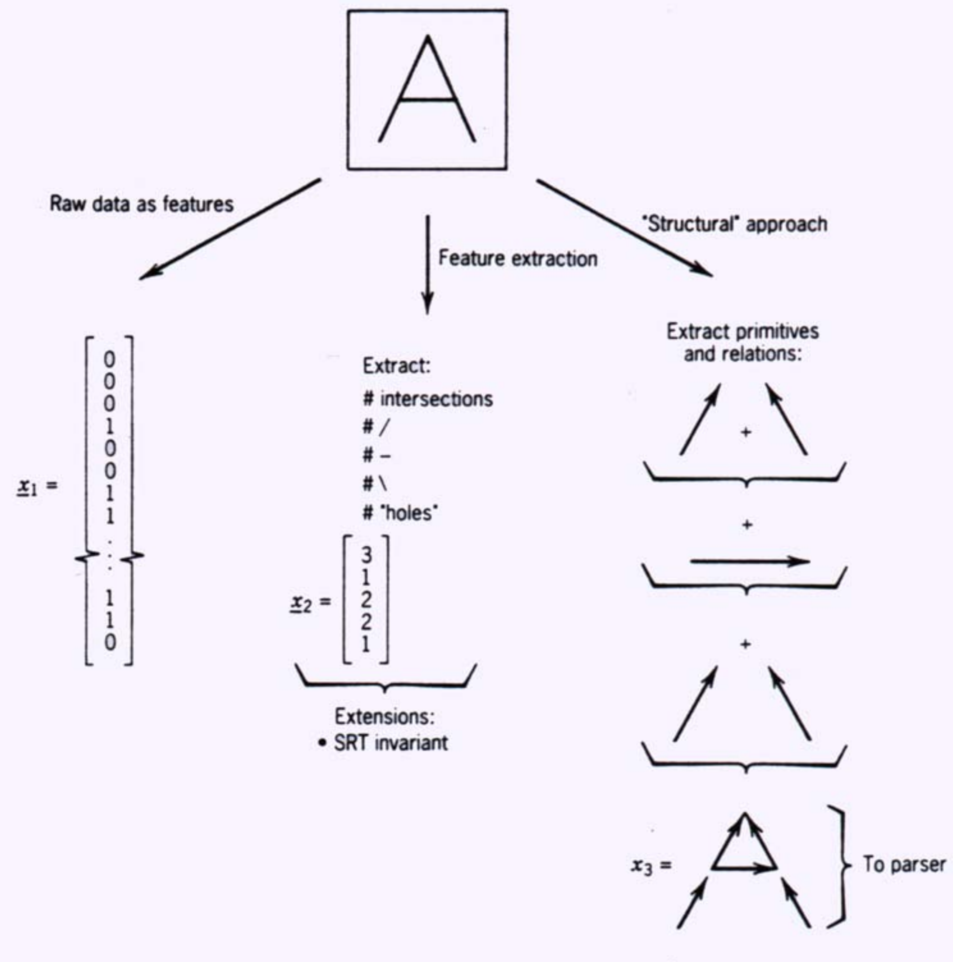


Figure 14: Example of PR approaches.
(a) Input pattern.
(b) Possible feature extraction approaches.

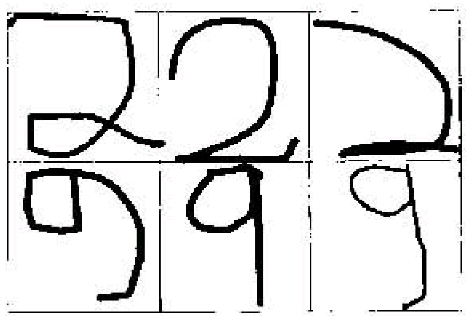
Pattern Recognition Applications

- Analysis of Remote Sensing Imagery
- Speech/speaker recognition
- Classification of seismic signals
- Medical waveform classification (EEG, ECG)
- Biometric ID: Identification of people from physical characteristics
- Medical image analyses (CAT scan, MRI)
- License plate recognition
- Decision making in stock trading

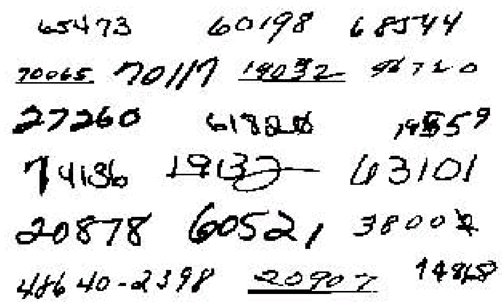
Optical Character Recognition (OCR)

Handwritten Digit Recognition

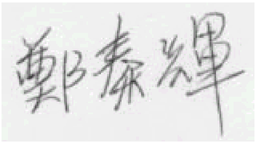
Segmented



Unsegmented



Handwritten Character Recognition



Chocolate Variety Pack

Play video

Face Detection & Recognition

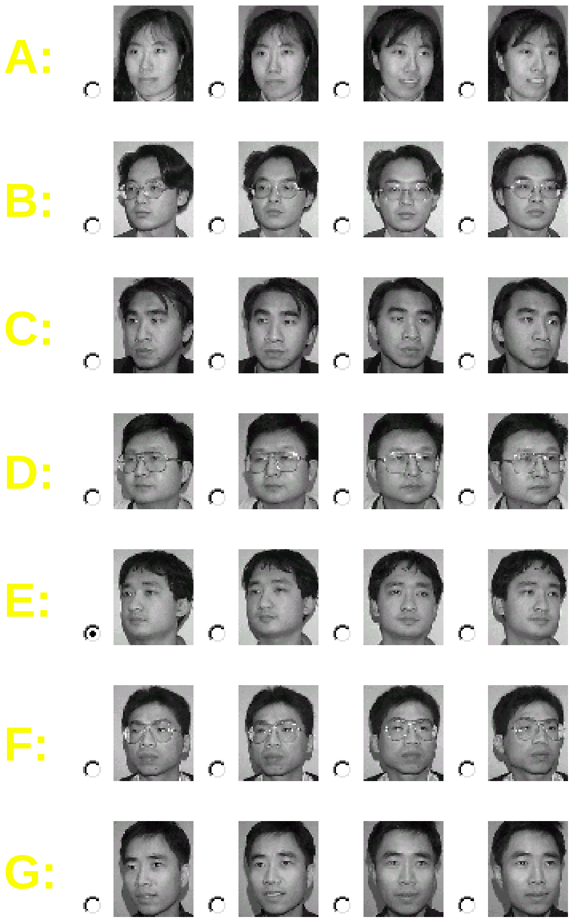
Test image:

Problems:

1. Is there a face?
2. If yes, how many?
3. How many are female?
4. Is Jones there? If yes, which one?

Face Recognition

Image databases:



Test image:

Who is this guy?

Tasks:

- **Face detection**
- **Local feature extraction**
- **Global feature extraction**
- **Data classification**

Face recognition homepage:

http://www.face-rec.org

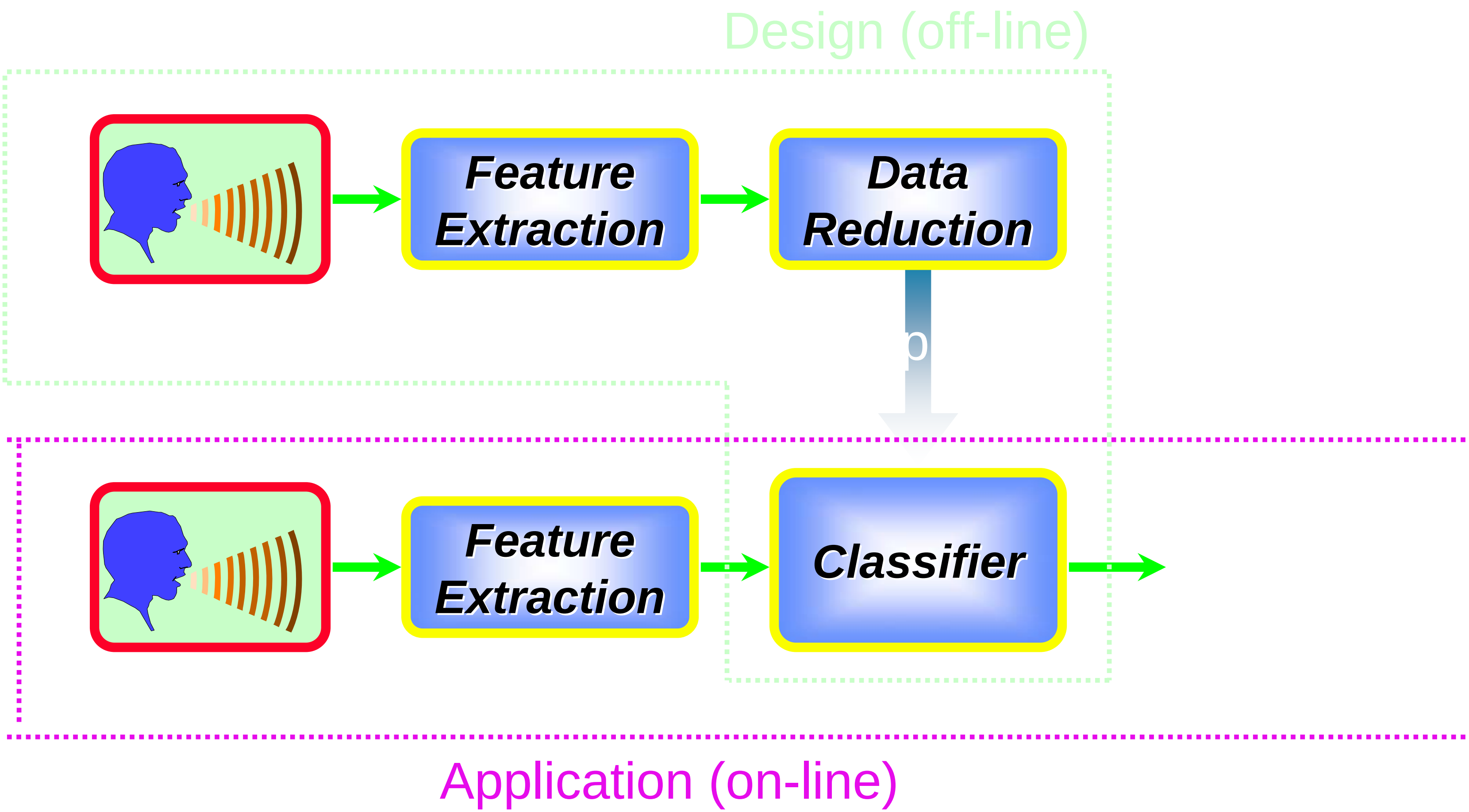
Biometric Identification

Identification of people from their physical characteristics, such as

- **faces**
- **voices**
- **fingerprints**
- **palm prints**
- **hand vein distributions**
- **hand shapes and sizes**
- **retinal scans**

Example: Speaker Recognition

Schematic diagram:



Example: Gender Classification

Goal:

Determine a person's gender from his/her profile data

Features collected

- Birthday
- Blood type
- Height and weight
- Hair length
- Voice pitch
- …
- Chromosome

Remote sensing image classification

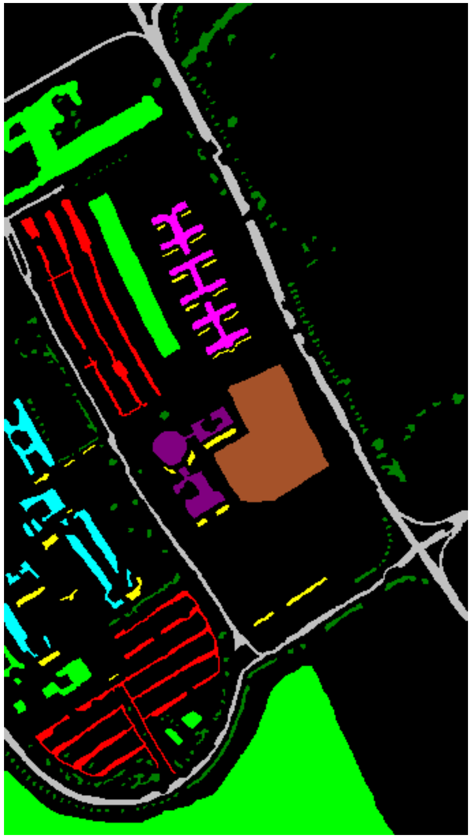
Input image (103 spectral channels)

Task:

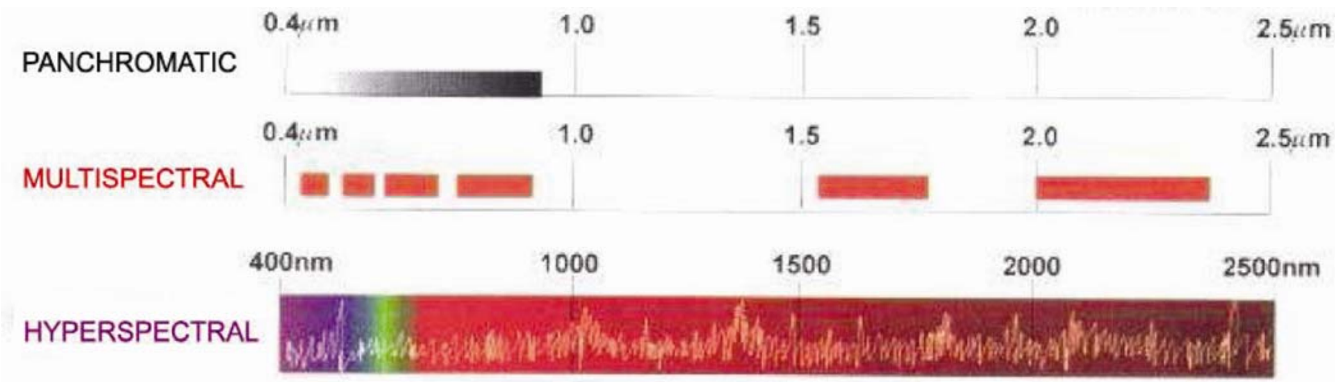
Assign *every* pixel to *one* of the *nine* classes:

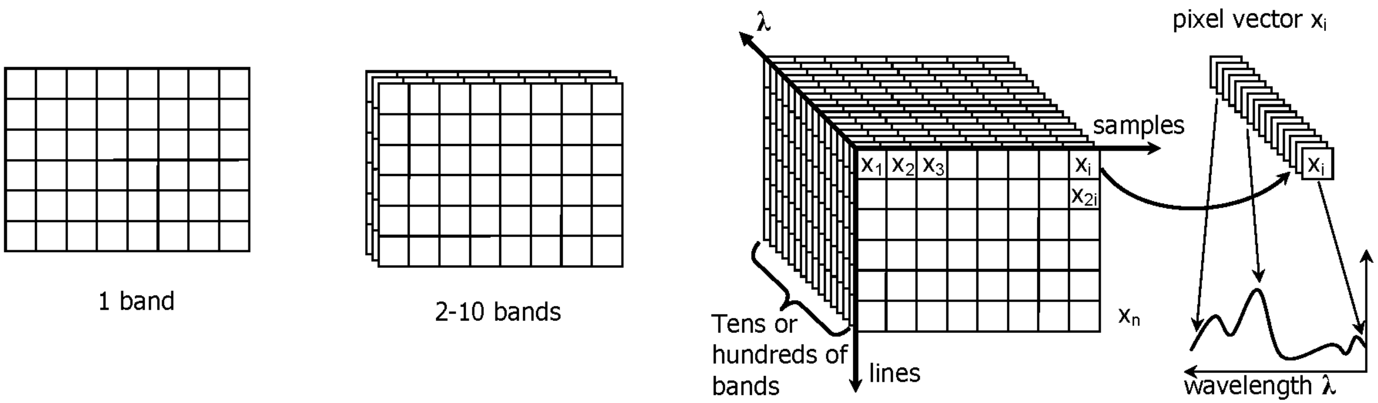
alphalt
meadows
gravel
trees
metal sheets
bare soil
bitumen
bricks
shadows

Reference data



Spectral context



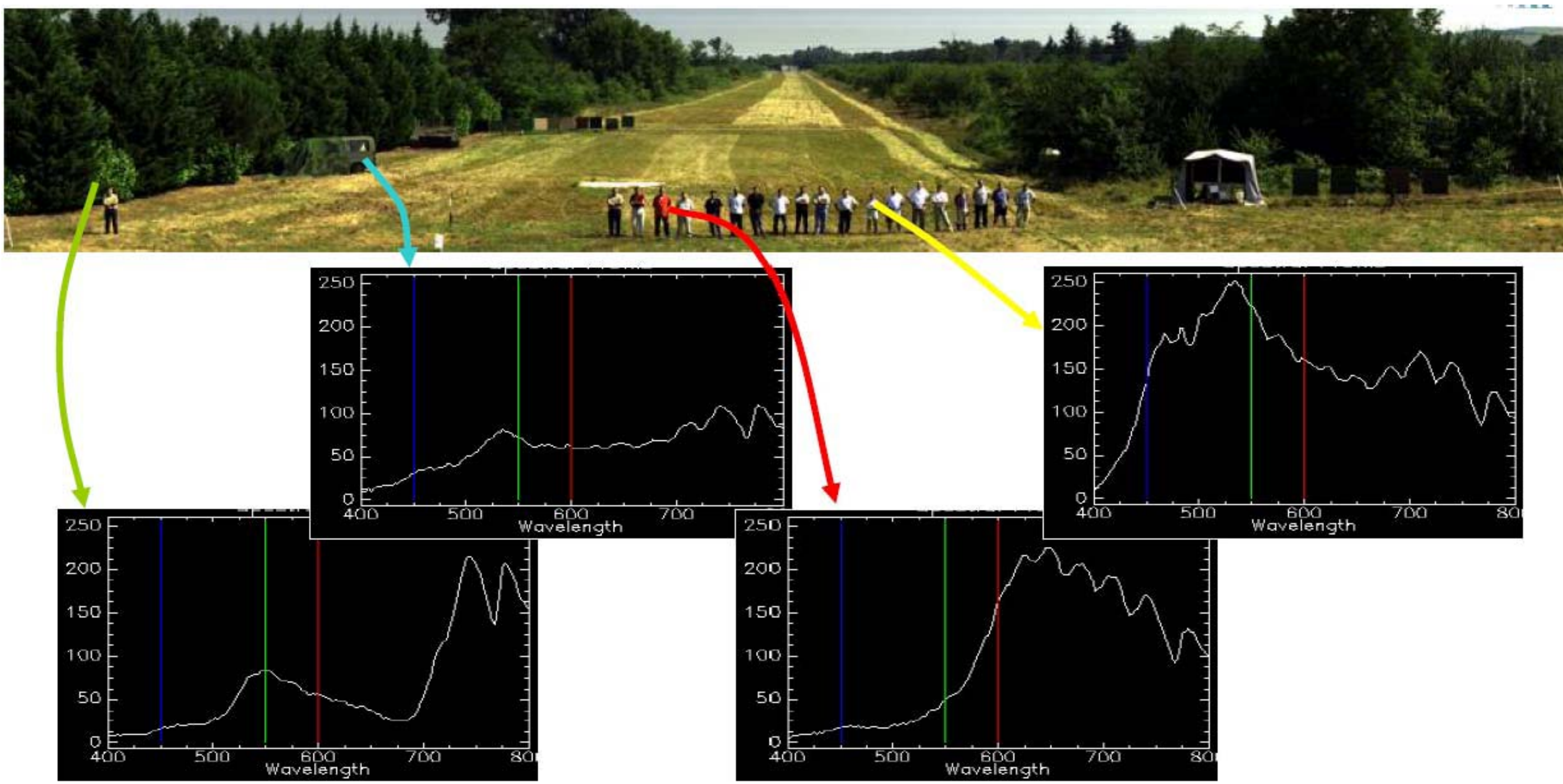


Panchromatic: one grey level value per pixel

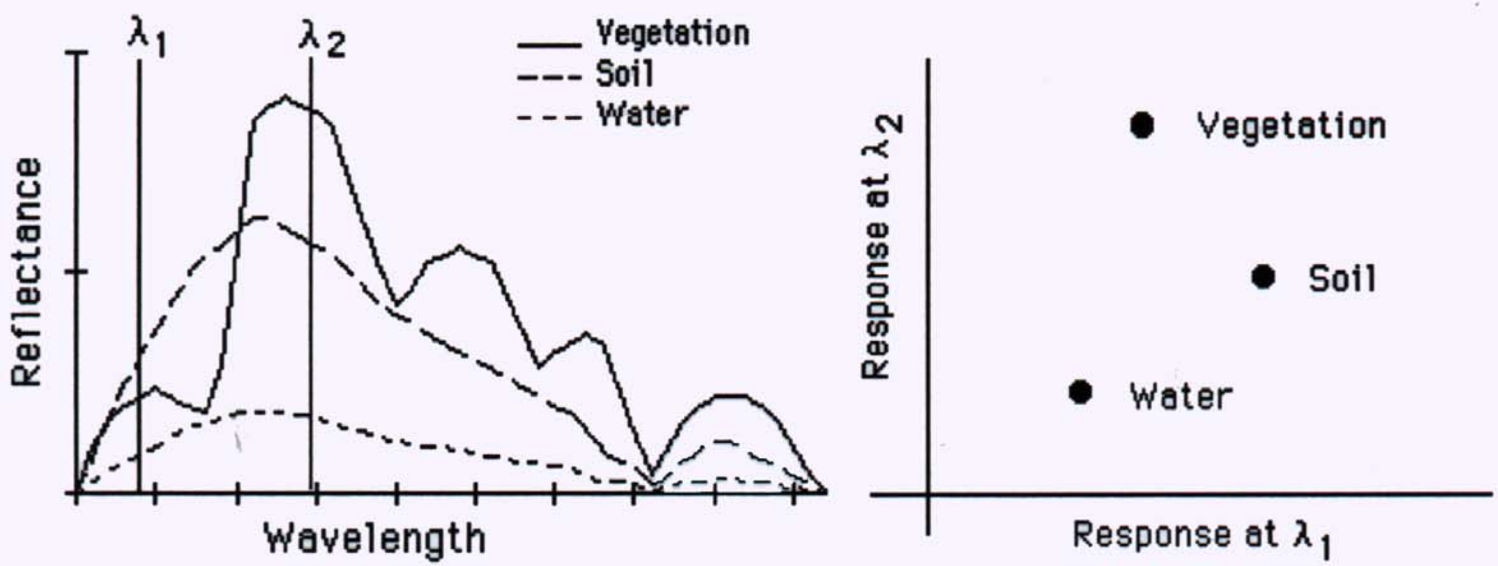
Multispectral: limited spectral info

Hyperspectral: detailed spectral info

Spectral context

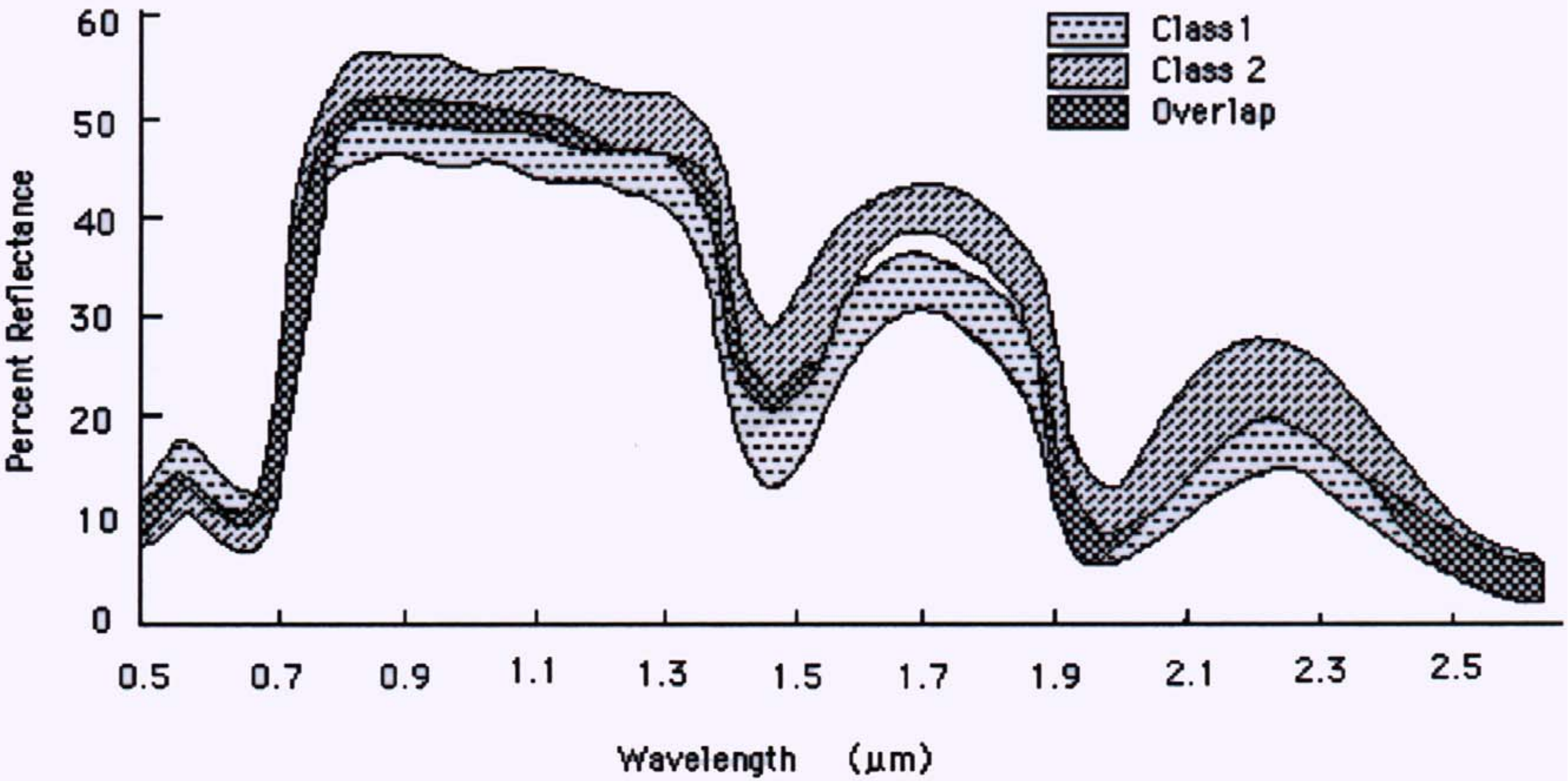


Feature Space



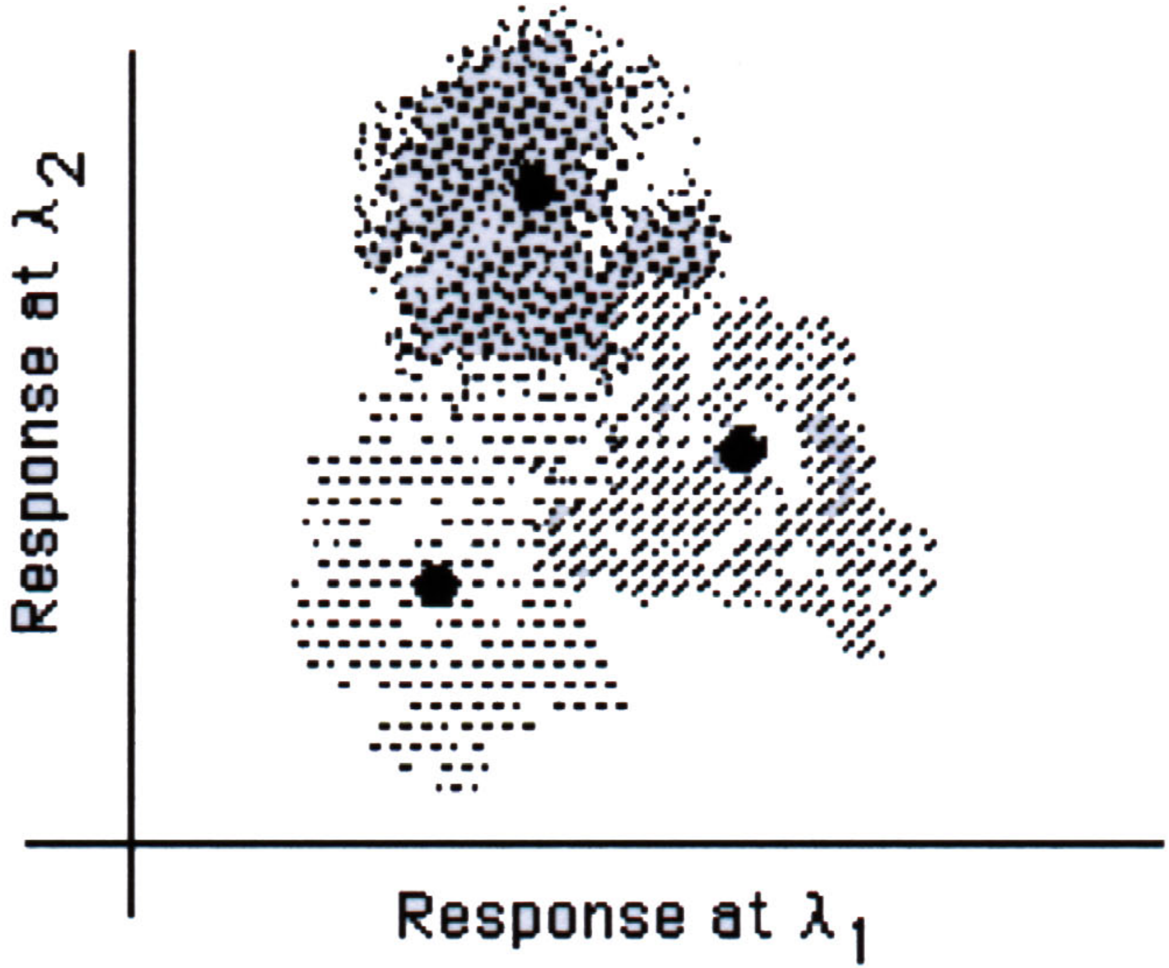
Three spectral responses mapped to a multivariant feature space. Though sampling at only two wavelengths is shown here, sampling at any number of wavelengths could be used, resulting in a correspondingly higher dimensional feature space.

There is no such thing as a clean pixel



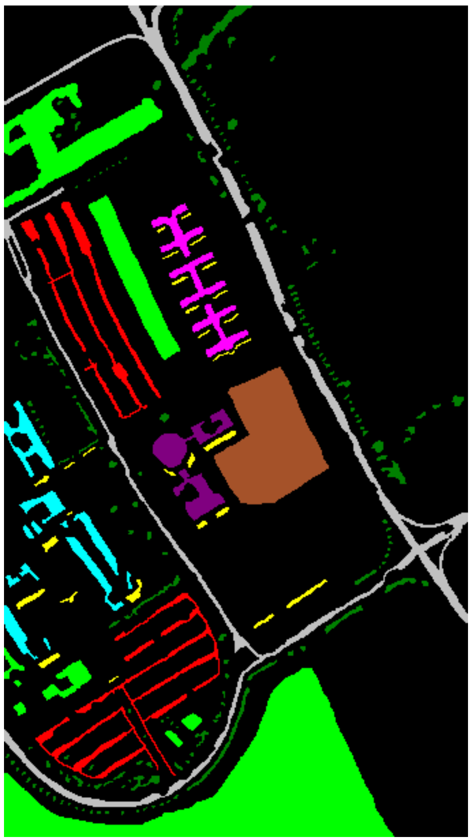
Spectral responses for a typical pair of classes, showing the interval at each wavelength into which they fall.

Feature Space



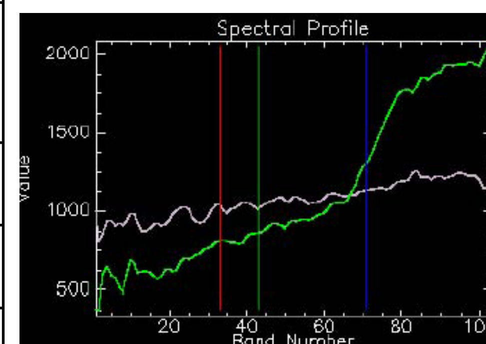
Data for classes tends to occupy a region in feature space rather than occurring at a single point.

Support Vector Machines classification



Class	Training samples	Test samples
Asphalt	**548**	**6304**
Meadows	**540**	**18146**
Gravel	**392**	**1815**
Trees	**524**	**2912**
Metal sheets	**265**	**1113**
Bare soil	**532**	**4572**
Bitumen	**375**	**981**
Bricks	**514**	**3364**
Shadows	**231**	**795**

Spectrum of each pixel is analyzed

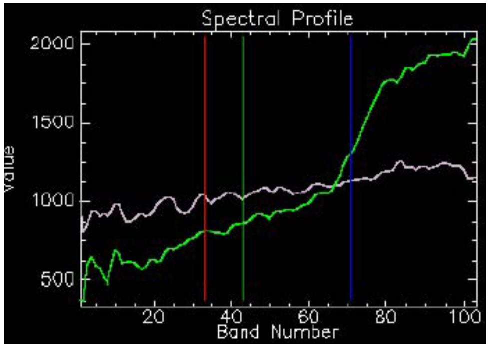


Support Vector Machines classification

alphalt, meadows, gravel, trees, metal sheets,
bare soil, bitumen, bricks, shadows

Overall accuracy = 81%

Spectral-spatial classification

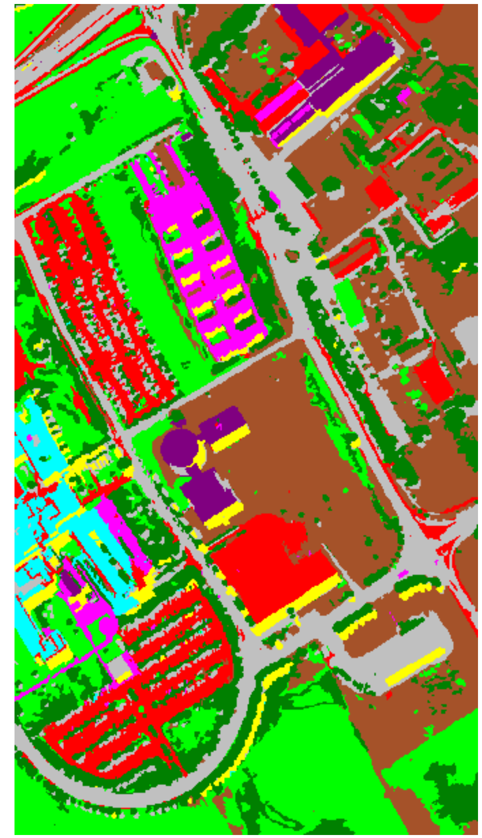


+

- Info about spatial structures included
  - For more accurate classification

- How to define spatial structures automatically?

Spectral-spatial classification

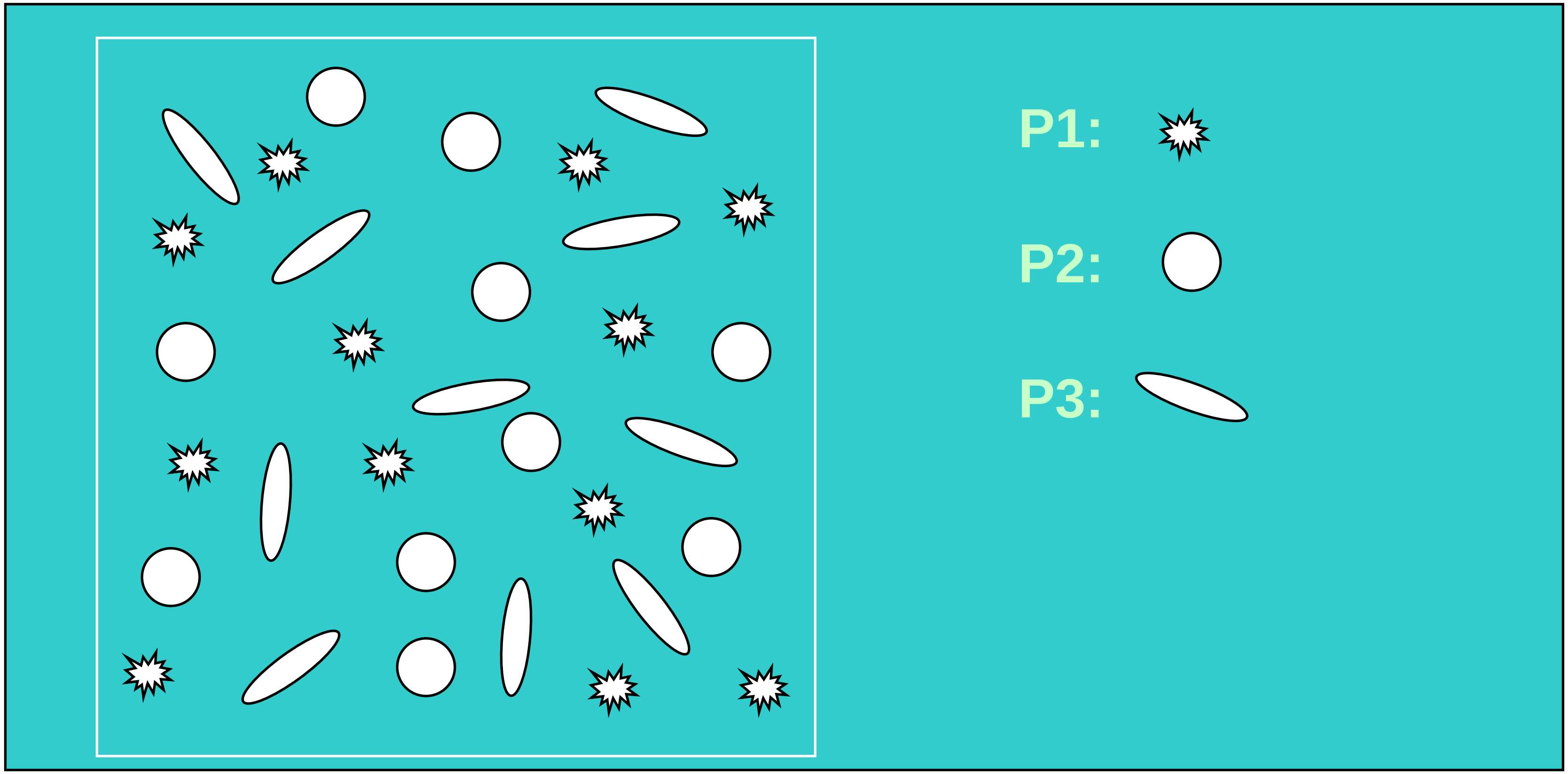


alphalt, meadows, gravel, trees, metal sheets, bare soil, bitumen, bricks, shadows

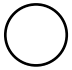
Overall accuracy = 94%

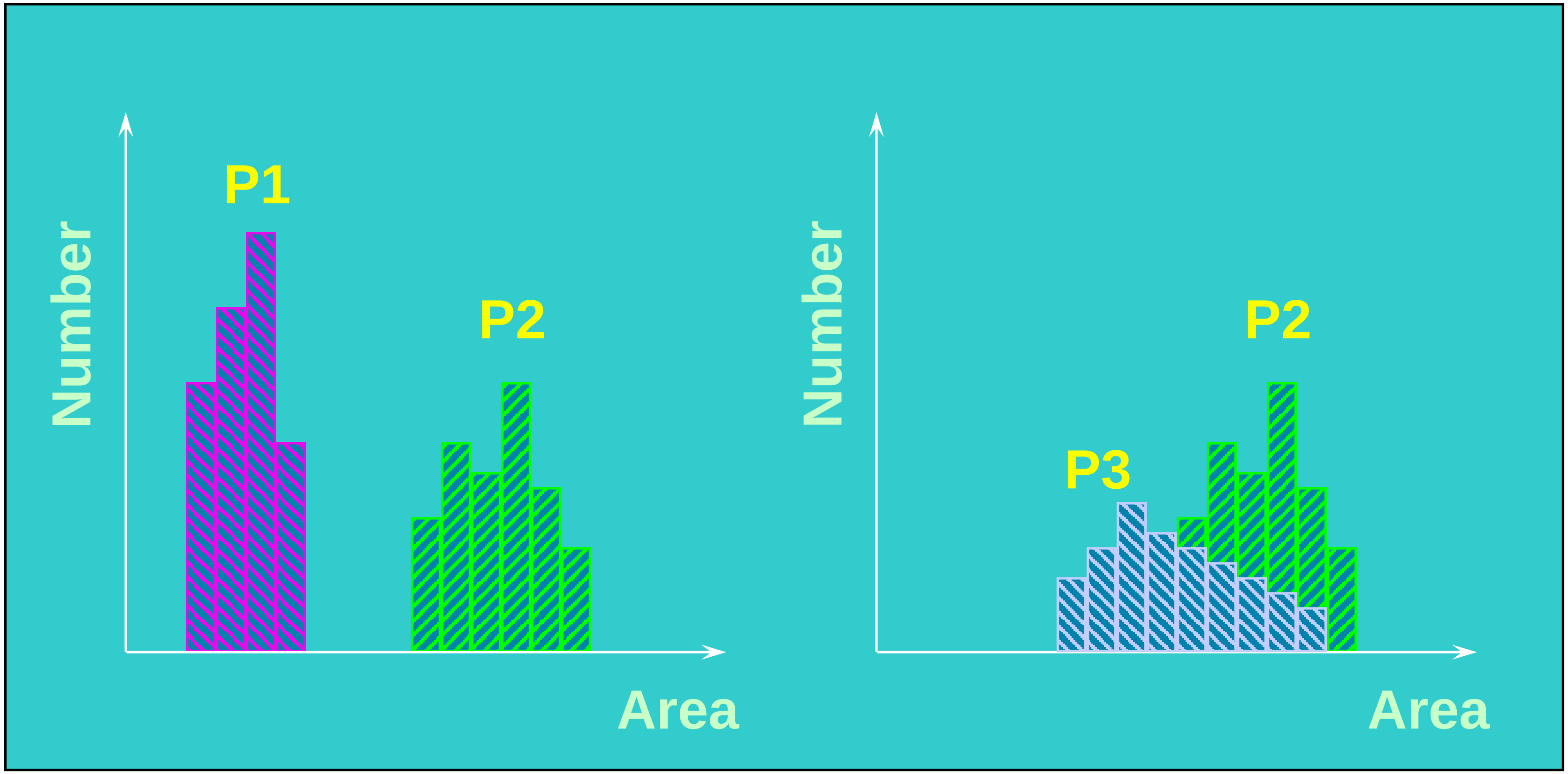
Example: Particle Classification

Particles on an air filter

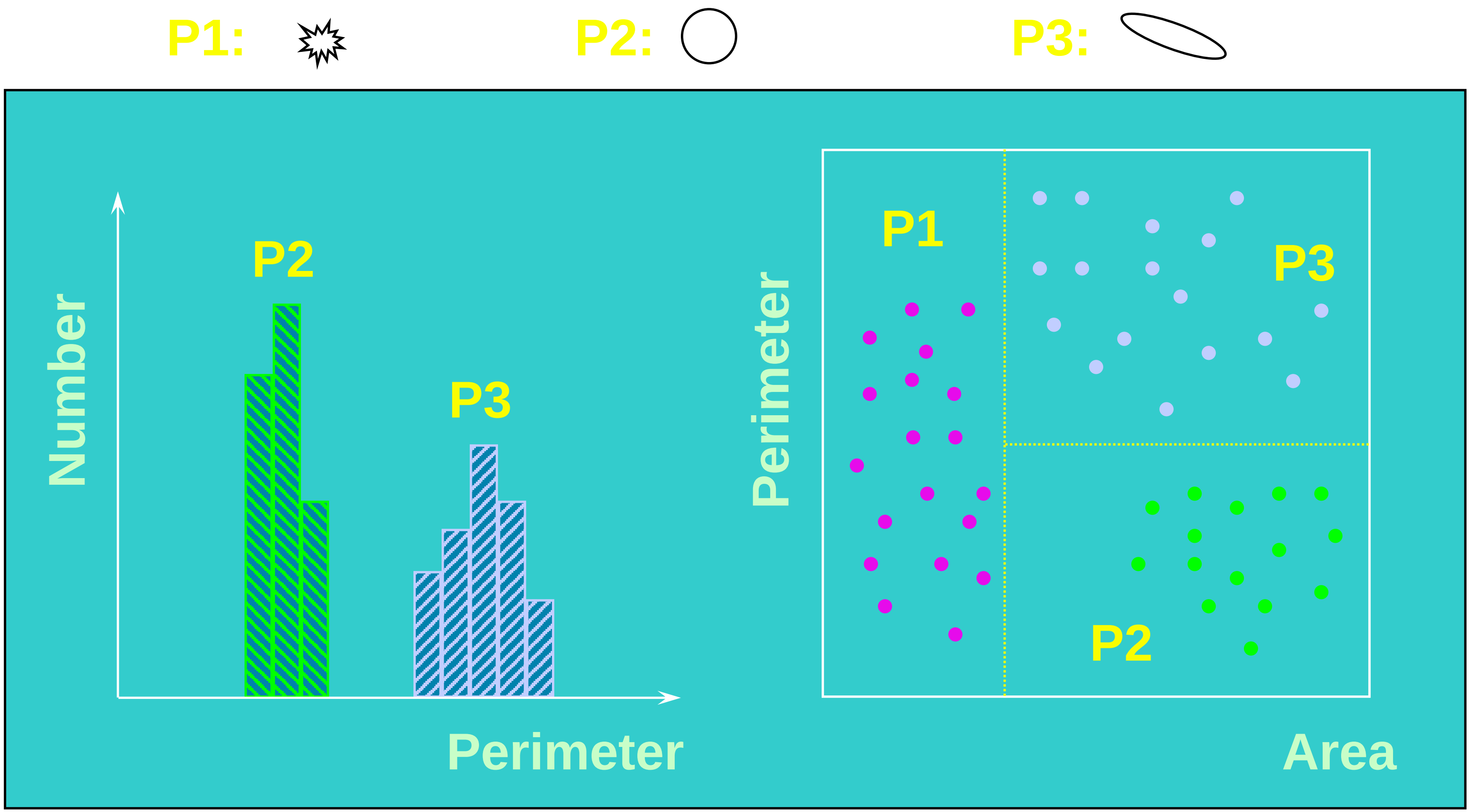


Histogram Analysis

P1:  P2:  P3: 



Histogram Analysis



Sample Data Set

Features **Class**

Area	Perimeter	Class
3	6	P1
5	7	P1
4	4	P1
7	6	P1
12	11	P2
15	10	P2
14	12	P2
17	13	P2
14	19	P3
13	20	P3
15	22	P3
12	18	P3
...	...	...

Design set:
Odd-indexed entries

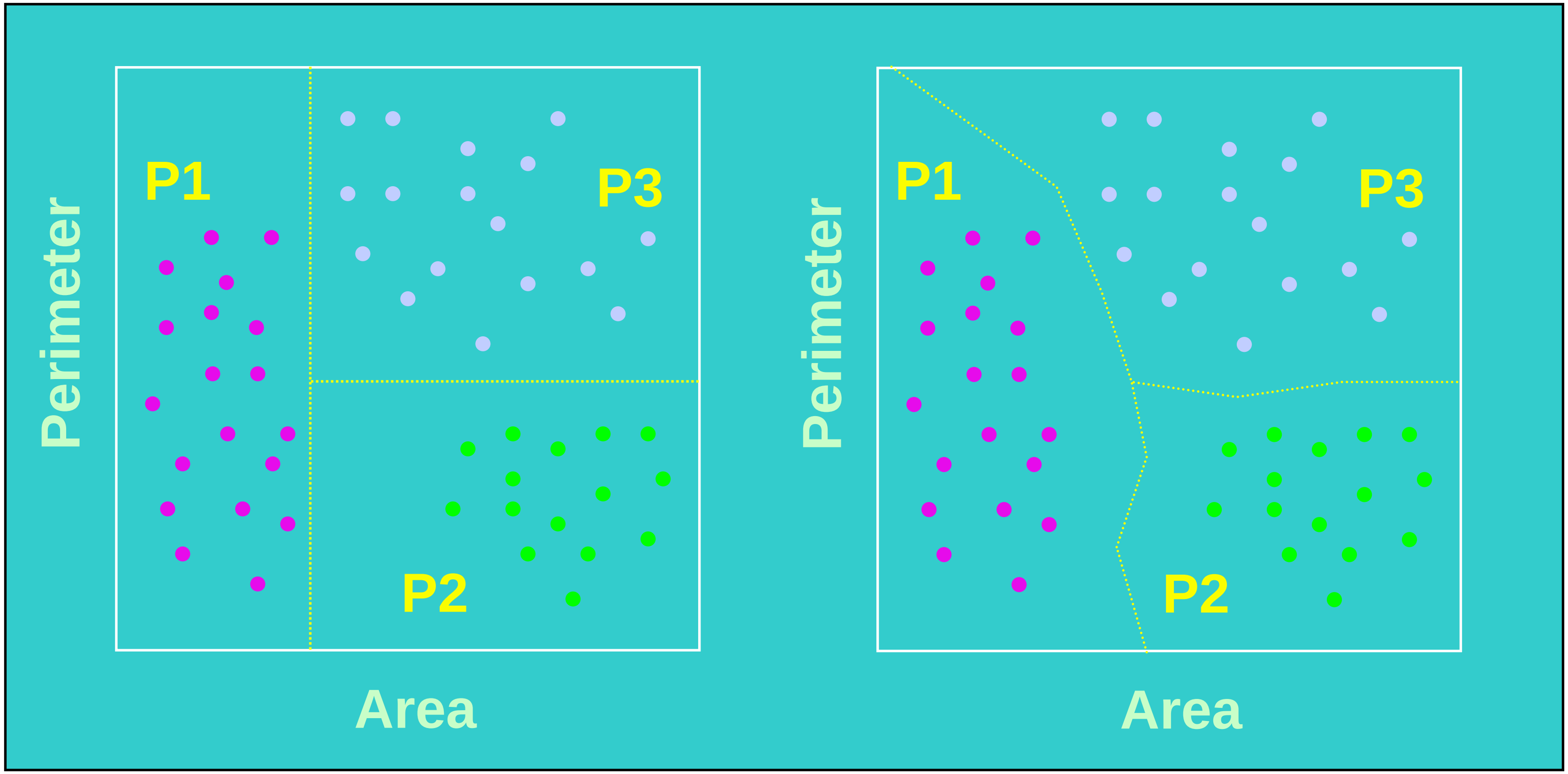
Test set:
Even-indexed entries

Flowchart for Analysis

General flowchart:	Particle example:
Feature extraction	From image to features
Data reduction	None
Probability estimate	Analysis

Input Space Partitioning

P1: P2: P3:



Fish Classification Example

"Sorting incoming Fish on a conveyor according to species using optical sensing"

Sea bass

Species ➔

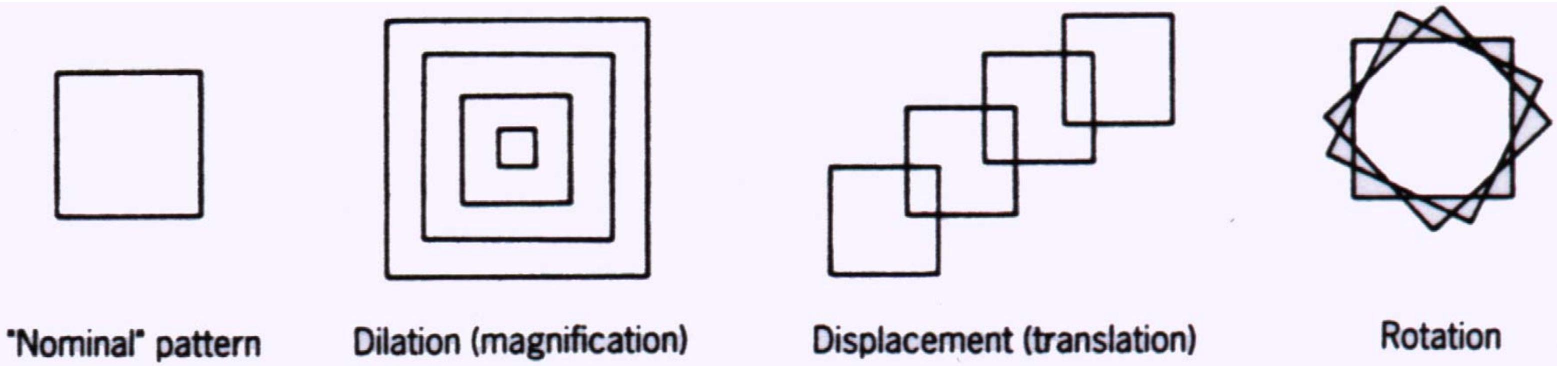
Salmon

Problem Analysis

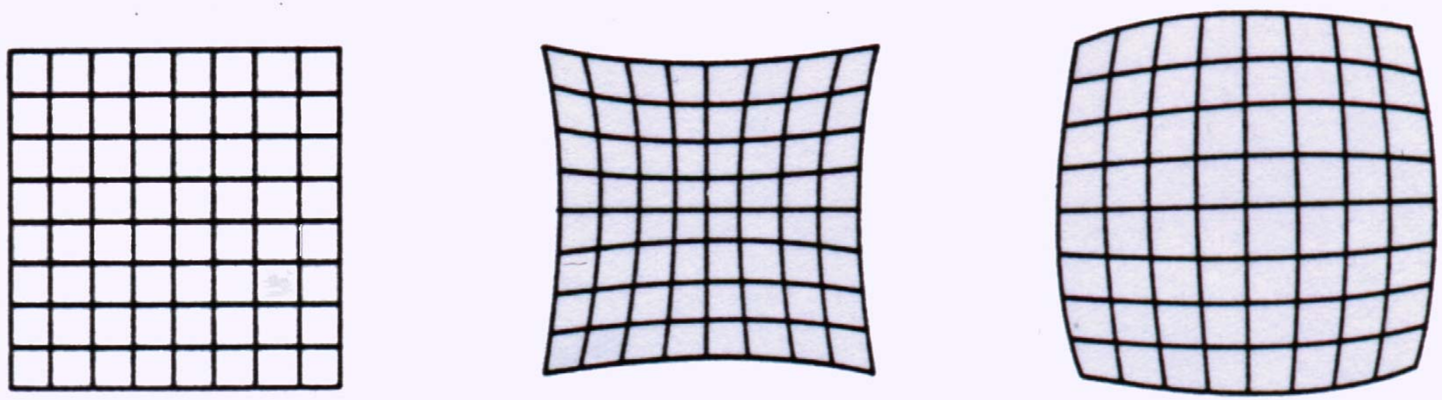
- Set up a camera and take some sample images to extract features
    - Length
    - Lightness
    - Width
    - Number and shape of fins
    - Position of the mouth, etc…

This is the set of all suggested features to explore for use in our classifier.

Possible Distortion



(a)



(b)

Possible Distortion

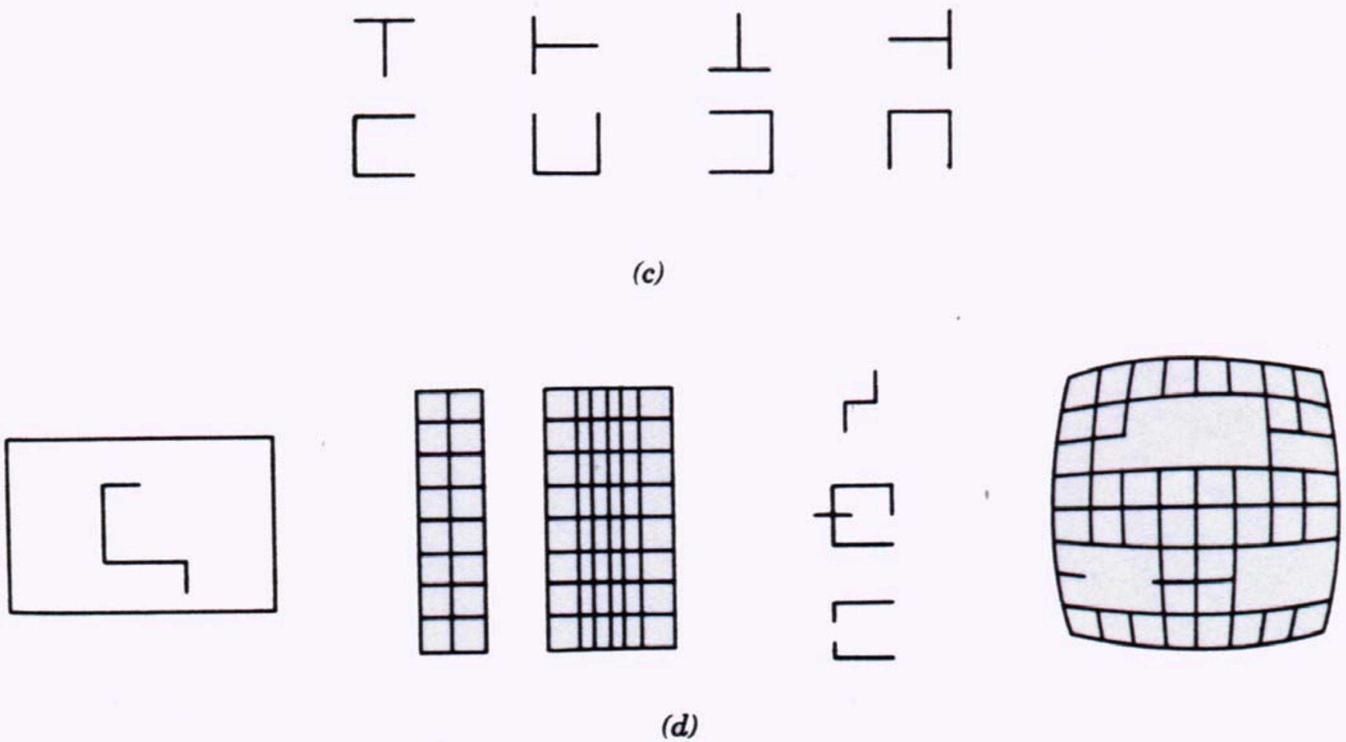
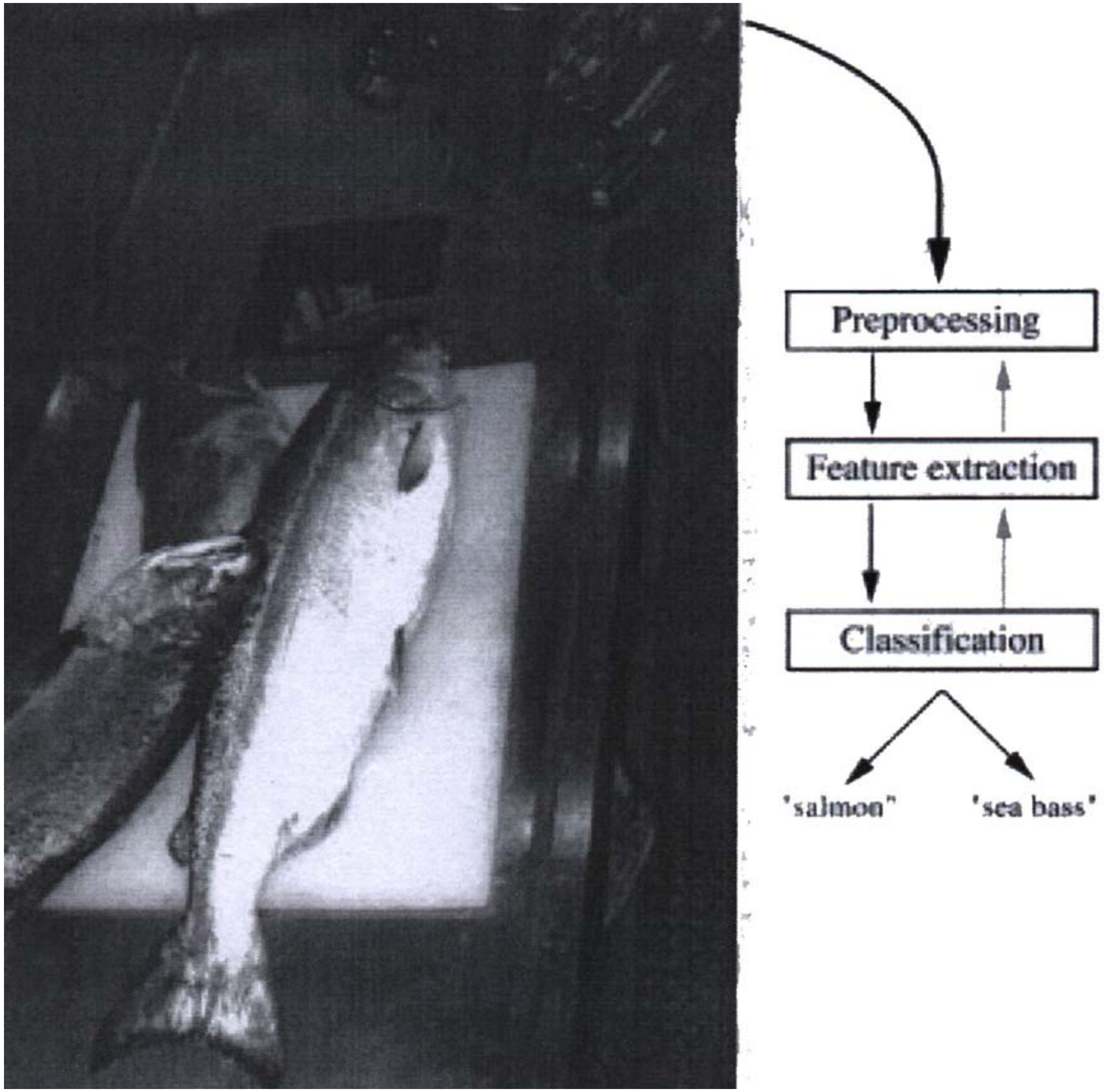


Figure 4: Examples of pattern distortions (visual patterns).
(a) Geometric distortions of visual (image) patterns-square.
(b) Geometric distortions of grid pattern.
(c) Geometric distortions of character patterns.
(d) More extreme pattern distortions (missing parts and extra parts) using the patterns of parts (a) to (c).

Preprocessing

- Use a segmentation operation to isolate fishes from one another and from the background
- Information from a single fish is sent to a feature extractor whose purpose is to reduce the data by measuring certain features
- The features are passed to a classifier

Fish Classification Example



Classification

- Select the length of the fish as a possible feature for discrimination

Fish Classification Example

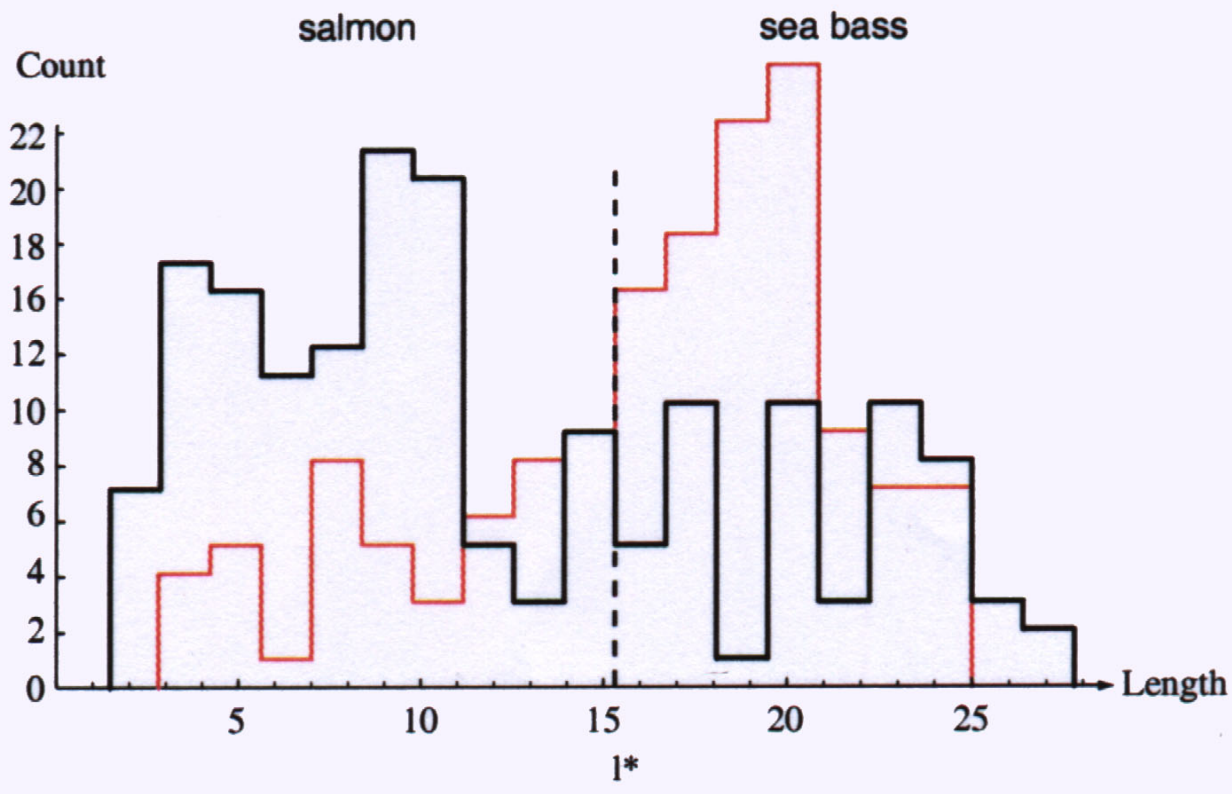


Figure 1.2: Histograms for the length feature for the two categories. No single threshold value l^* (decision boundary) will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value l^* marked will lead to the smallest number of errors, on average.

Fish Classification Example

- The *length* is a poor feature alone

- Select the *lightness* as a possible feature

Fish Classification Example

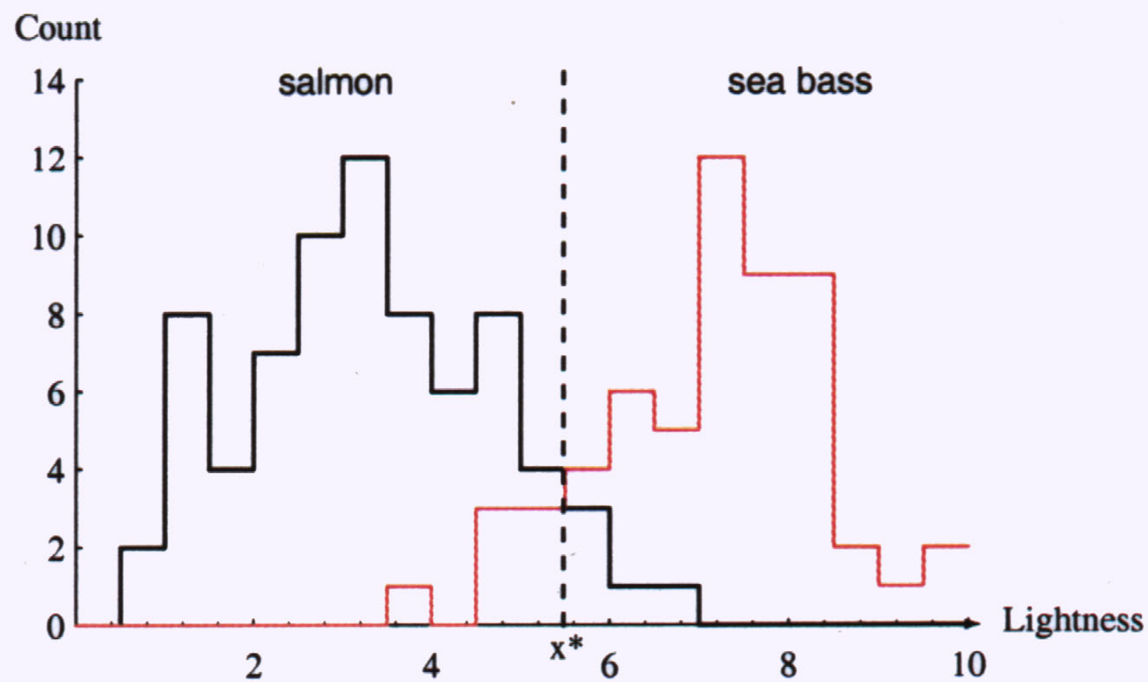


Figure 1.3: Histograms for the lightness feature for the two categories. No single threshold value x^* (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value x^* marked will lead to the smallest number of errors, on average.

Task of Decision Theory

- **Decision boundary and cost relationship**
  - **Move the decision boundary toward smaller values of lightness in order to minimize the cost (reduce the number of sea bass that are classified salmon)**

This is the task of decision theory

Two Features

- Adopt the lightness and add the width of the fish

Fish ➔ $x = [x_1, x_2]$ (Lightness, Width)

Decision Boundary

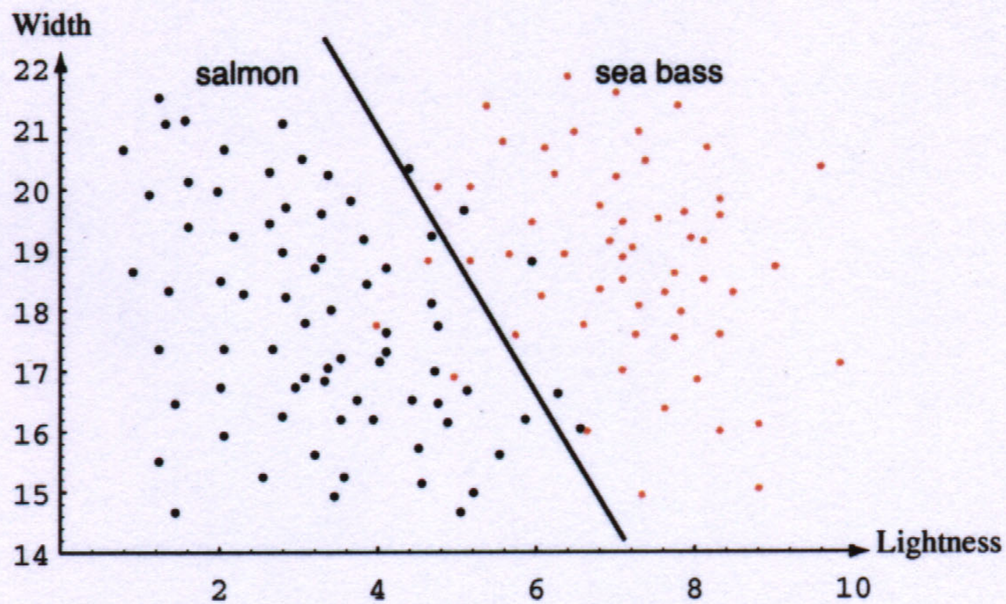


Figure 1.4: The two features of lightness and width for sea bass and salmon. The dark line might serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors.

More Features

- We might add other features that are not correlated with the ones we already have. A precaution should be taken not to reduce the performance by adding such "noisy features"

- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure:

Decision Boundary

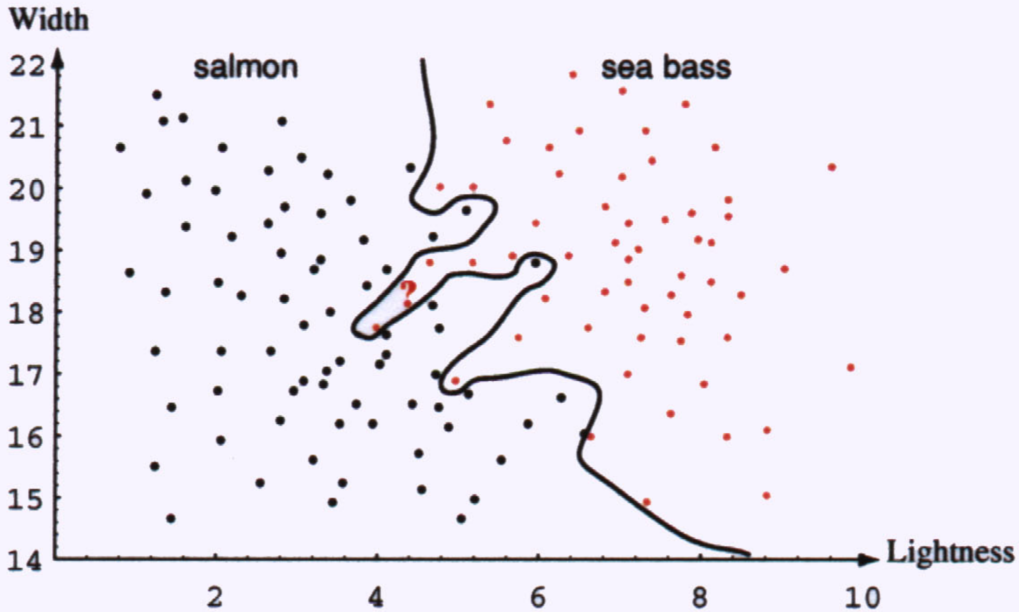


Figure 1.5: Overly complex models for the fish will lead to decision boundaries that are complicated. While such a decision may lead to perfect classification of our training samples, it would lead to poor performance on future patterns. The novel test point marked ? is evidently most likely a salmon, whereas the complex decision boundary shown leads it to be misclassified as a sea bass.

Issue of Generalization

- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel input.

Decision Boundary

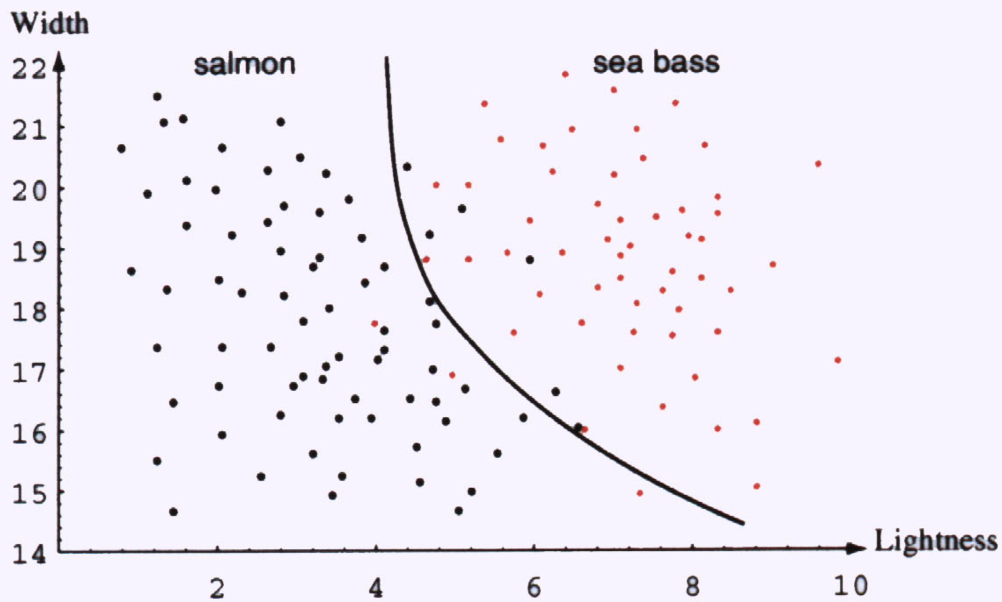
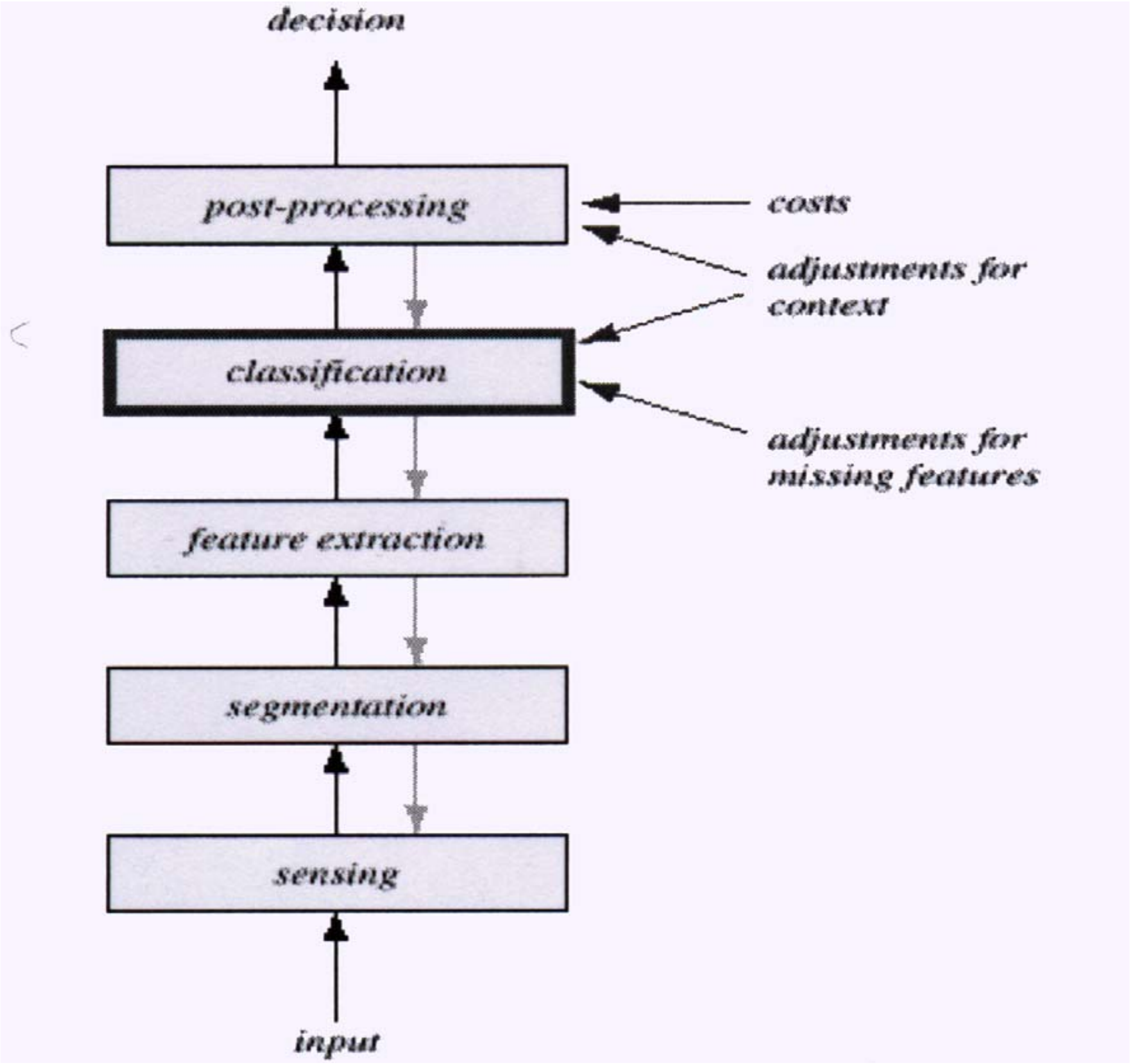


Figure 1.6: The decision boundary shown might represent the optimal tradeoff between performance on the training set and simplicity of classifier.

Pattern Recognition Systems

- **Sensing**
  - Use of a transducer (camera or microphone)
  - PR system depends on the bandwidth, the resolution, the sensitivity, and the distortion of the transducer
- **Segmentation and grouping**
  - Patterns should be well separated and should not overlap

Pattern Recognition Systems



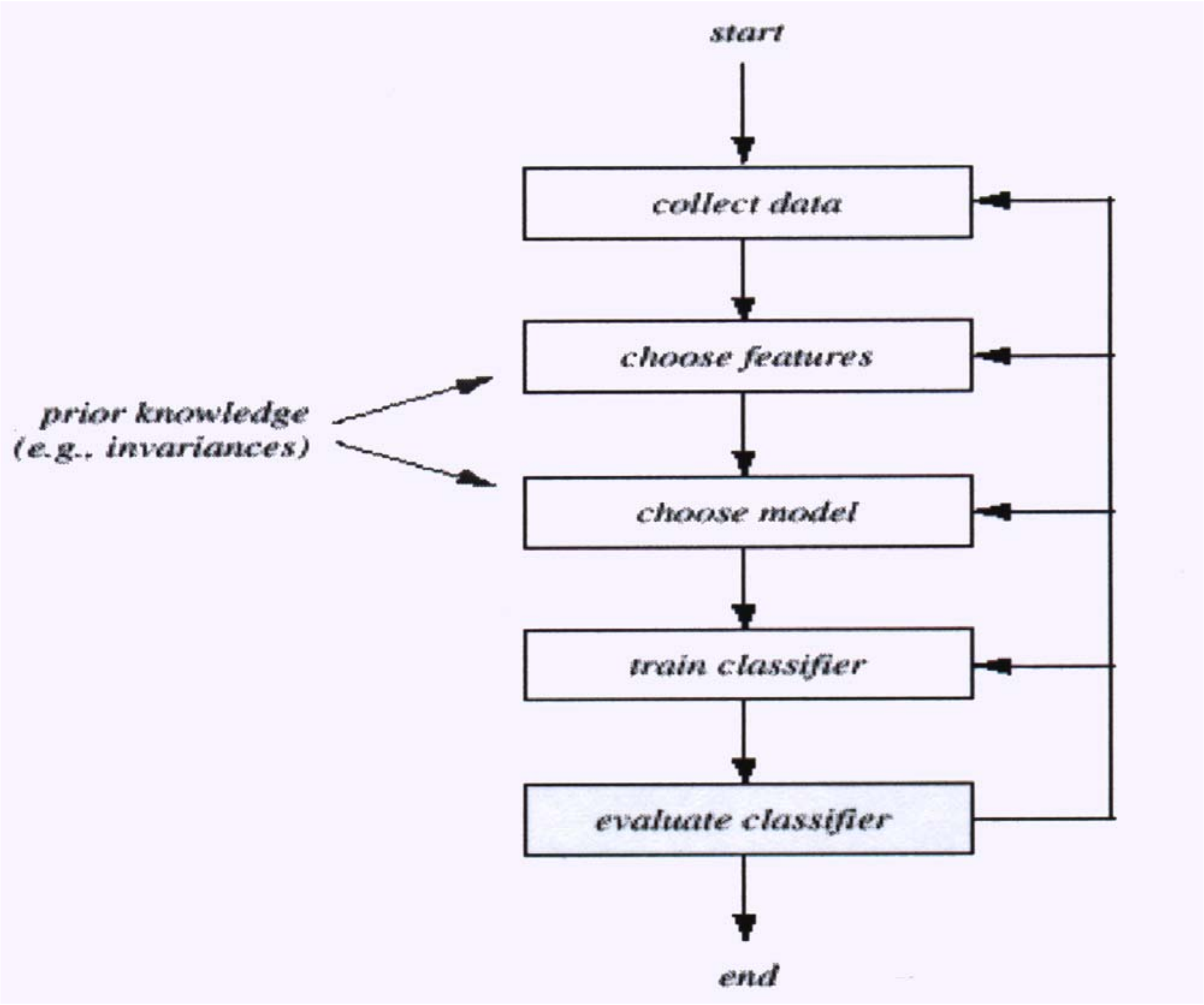
Pattern Recognition Systems

- **Feature extraction**
  - Discriminative features
  - Invariant features with respect to translation, rotation and scale.
- **Classification**
  - Use a feature vector provided by a feature extractor to assign the object to a category
  - Perfect classification is often impossible → to determine the probability for each of the possible categories
- **Post Processing**
  - Exploit context input dependent information other than from the target pattern itself to improve performance

The Design Cycle

- Data Collection
- Feature Choice
- Model Choice
- Training
- Evaluation
- Computational Complexity

The Design Cycle



The Design Cycle

Data Collection

- How do we know when we have collected an adequately large and representative set of examples for training and testing the system?

The Design Cycle

Feature Choice

- Depends on the characteristics of the problem domain.

- Preferably:
  - Simple to extract,
  - invariant to irrelevant transformation,
  - insensitive to noise,
  - useful for discriminating patterns in different categories.

The Design Cycle

Model Choice

- Unsatisfied with the performance of our fish classifier and want to jump to another class of model

The Design Cycle

Training

- Use data to determine the classifier. Many different procedures for training classifiers and choosing models

The Design Cycle

Evaluation

- Measure the error rate (or performance) and switch from one set of features to another one

The Design Cycle

Computational Complexity

- What is the trade off between computational ease and performance? (How an algorithm scales as a function of the number of features, patterns or categories?)

The Design Cycle

- **Learning and Adaptation**

- **Supervised learning**
  - **A teacher provides a category label or cost for each pattern in the training set**

- **Unsupervised learning**
  - **The system forms clusters or "natural groupings" of the input patterns**

Conclusion

- The number, complexity and magnitude of the sub-problems of PR should not be considered overwhelming
- Many of these sub-problems can be solved
- However, many fascinating unsolved problems still remain