Data processing framework for decision making

Jan Larsen

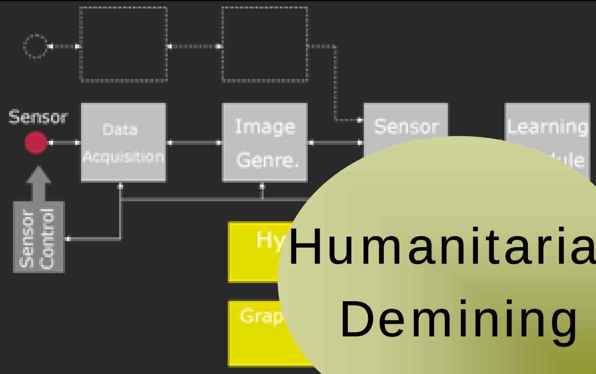
Intelligent Signal Processing Group

Department of Informatics and Mathematical Modelling

Technical University of Denmark

jl@imm.dtu.dk, www.imm.dtu.dk/~jl

ISP Group Activities



Multimedia

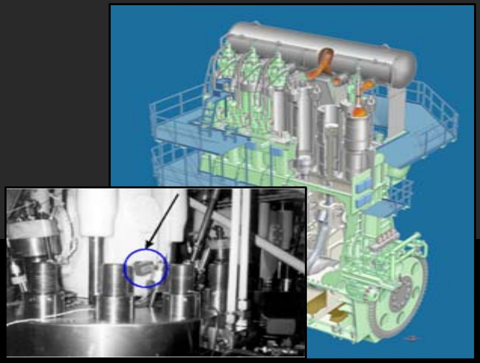
Humanitarian Demining

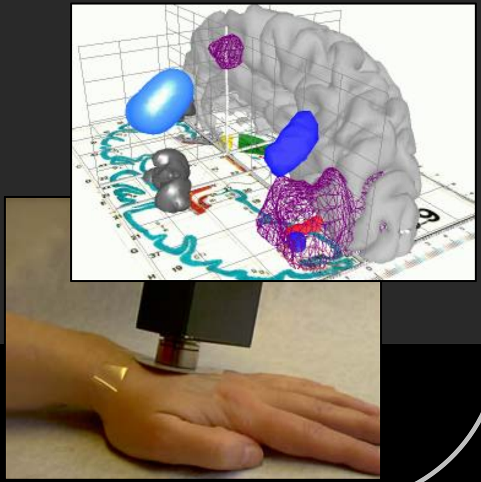
Neuroinformatics

Methods Algorithms

Monitor Systems

Biomedical





DTU

DeFuse

- Obtain general scientific knowledge about advantages (and drawbacks) in detection and clearance of mine like test objects by deploying a combined approach of complementary methods
- To lay the foundation for new practices

Xsense

- The scope of the Xsense program is to realize a reliable, sensitive, portable and low-cost explosive detector
- The detector will be miniaturized and will therefore be highly suitable for use in anti terror efforts, boarder control, environmental monitoring and demining
- The sensitivity will be optimized by a concentrated effort in data processing (reducing noise and pattern recognition) and emerging sensing principles
- The reliability of the detector will be ensured by combining several independent sensor technologies

Objective of this talk

- To provide insight into some of the issues in data processing and detection systems
- To hint at possible solutions using statistical signal processing and machine learning methodologies
- To facilitate the discussion – the good solution requires a cross-disciplinary effort

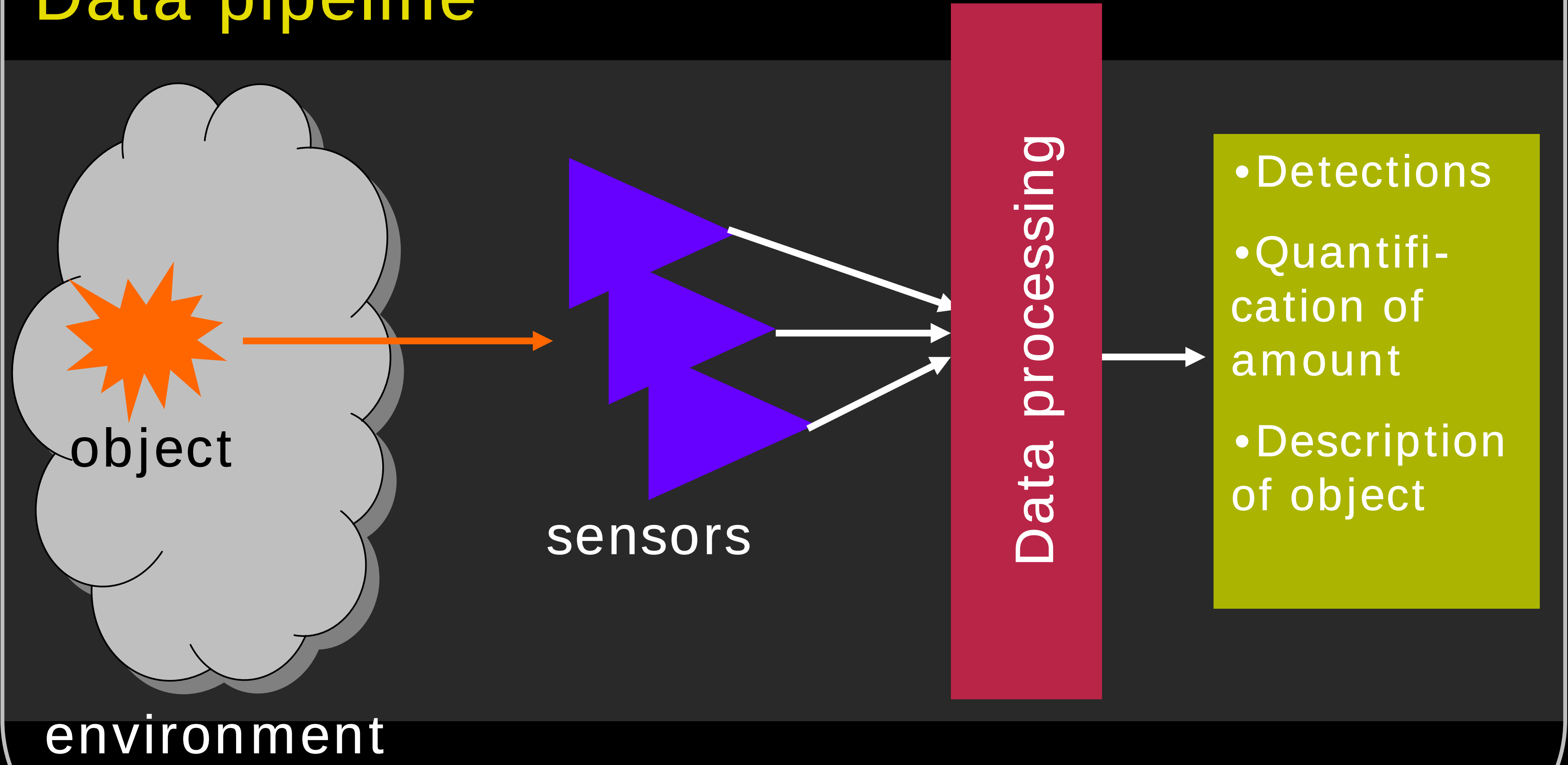
No math!

$$P(\theta \mid y) = \frac{P(y \mid \theta)p(\theta)}{P(y)}$$

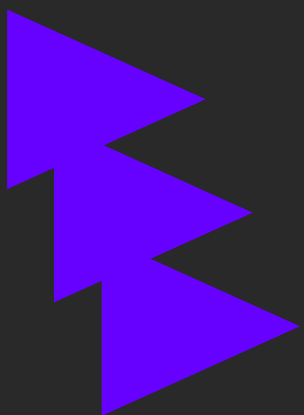
Outline

- The data processing pipeline
- Methods for taking up the challenge: reliable detection
- Summary

Data pipeline



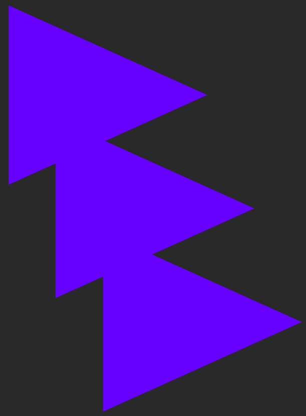
Sensing



sensors

- Sensing specific primary property of the object (e.g. odor component)
- Sensing a related property (e.g. reflected light)
- Sensing a mixture of properties – maybe only one is relevant
- Multiple sensors can sense different aspects

Sensing errors



sensors

- Various factors and other objects in the environment disturb the sensing
  - masking of related or primary property
  - other properties might be too strong
  - the environment is different from the environment in which the sensor was designed to work
- Errors in the sensors
  - Electrical noise
  - Drift
  - Degradation

Data processing

Data processing

- Extracting relevant features from sensor data
- Suppressing noise and error
- Segregation of relevant components from a mixture
- Integration of sensor data
- Prediction:
  - Presence of object
  - Classification of object type
  - Quantification of properties of the object (e.g. amount, size)
  - Description of object

Data processing errors

Data processing

- The sensed expression is too weak to make a reliable prediction of objects presence or quantification of an object property
- The processing device misinterprets the sensed expression
  - Maybe an unknown object in the environment
  - Not able to sufficiently suppress noise and errors
  - The processing can never done with 100% accuracy

Outline

- The data processing pipeline
- Methods for taking up the challenge: reliable detection
- Summary

How do we construct a reliable detector?

- Empirical method: systematic acquisition of knowledge which is used to build a mathematical model
- Specifying the relevant scenarios and performance measures – end user involvement is crucial!!!
- Cross-disciplinary R&D involving very competences

Mathematical models are prevalent: you need them to generate reliable results in a real use case

Knowledge acquisition

Physical modeling

- Study physical properties and mechanism of the environment and sensors
- Describe the knowledge as a mathematical model

Statistical modeling

- Require real world related data
- Use data to learn e.g. the relation between the sensor reading and the presence/absence of explosives

Why do we need statistical models?

> Scientist and engineers are born sceptical: they don't believe facts unless they see them often enough

- The process is influenced by many uncertain factors which makes classical physical modeling insufficient
- We can never achieve 100% accuracy – hence an estimate of the reliability is needed

There is no such thing as facts to spoil a good explanation!

- Pitfalls and misuse o tatistical methods sometimes w r nclusion that they are of little pr c

U si
n ev
h az

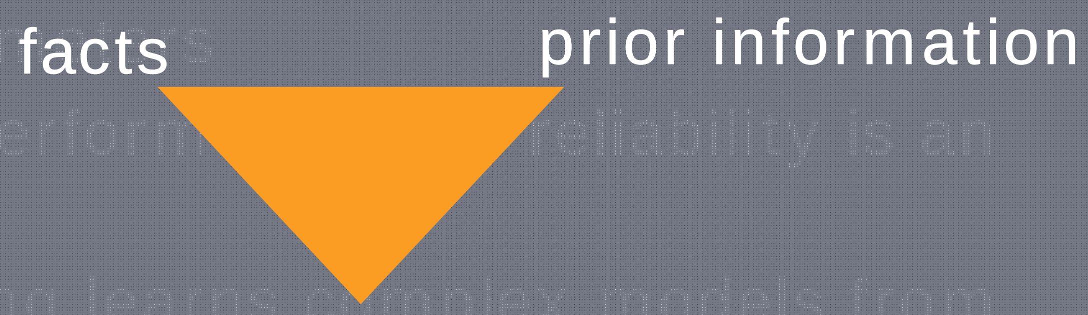
Some data are in the tail of the distribution: generalization from few examples is not possible

The number of hazardous objects is very small

his live

Why do we need statistical models and machine learning?

- statistical modeling is the principled framework to handle uncertainty and complexity
- Statistic modeling usually focuses on identifying important parameters
- Estimation of performance and reliability is an integral part
- machine learning learns complex models from collections of data to make optimal predictions in new situations



consistent and robust information and decisions with associated risk estimates

Three examples of using statistical modeling

- Reliable detection
- Increasing detection rate by combining sensors
- Segregation of mixed signals in order to reduce disturbances

Reliable detection of hazardous object – tossing a coin

$$Frequency = \frac{\text{no of heads}}{\text{no of tosses}}$$

$$probability = frequency \text{ when infinitely many tosses}$$

To achieve 99,6% detection probability

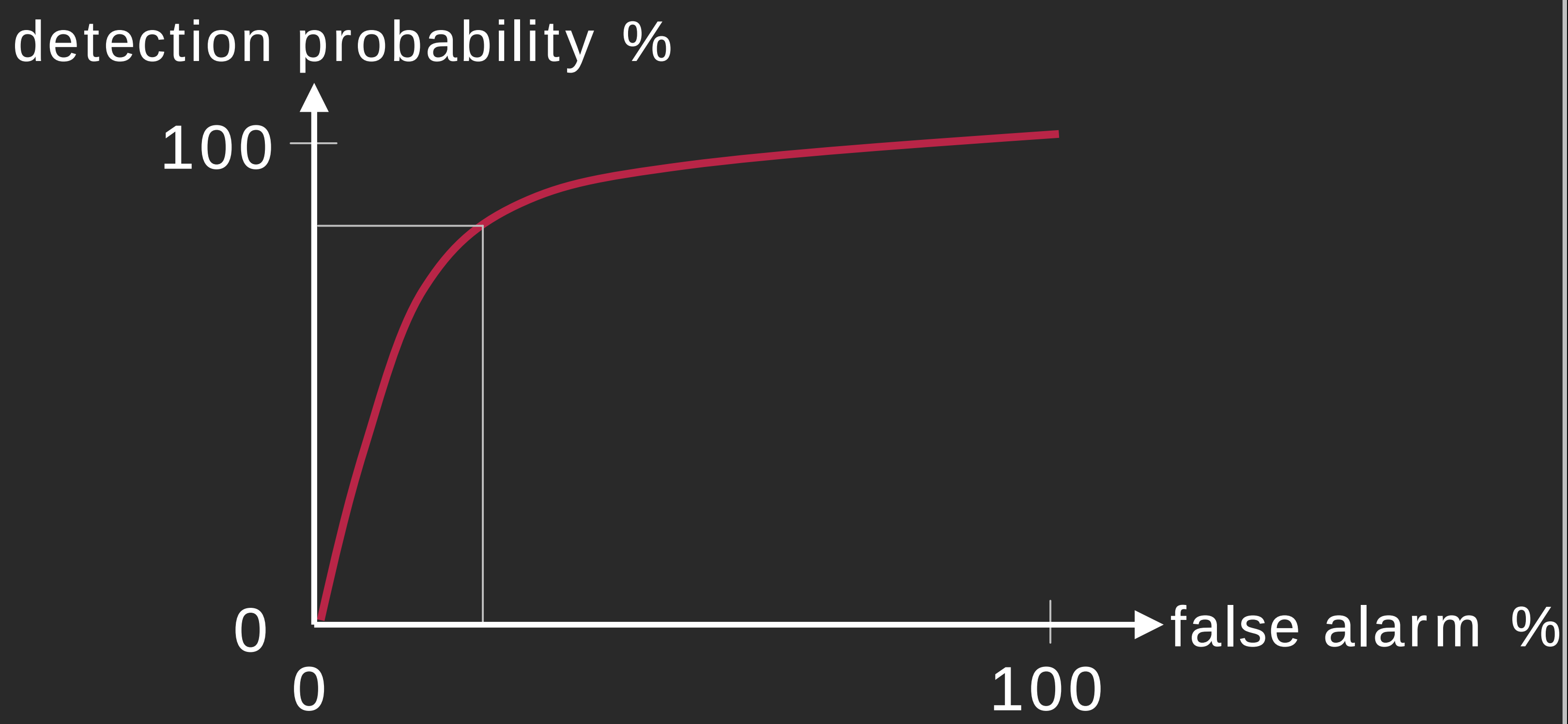
$$Frequency = \frac{996}{1000} = 99{,}6\%$$

$$Frequency = \frac{9960}{10000} = 99{,}60\%$$

One more (one less) count will change the frequency a lot!

You need 747 examples to be 95% sure that detection is better than 99,6% even if you detected all cases

Receiver operation characteristic (ROC)



Two types of errors in relation to ROC

Example: odor

•Sensing error: t
has little explosiv
content

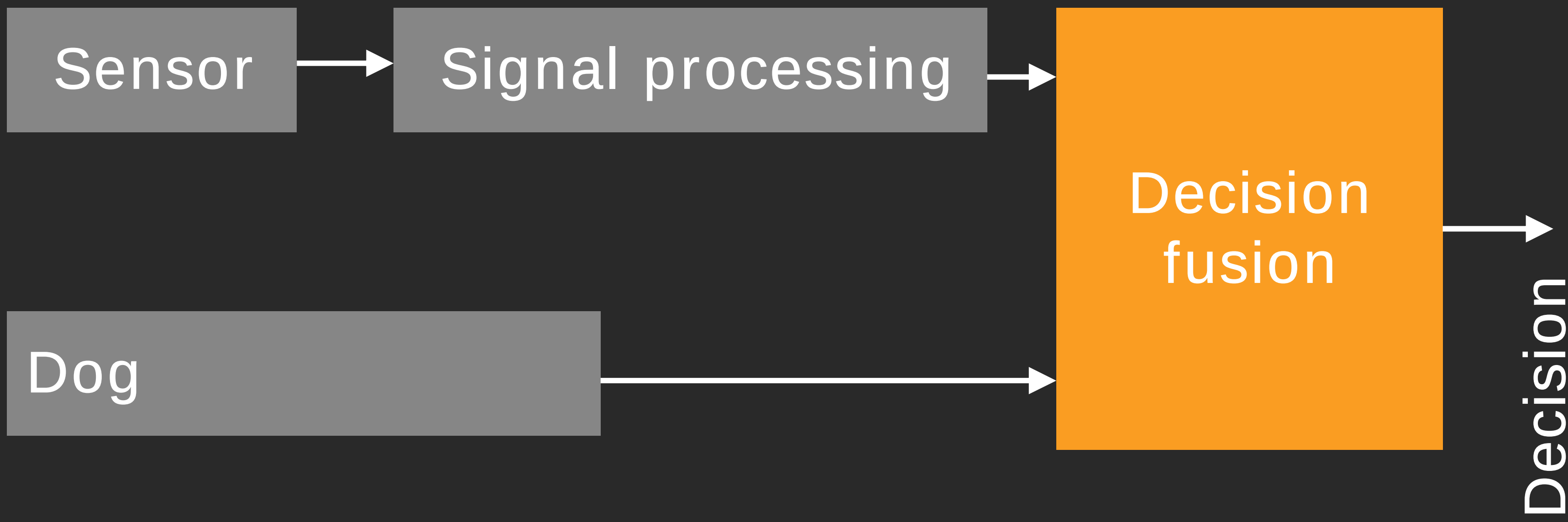
•Decision error:
bee-wax was fou

Example: dog

•Sensing error: the TNT leakage from the object was too low

•Decision error: the dog handler misinterpreted the dogs indication

Late integration – decision fusion



Independent error assumption

- Combination leads to a possible exponential increase in detection performance
  - System 1: 80%
  - System 2: 70%
  - Combined system: 94%
- Combination leads to better robustness against changes in environmental conditions

Segregation of signals

- Independent Component Analysis of audio signals
  - Cocktail Party Problem
  - Two people talking together, recording two mixtures
  - Example: Molgedey and Schuster’s algorithm (1994)

Summary

- A cross-disciplinary effort is required to obtain sufficient knowledge about physical, operational and processing possibilities and constraints as well as clear definition of a measurable goal – the right tool for the right problem
- Use of statistical modeling is a principled framework for
  - optimally combining all available information and sensor data
  - handling uncertainty
  - enhancing robustness