AARHUS
UNIVERSITET

Computer Vision & Machine Learning

Alexandros Iosifidis
@
Department of Electrical and Computer Engineering
Aarhus University

This week

Introduction to:
- Computer Vision
- Machine Learning

The course:
- Structure
- Learning objectives
- Exams

Introduction to Image Processing

Computer Vision

Some definitions:

- Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images/videos.

Computer Vision

Some definitions:

- Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images/videos.
- From the perspective of engineering, it seeks to automate tasks that the human visual system can do:
  - Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from one/multiple images
  - It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding.

Computer Vision

Some definitions:

- As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.

- As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems

Computer Vision

Past and Present of Computer Vision:

- Computer Vision began (1960s) at universities that were pioneering artificial intelligence as an attempt to mimic the human visual system, as a steppingstone to endowing robots with intelligent behavior.

Computer Vision

Past and Present of Computer Vision:

- Computer Vision began (1960s) at universities that were pioneering artificial intelligence as an attempt to mimic the human visual system, as a steppingstone to endowing robots with intelligent behavior.
- Differences from Image Processing:
  - Computer Vision's target was to extract three-dimensional structure from images with the goal of achieving full scene understanding.

Computer Vision

Past and Present of Computer Vision:

- Computer Vision began (1960s) at universities that were pioneering artificial intelligence as an attempt to mimic the human visual system, as a steppingstone to endowing robots with intelligent behavior.
- Differences from Image Processing:
  - Computer Vision's target was to extract three-dimensional structure from images with the goal of achieving full scene understanding.
- Since it tries to understand the real world, Computer Vision is directly connected to Visual Perception and Machine Learning:

Computer Vision

Past and Present of Computer Vision:

- Computer Vision began (1960s) at universities that were pioneering artificial intelligence as an attempt to mimic the human visual system, as a steppingstone to endowing robots with intelligent behavior.
- Differences from Image Processing:
  - Computer Vision's target was to extract three-dimensional structure from images with the goal of achieving full scene understanding.
- Since it tries to understand the real world, Computer Vision is directly connected to Visual Perception and Machine Learning:
  - Visual Perception is the ability to interpret the surrounding environment

Computer Vision

Past and Present of Computer Vision:

- Computer Vision began (1960s) at universities that were pioneering artificial intelligence as an attempt to mimic the human visual system, as a steppingstone to endowing robots with intelligent behavior.
- Differences from Image Processing:
  - Computer Vision's target was to extract three-dimensional structure from images with the goal of achieving full scene understanding.
- Since it tries to understand the real world, Computer Vision is directly connected to Visual Perception and Machine Learning:
  - Visual Perception is the ability to interpret the surrounding environment
  - Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed

Computer Vision

Connection with other fields in Visual Perception:

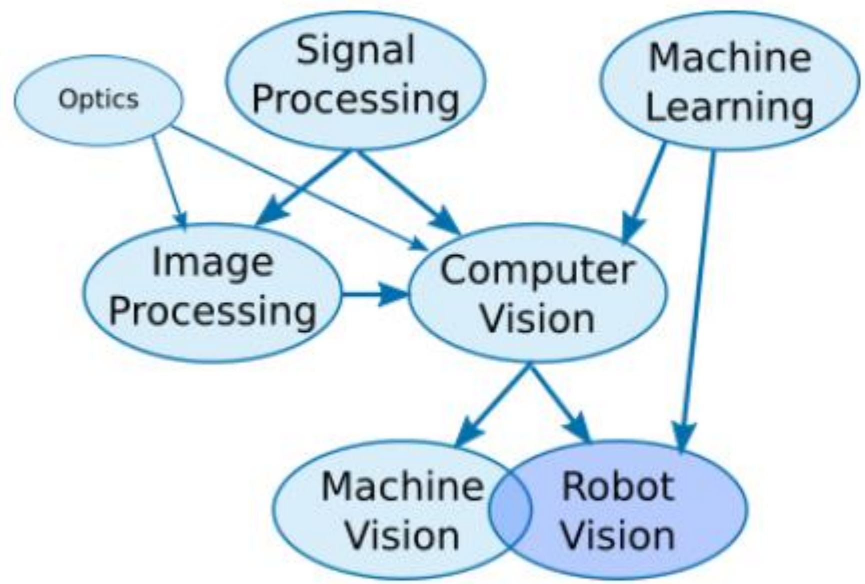
- Related fields (and usually confused as being the same) are:
  - Machine Vision
  - Robot Vision
  - Image Analysis (Processing)
  - Signal Processing

Computer Vision

Connection with other fields in Visual Perception:

- Related fields (and usually confused as being the same) are:
  - Machine Vision
  - Robot Vision
  - Image Analysis (Processing)
  - Signal Processing

A relation scheme drawn from here



Machine Learning

<u>Some definitions in the form of Q/A:</u>

- What is Artificial Intelligence?

Machine Learning

Some definitions in the form of Q/A:

- What is Artificial Intelligence?
  - AI is intelligence displayed by machines

Machine Learning

Some definitions in the form of Q/A:

- What is Artificial Intelligence?
  - AI is intelligence displayed by machines
  - The act of a machine perceiving its environment and taking actions that maximize its chance to succeed at some goal

Machine Learning

Some definitions in the form of Q/A:

- What is Artificial Intelligence?
  - AI is intelligence displayed by machines
  - The act of a machine perceiving its environment and taking actions that maximize its chance to succeed at some goal
  - The act of a machine that is mimicking cognitive functions that humans associate with other intelligent beings (mostly humans), such as problem solving and/through learning

Machine Learning

Some definitions in the form of Q/A:

- What is Machine Learning?

Machine Learning

Some definitions in the form of Q/A:

- What is Machine Learning?
  - ML is the field of computer science that gives computers the ability to learn without being explicitly programmed

Machine Learning

Some definitions in the form of Q/A:

- What is Machine Learning?
  - ML is the field of computer science that gives computers the ability to learn without being explicitly programmed
  - Learning is done in the form of:
    - estimating the optimal values (parameters) of a human-made model that tries to explain a real phenomenon/scenario using data

Machine Learning

Some definitions in the form of Q/A:

- What is Machine Learning?
  - ML is the field of computer science that gives computers the ability to learn without being explicitly programmed
  - Learning is done in the form of:
    - estimating the optimal values (parameters) of a human-made model that tries to explain a real phenomenon/scenario using data
    - teaching the machines by providing examples followed by good responses and/or a goodness metric

Machine Learning

Some definitions in the form of Q/A:

- What is Machine Learning?
  - ML is the field of computer science that gives computers the ability to learn without being explicitly programmed
  - Learning is done in the form of:
    - estimating the optimal values (parameters) of a human-made model that tries to explain a real phenomenon/scenario using data
    - teaching the machines by providing examples followed by good responses and/or a goodness metric
    - letting the machines to evolve by themselves through understanding their environment by analyzing data

Machine Learning

Some definitions in the form of Q/A:

- What is Data Analytics?

Machine Learning

Some definitions in the form of Q/A:

- What is Data Analytics?
  - is the discovery, interpretation and communication of meaningful patterns in data

Machine Learning

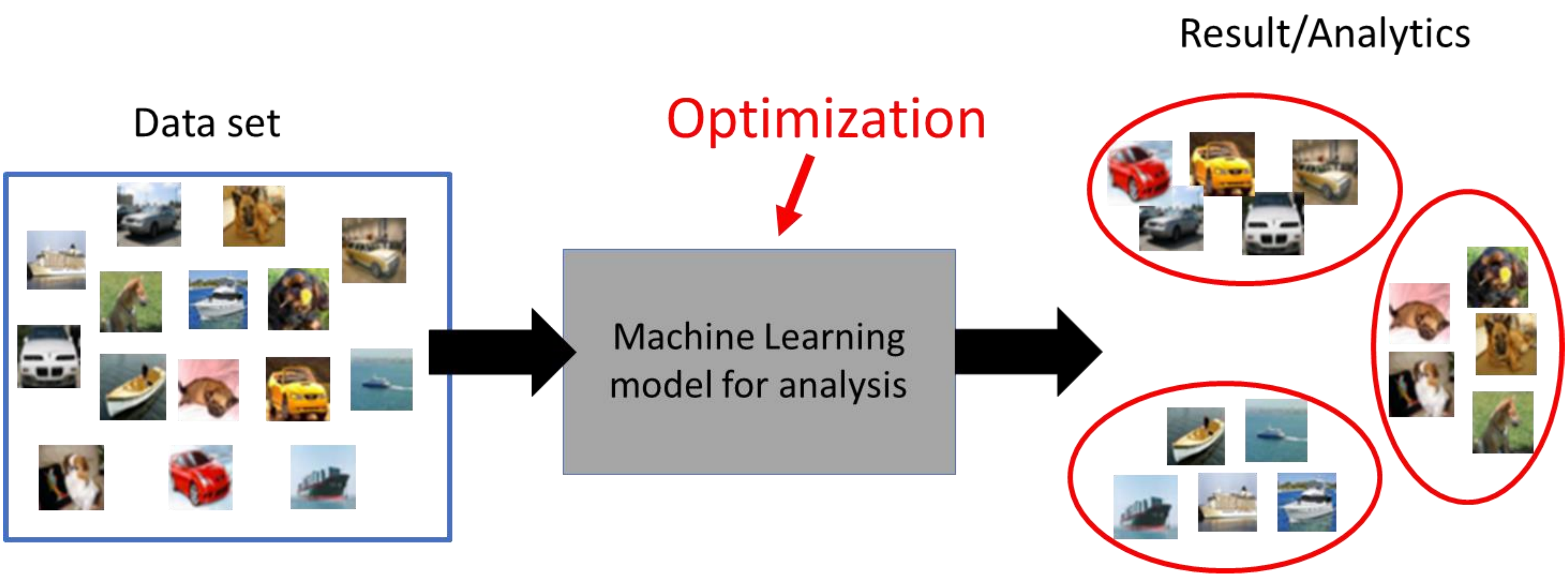
Some definitions in the form of Q/A:

- What is Data Analytics?
  - is the discovery, interpretation and communication of meaningful patterns in data

- Optimization:
  - is the process for selecting the best element (with regard to some criterion) from some set of available alternatives

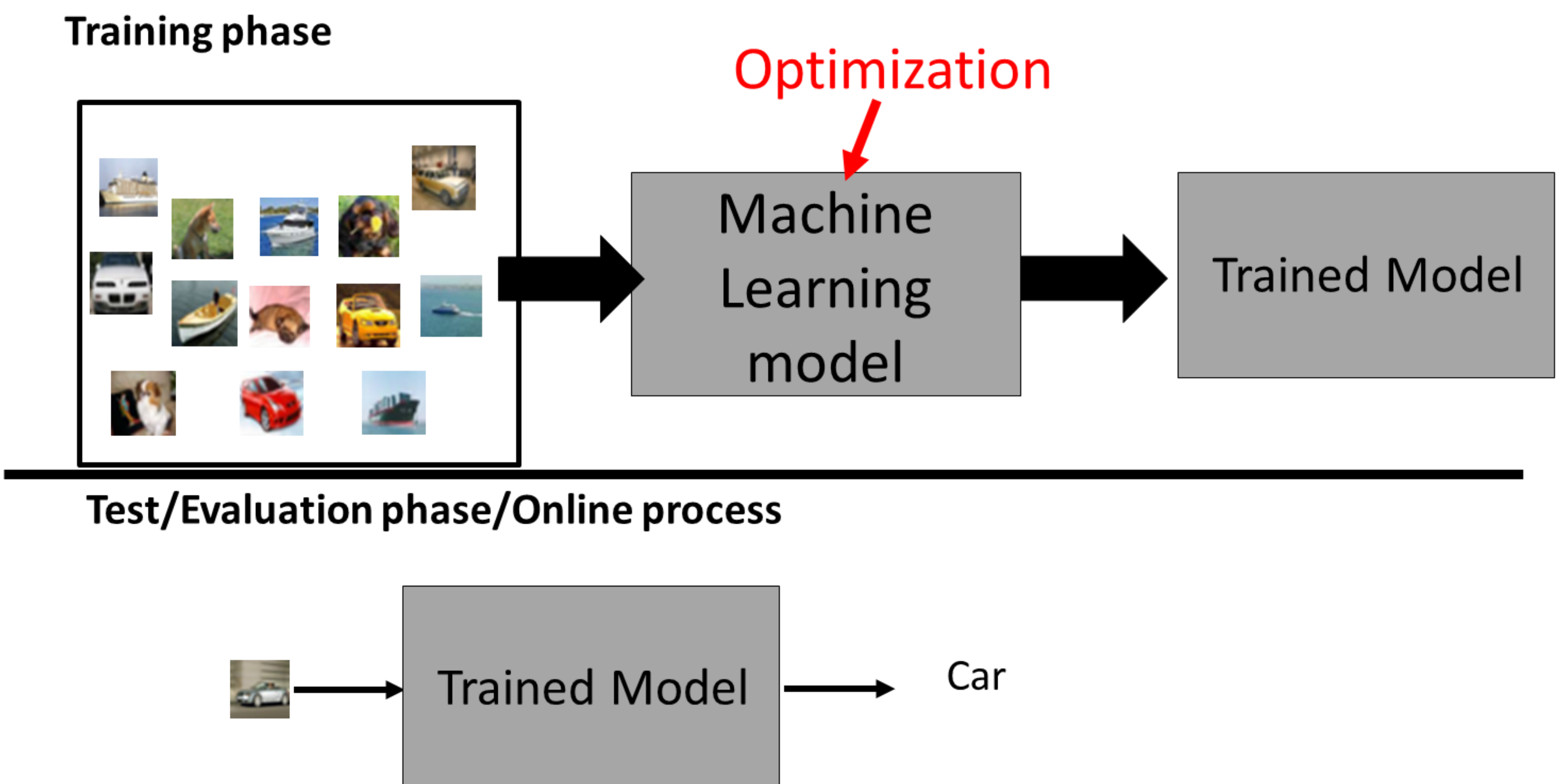
Connection of ML to Optimization

Data analysis (in-sample analysis)

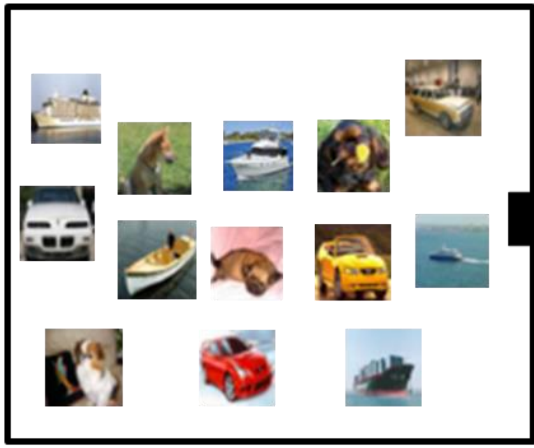


Connection of ML to Optimization

Learning (out-of-sample generalization)



Bottom-up view of classical ML models



Machine Learning model → Trained Model

Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

Data pre-processing
For example:
- Image segmentation
- Temp. video segm.

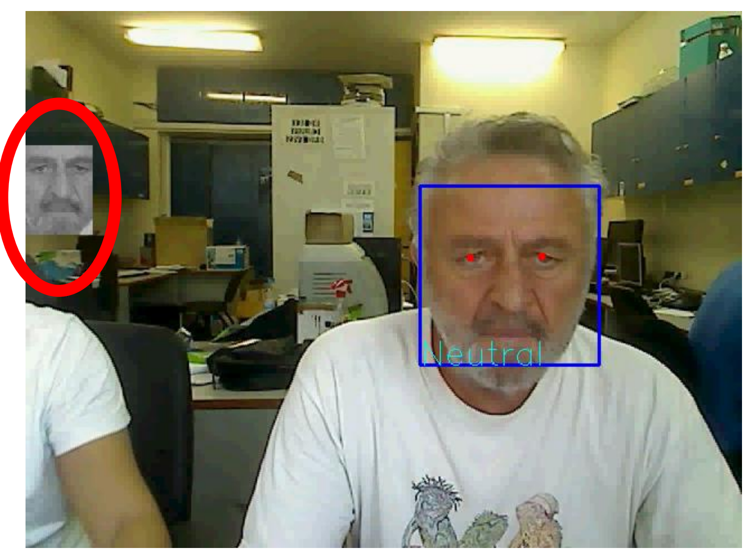
Data representation
For example:
- Vector per sample
- Vectors per sample

Repr. pre-processing
For example:
- Centering
- Standardization

Model selection/ training
For example:
- Classification
- Clustering

Bottom-up view of classical ML models

Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

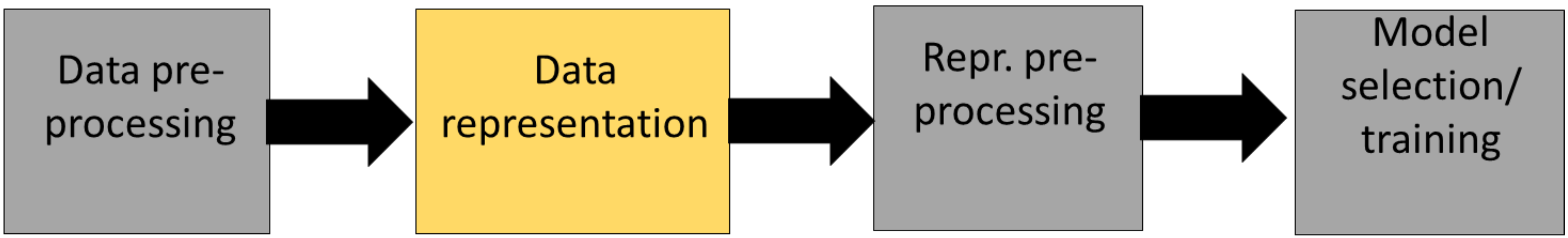


MOBISERV FP7-248434

Data pre-processing is a process which is guided by the domain of the problem to be solved.

For example in facial image analysis:
- Face detection and segmentation
- (Possibly) image frontalization
- Illumination normalization
- Facial image resizing

Bottom-up view of classical ML models

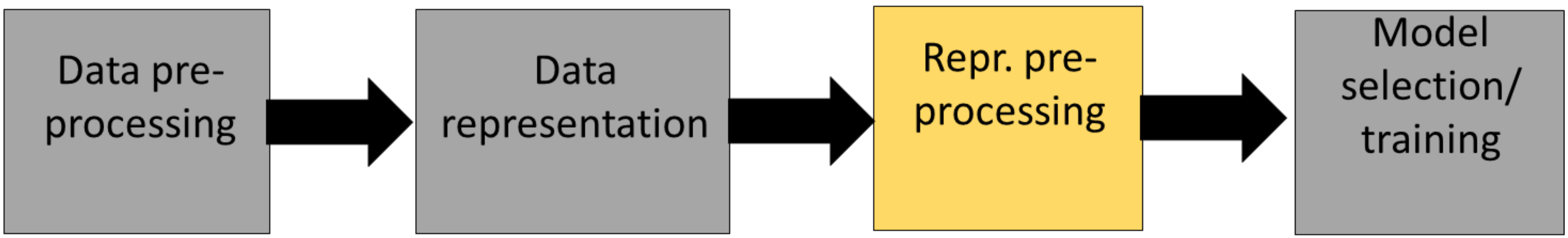


Data representations can also be guided by the domain of the problem to be solved.

For example we can represent a facial image based on:

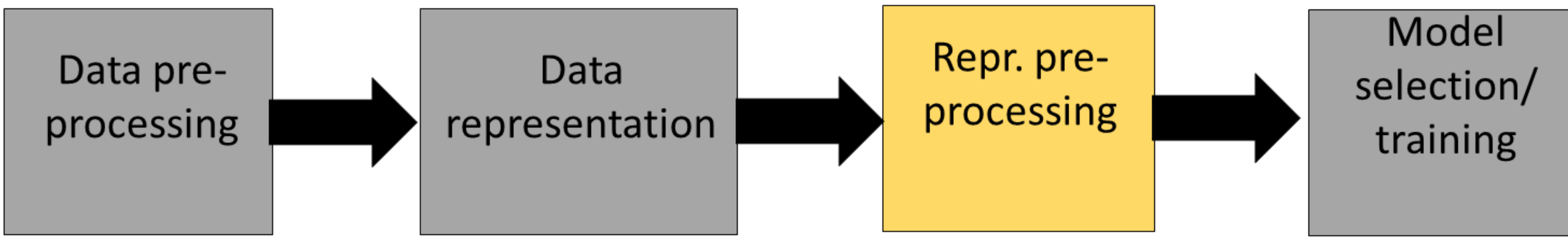
- Image intensities
- Edges → texture
- Locations and sizes of each facial feature
- A set of local features (SIFT, HOG, LBP) encoded, e.g. with Bag of Words model

Bottom-up view of classical ML models



After obtaining data representations, we (usually) obtain a D-dimensional vector **x** for each sample.

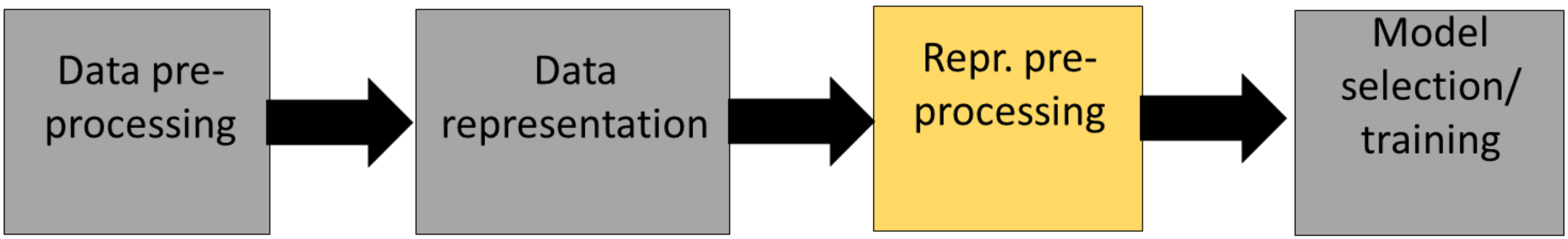
Bottom-up view of classical ML models



After obtaining data representations, we (usually) obtain a D-dimensional vector $\mathbf{x}$ for each sample.

Let us assume that we have a set of N samples. We use a subscript i=1,…,N to denote a specific sample. That is, we write $\mathbf{x}_i$ to denote the i-th sample in the data set. Some times it is convenient to write $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_N]$.

Bottom-up view of classical ML models

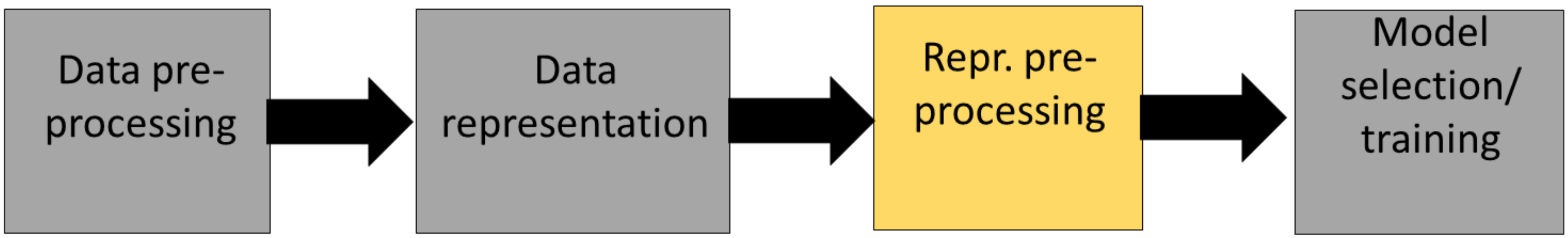


After obtaining data representations, we (usually) obtain a D-dimensional vector $\mathbf{x}$ for each sample.

Let us assume that we have a set of N samples. We use a subscript i=1,…,N to denote a specific sample. That is, we write $\mathbf{x}_i$ to denote the i-th sample in the data set. Some times it is convenient to write $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_N]$.

If D>2, sometimes it is convenient to 'map' the D-dimensional data to the 2-dimensional space. We say that we map (or project) the data from $\mathbb{R}^D$ to $\mathbb{R}^d$ (d<D), e.g. by applying a linear mapping $\mathbf{z}_i = \mathbf{W}^T \mathbf{x}_i$

Bottom-up view of classical ML models



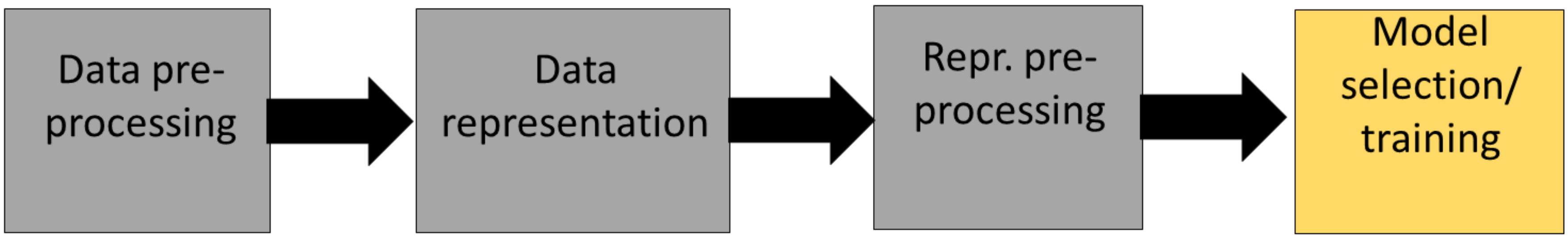
Other data representations pre-processing types:

Data centering $\bar{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}, \;\; i = 1, \ldots, N$, where $\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$

Data standardization $\hat{x}_{id} = \frac{x_{id} - \mu_d}{s_d}$, where $s_d = \frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(x_{id} - \mu_d)^2}$

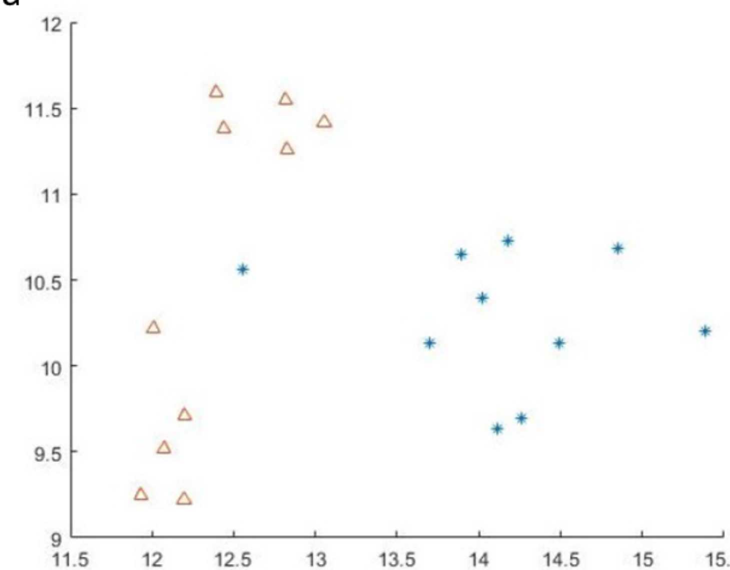
Data normalization (e.g. l2-norm) $\tilde{\mathbf{x}}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2} = \frac{\mathbf{x}_i}{\sqrt{\mathbf{x}_i^T \mathbf{x}_i}}$

Bottom-up view of classical ML models

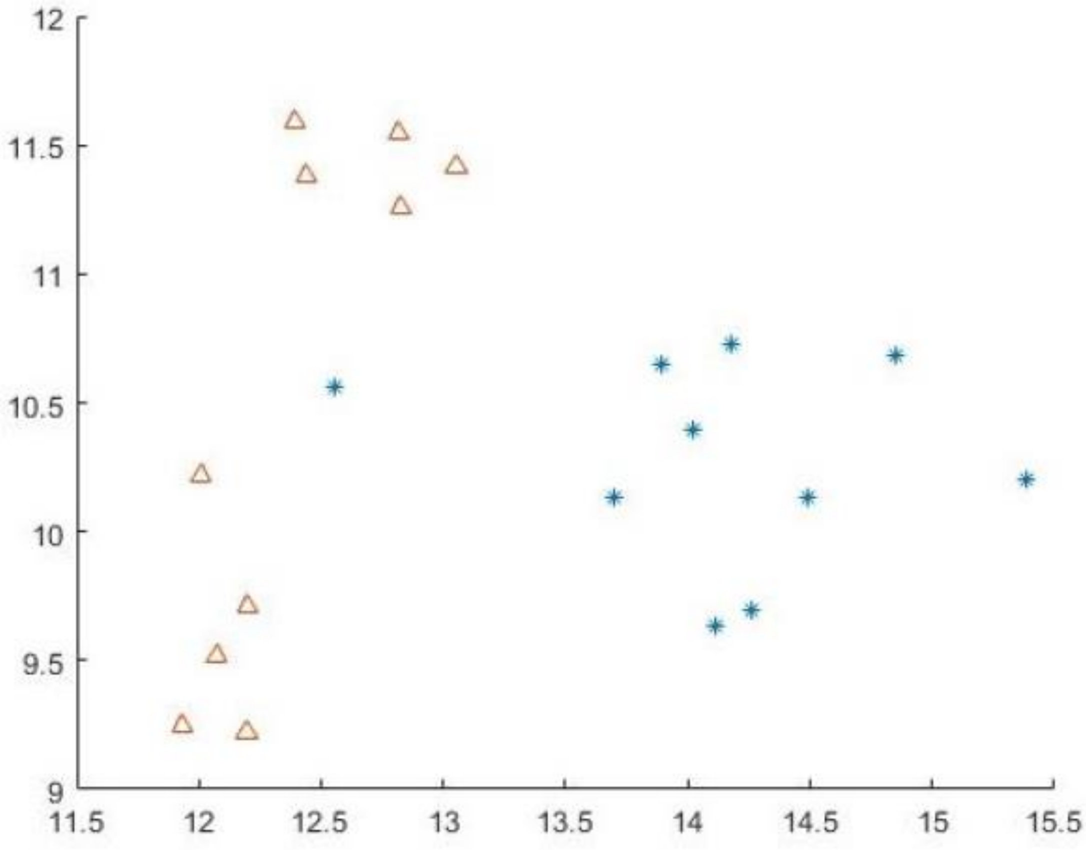


Let us assume that we have the following set of data

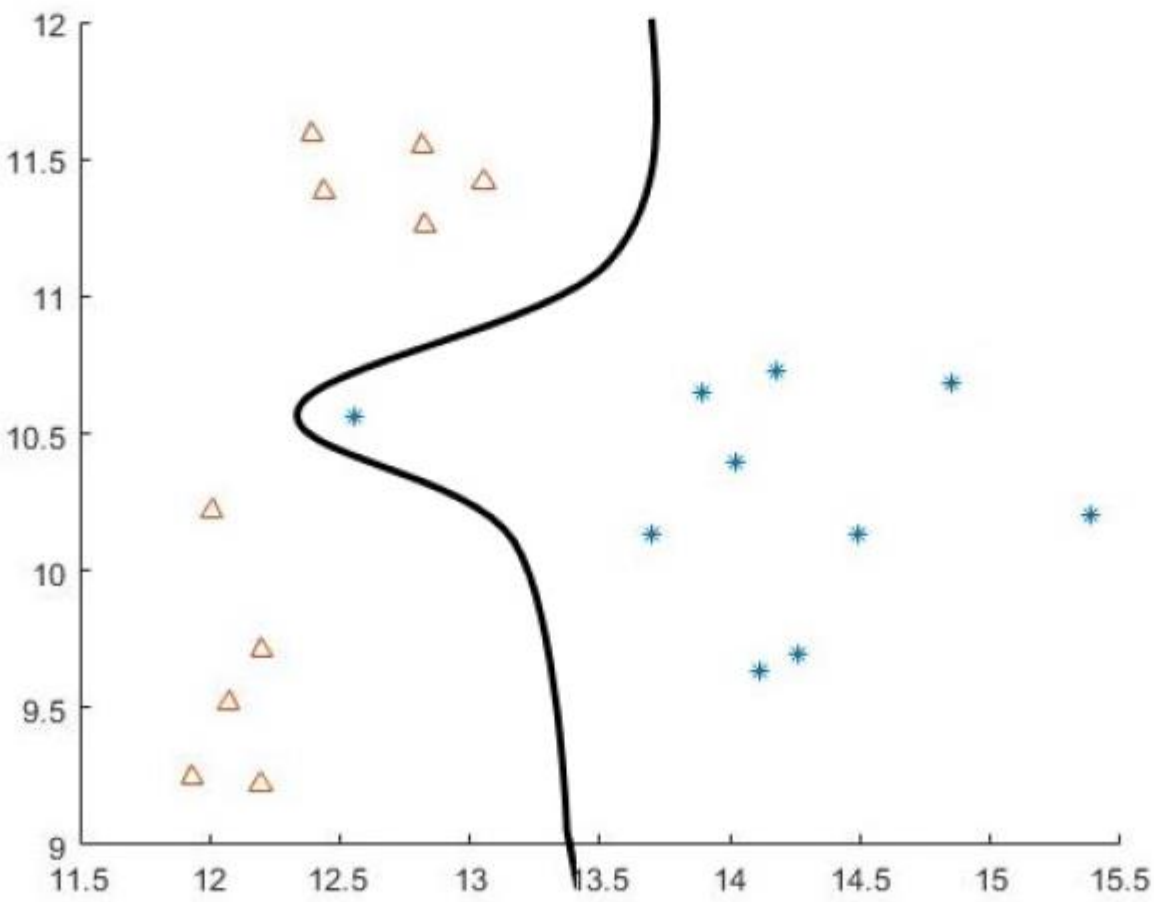
Can you draw a 'good decision function' classifying the stars and the triangles correctly?



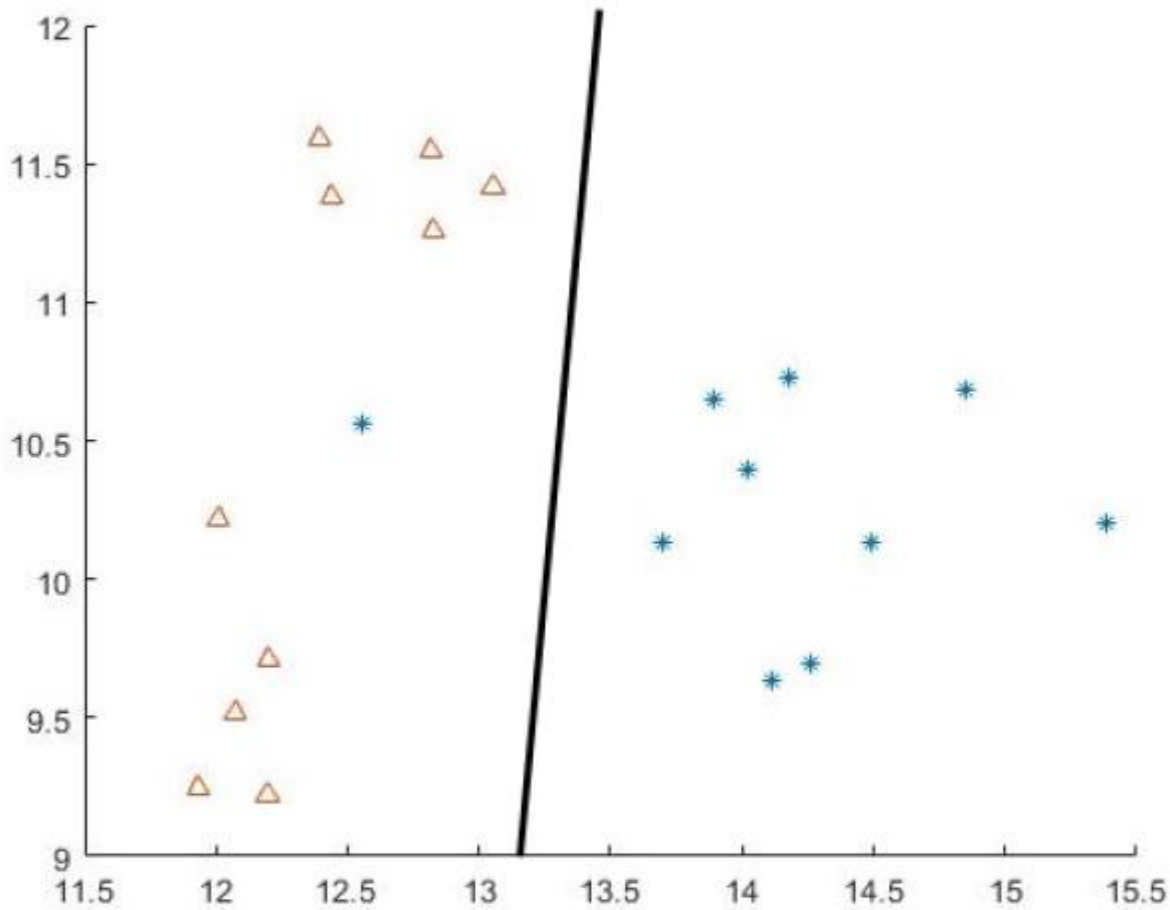
Model selection



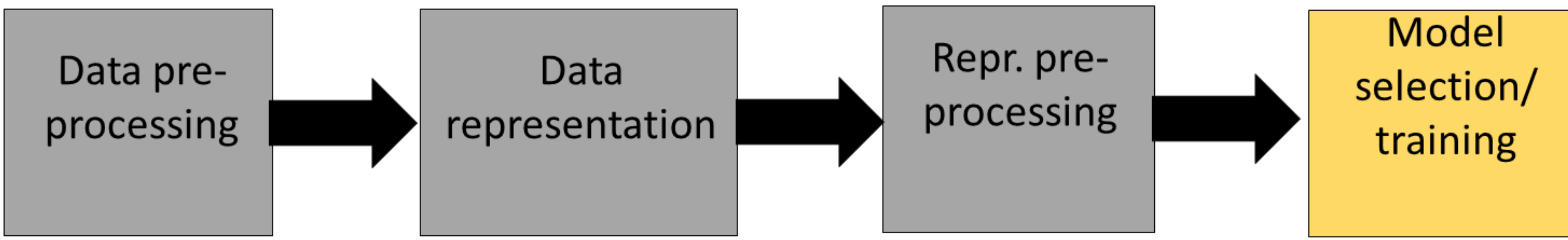
Model selection



Model selection



Bottom-up view of classical ML models



Model selection is the process of:

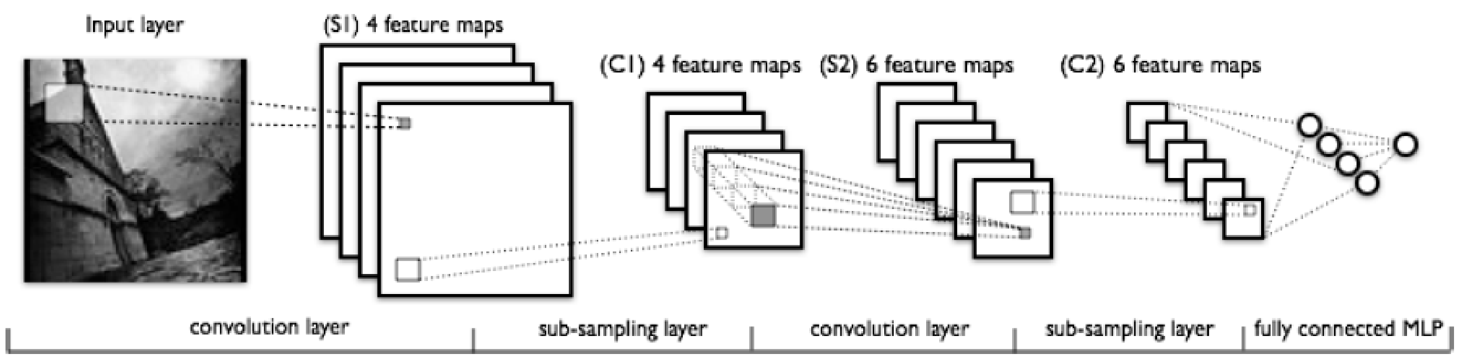
- Defining what technique will be used (e.g. linear classifier, nonlinear classifier, etc.)
- Defining the optimal parameters of the model. This includes:
    - The best architecture of the model (e.g. in neural networks)
    - The optimal values of the parameters of the model

End-to-end learning models

Classical machine learning models require human design and input at all stages of the pipeline

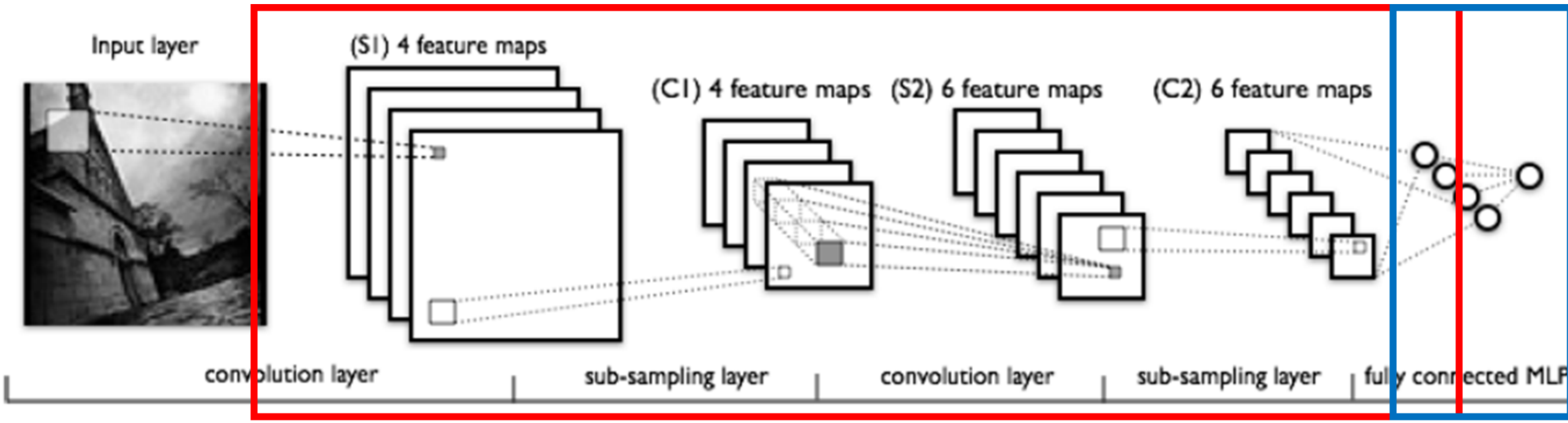
Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

End-to-end learning (also called data-driven) models require low human intervention



End-to-end learning models

In end-to-end learning models, all steps are optimized in a combined manner in order to obtain the best possible performance



Simultaneous:
- data pre-processing
- representation learning
- representation pre-processing

Classification/Regression:
- used to guide representation learning

Categorization of ML models

Three broad categories:

- Supervised Learning (labeling information available):
  - Training examples (e.g. N images)
  - Associated (class) labels (e.g. Image1 belongs to class ‘car’)

Categorization of ML models

Three broad categories:

- Supervised Learning (labeling information available):
  - Training examples (e.g. N images)
  - Associated (class) labels (e.g. Image1 belongs to class ‘car’)
- Unsupervised Learning (no labeling information):
  - There is the assumption that samples naturally form groups based on a similarity criterion and the model tries to identify the underlying groups and patterns

Categorization of ML models

Three broad categories:

- Supervised Learning (labeling information available):
  - Training examples (e.g. N images)
  - Associated (class) labels (e.g. Image1 belongs to class ‘car’)
- Unsupervised Learning (no labeling information)
  - There is the assumption that samples naturally form groups based on a similarity criterion and the model tries to identify the underlying groups and patterns
- Reinforcement Learning:
  - The model is guided by an expert who provides positive/ negative feedback
  - The model (agent) explores the space by making actions (guesses) and accumulating expert feedback
  - Given that feedback, the model is updated so that it can maximize its future reward

Unsupervised Learning

Principal Component Analysis for dimensionality reduction

Clustering with K-Means algorithm Link1

Supervised Learning

Linear Discriminant Analysis

Multi-layer Neural Network

Support Vector Machine

Support Vector Machine

Convolutional Neural Network

Convolutional Neural Network

Contents of the course

Structure of the course:

- 14 weeks (4 hours of lectures and 2 hours of exercises)
- Lectures on Tuesdays
- Hands-on exercises (in Matlab) on Fridays (illustration of some topics of the course)
- Project in the form of a Kaggle inClass competition and report

You can use any programming language for solving the exercises. However, since some of them are based on complex Computer Vision steps (provided by libraries in Matlab), it is advised to follow the provided structure and code.

Exams:

- Approval of the project is a pre-requisite for giving the exam.
- Oral examination

Contents of the course

1 Introduction and Intro to Image Processing

2 Image features, Scale space

3 Camera model, homographies, camera calibration, shading

4 Stereo and epipolar geometry & Structure from motion

5 Edges and lines

6 Image segmentation

Contents of the course

7 Intro to Machine Learning (Classification and Clustering models)

8 Dimensionality Reduction

9 Multi-view Learning and Neural Networks

10 Object detection and recognition

11 Scene classification (Multiple Instance Learning)

12 Visual Tracking

13 Salient Object segmentation

14 Human Action Recognition

Fast forward

1 Introduction and Intro to Image Processing

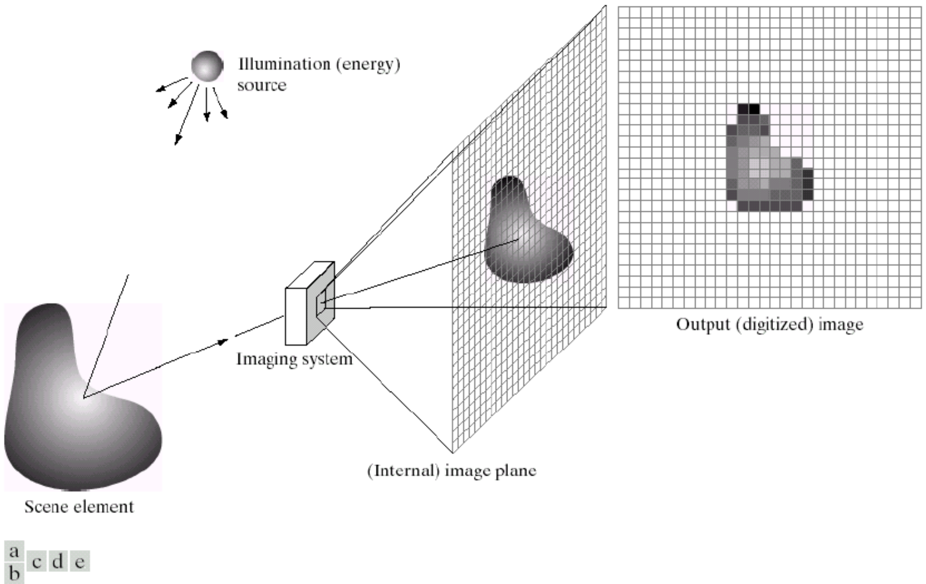
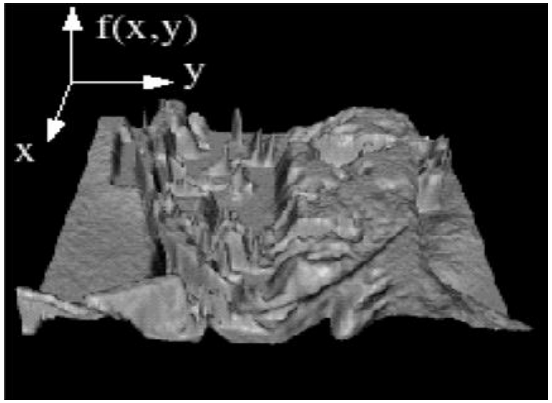
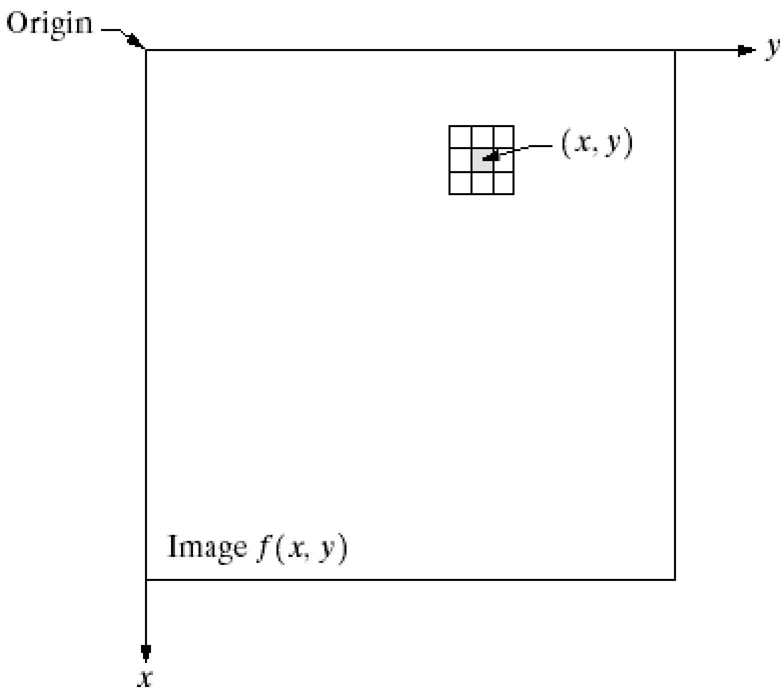


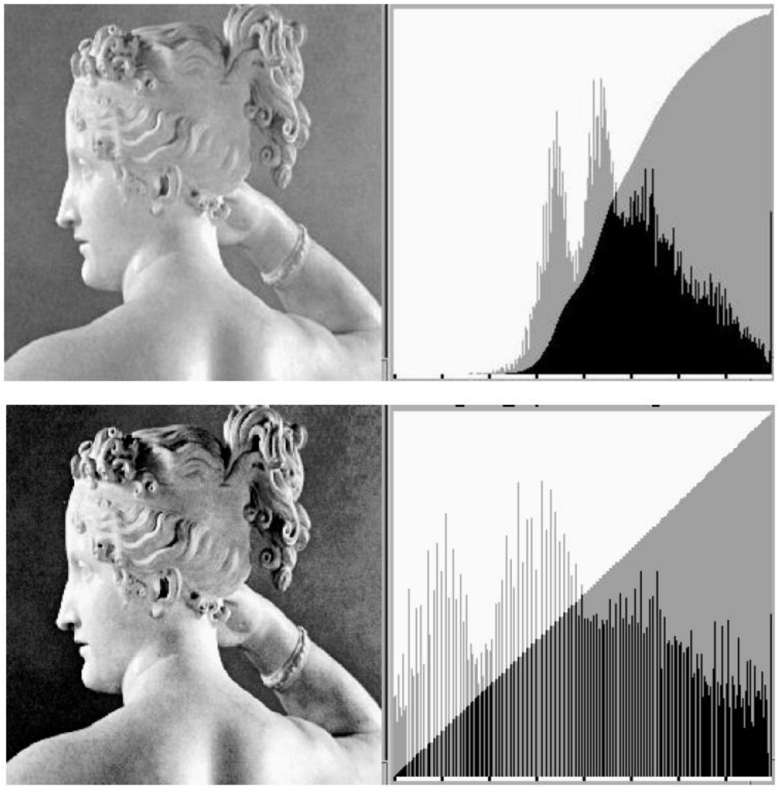
FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.



Fast forward

1 Introduction and Intro to Image Processing



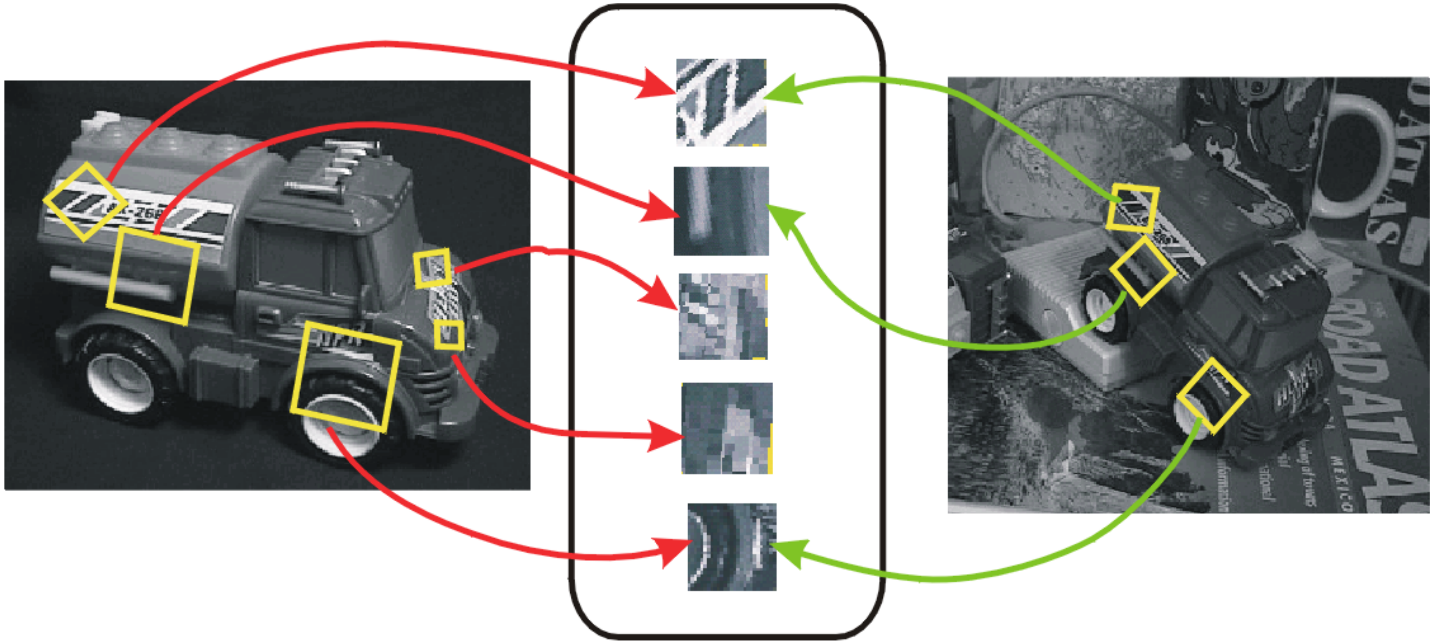


Contents of the course

2 Image features, Scale space

Contents of the course

2 Image features, Scale space



Contents of the course

2 Image features, Scale space

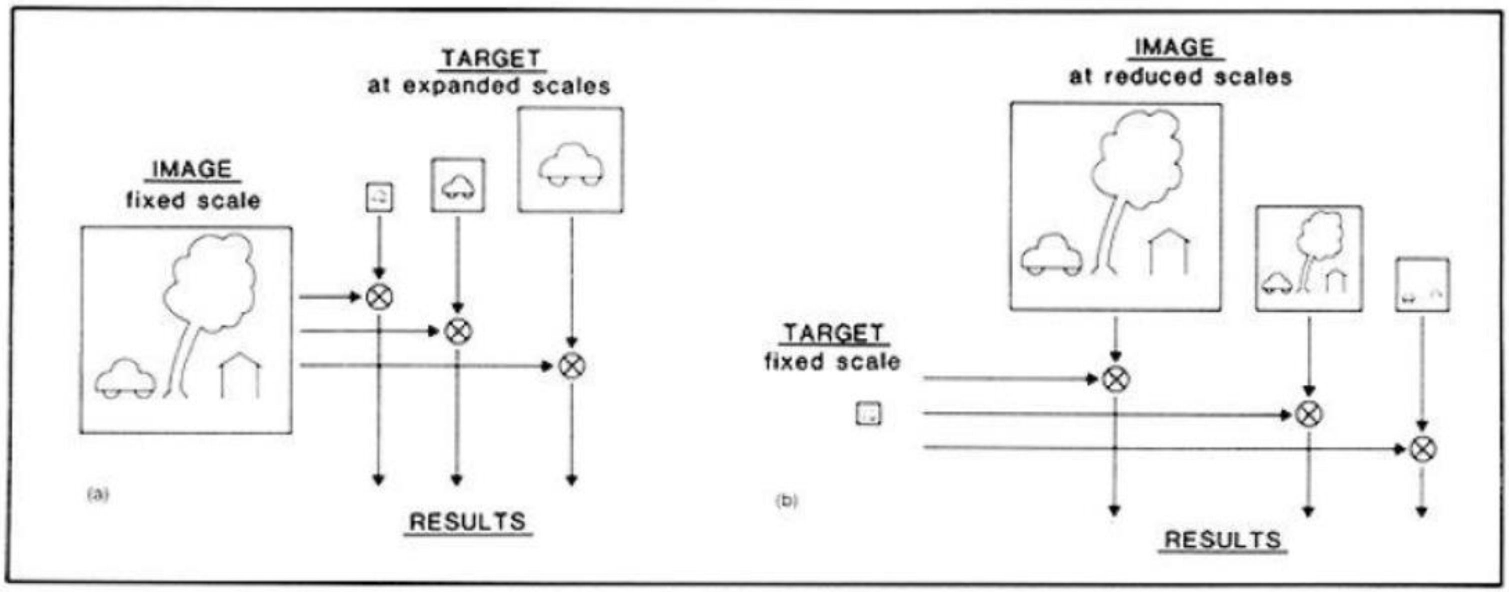
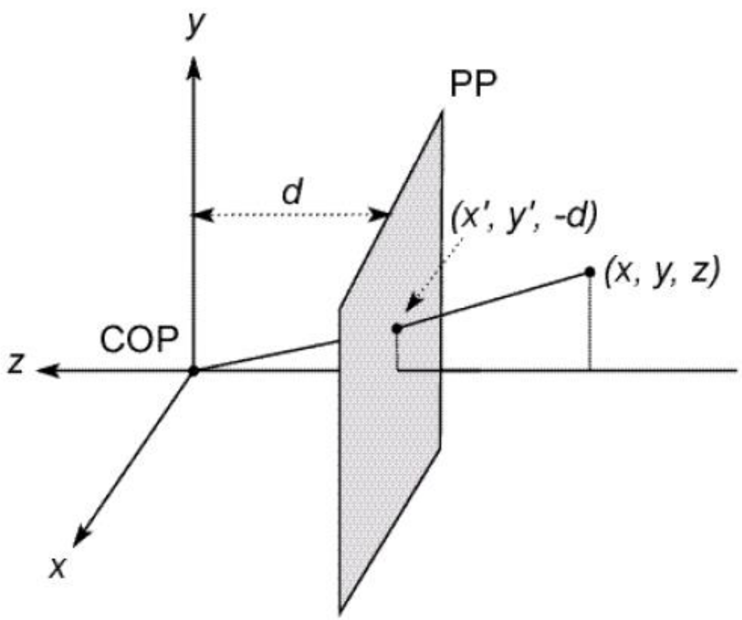
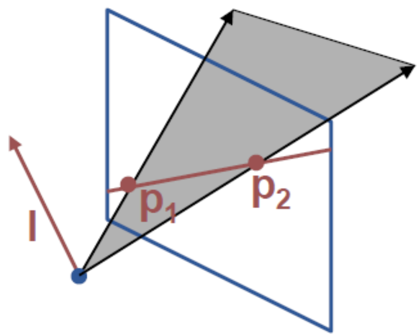


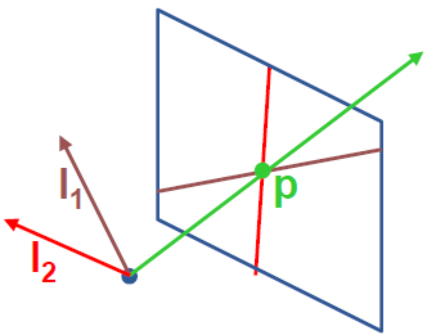
Fig. 1. Two methods of searching for a target pattern over many scales. In the first approach, (a), copies of the target pattern are constructed at several expanded scales, and each is convolved with the original image. In the second approach, (b), a single copy of the target is convolved with copies of the image reduced in scale. The target should be just large enough to resolve critical details The two approaches should give equivalent results, but the second is more efficient by the fourth power of the scale factor (image convolutions are represented by 'O').

Contents of the course

3 Camera model, homographies, camera calibration, shading

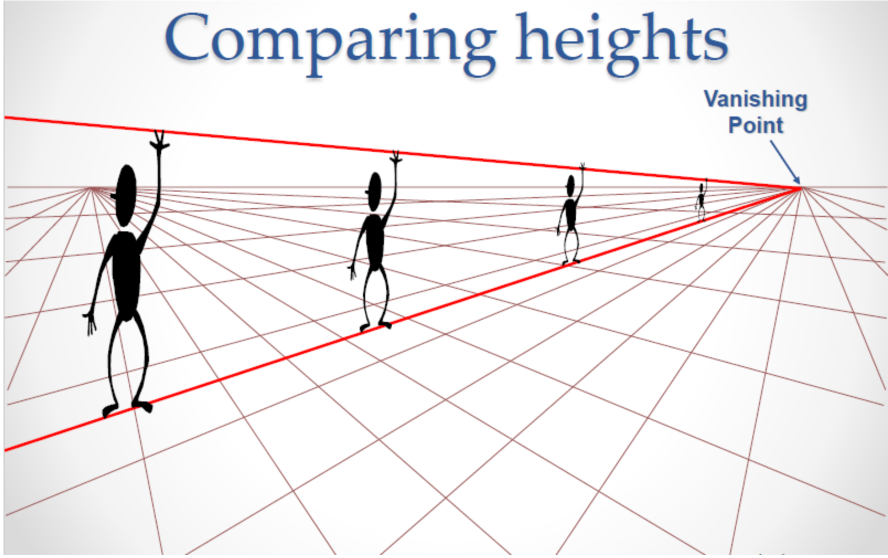






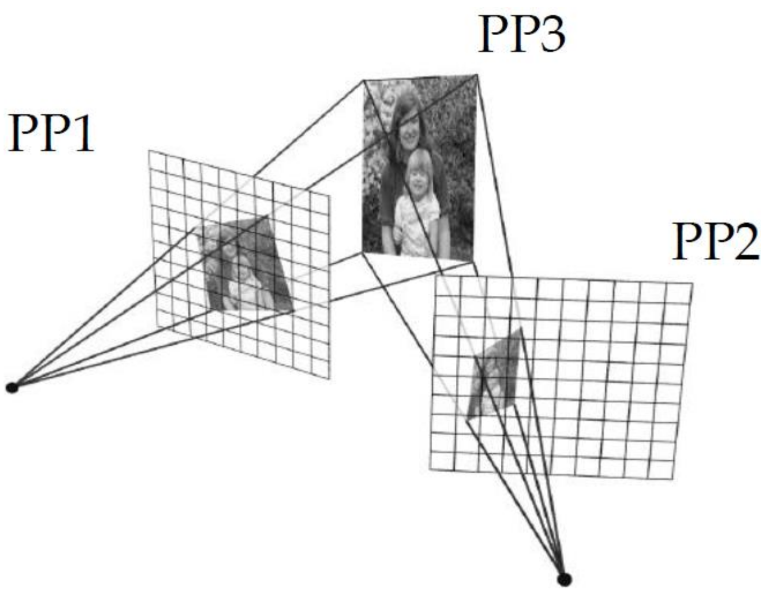
Contents of the course

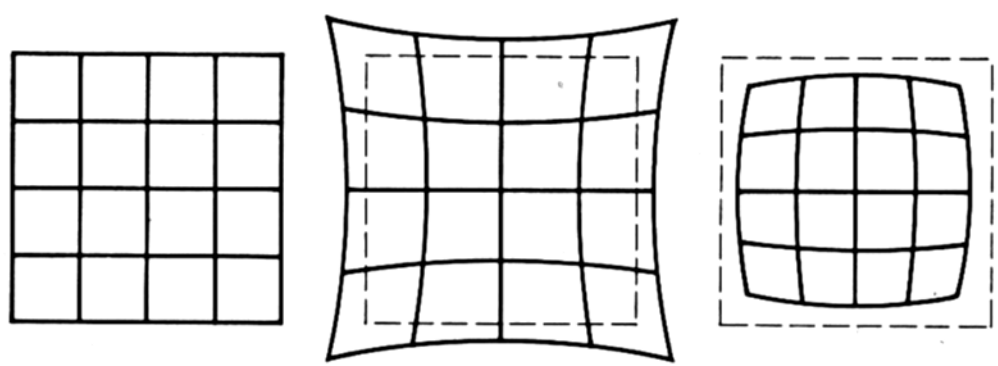
3 Camera model, homographies, camera calibration, shading



Contents of the course

3 Camera model, homographies, camera calibration, shading

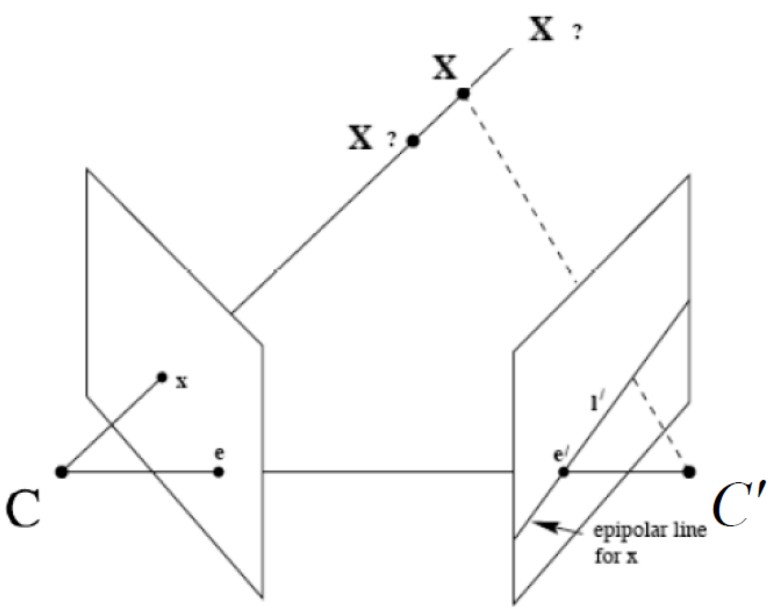


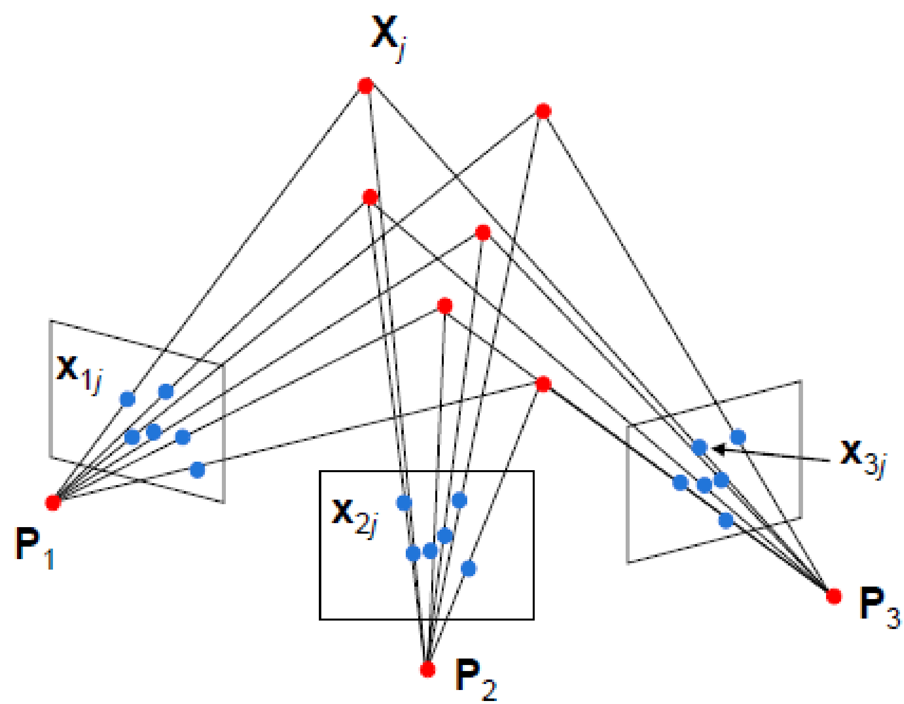


No distortion Pin cushion Barrel

Contents of the course

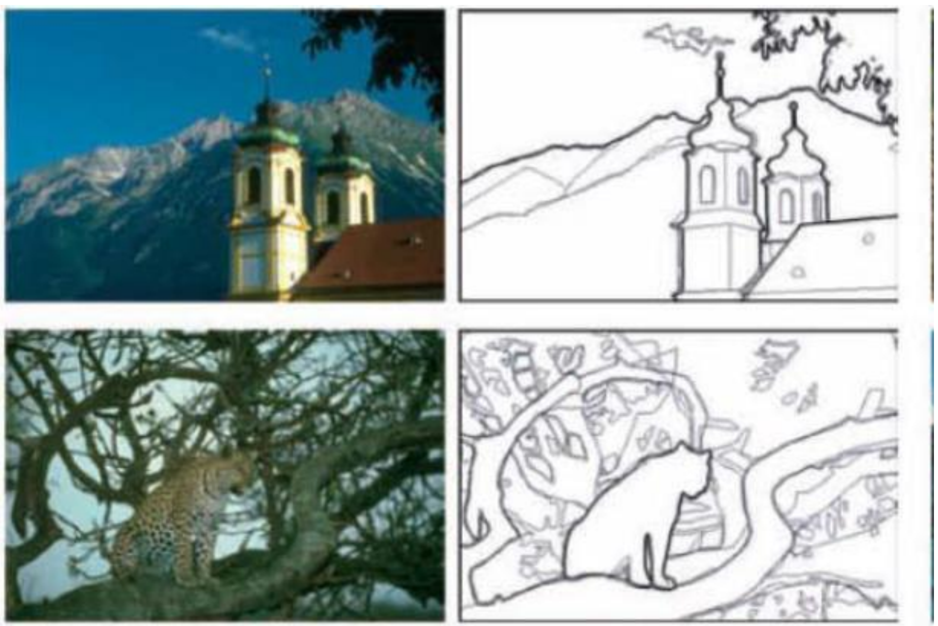
4 Stereo and epipolar geometry & Structure from motion





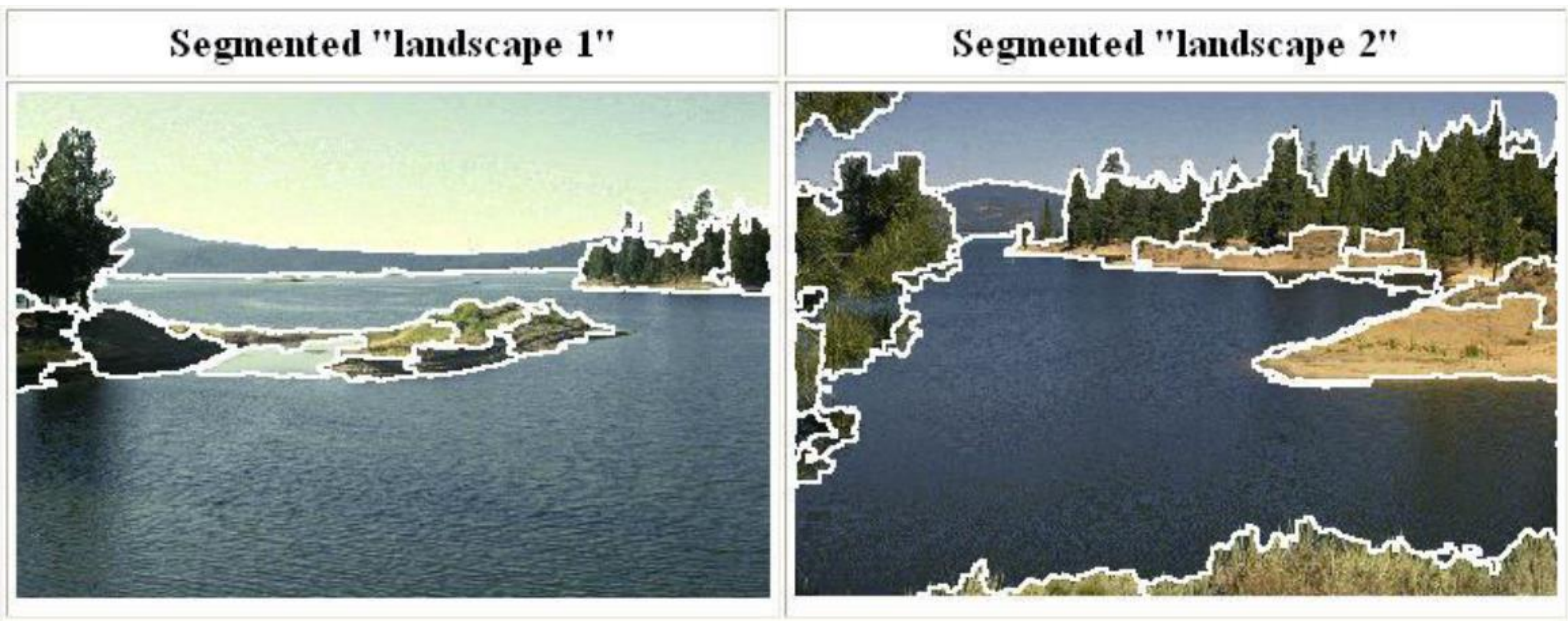
Contents of the course

5 Edges and lines



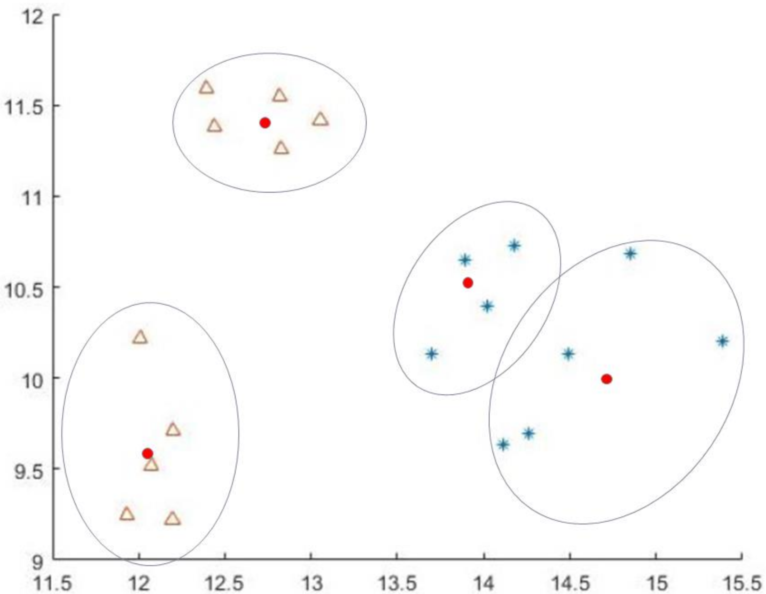
Contents of the course

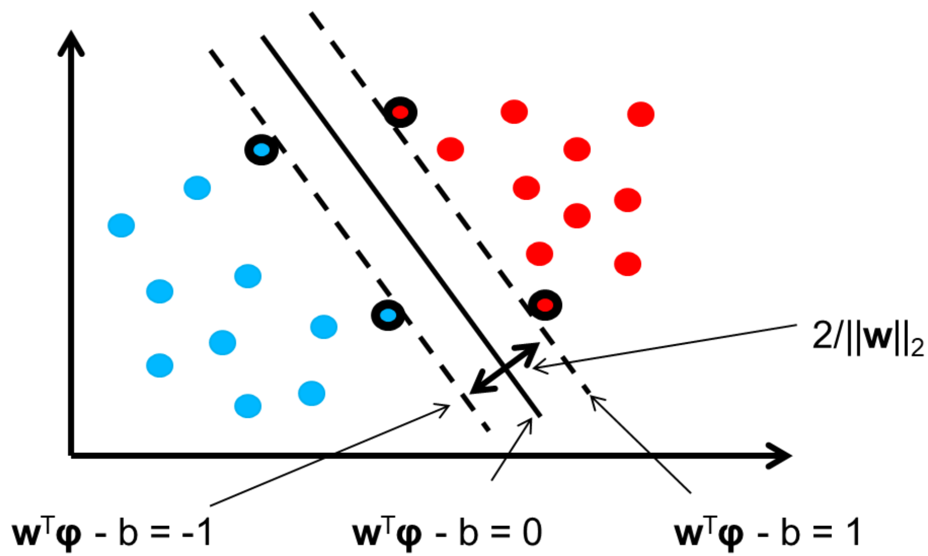
6 Image segmentation



Contents of the course

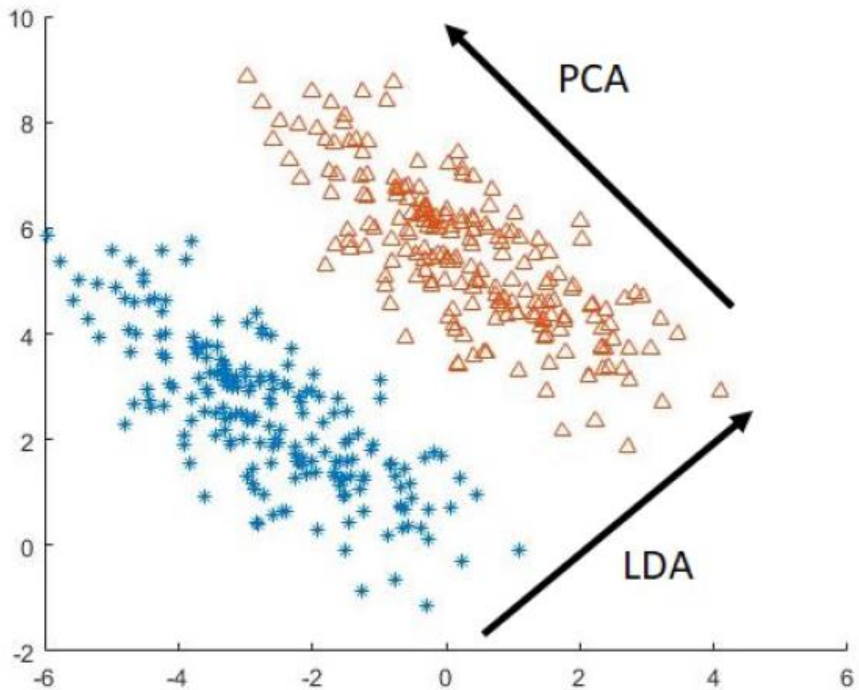
7 Intro to Machine Learning (Classification and Clustering models)





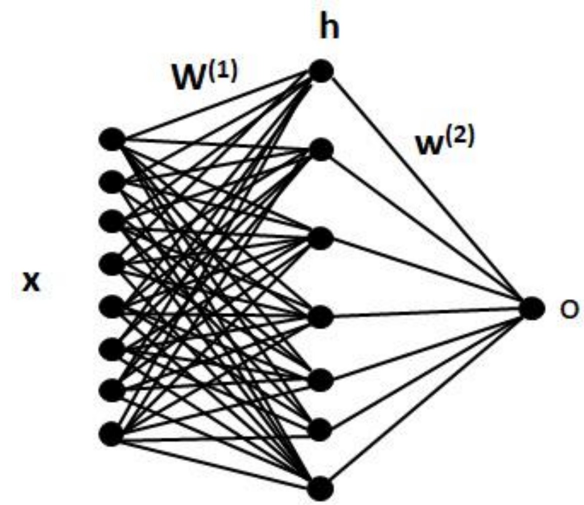
Contents of the course

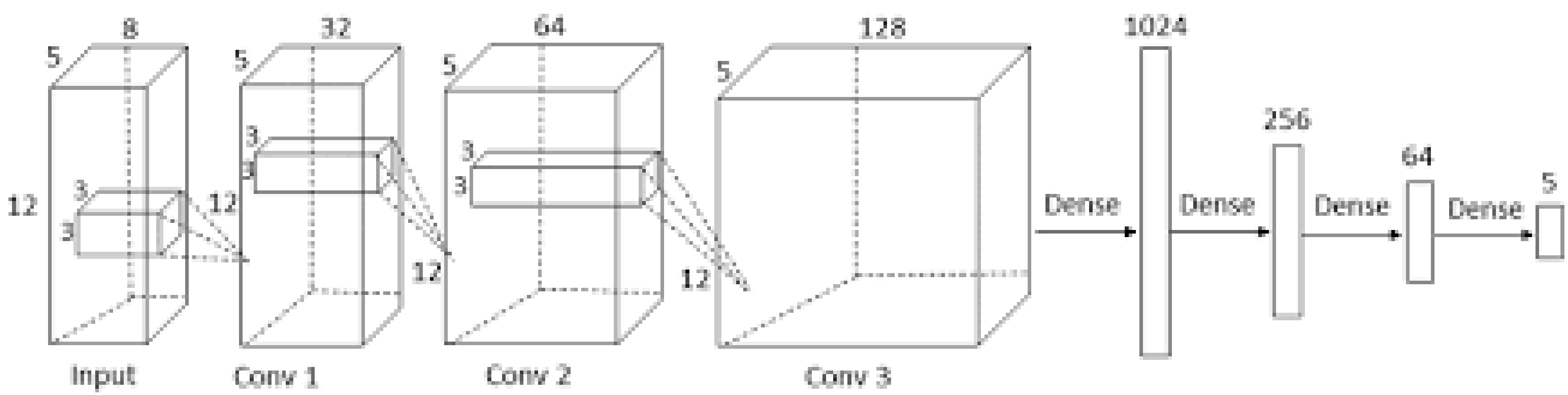
8 Dimensionality Reduction



Contents of the course

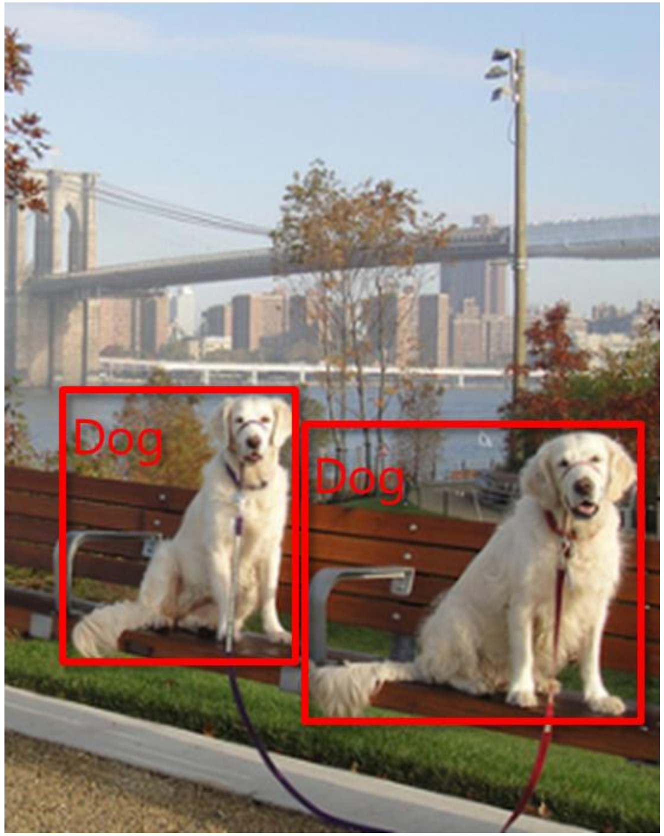
9 Multi-view Learning and Neural Networks





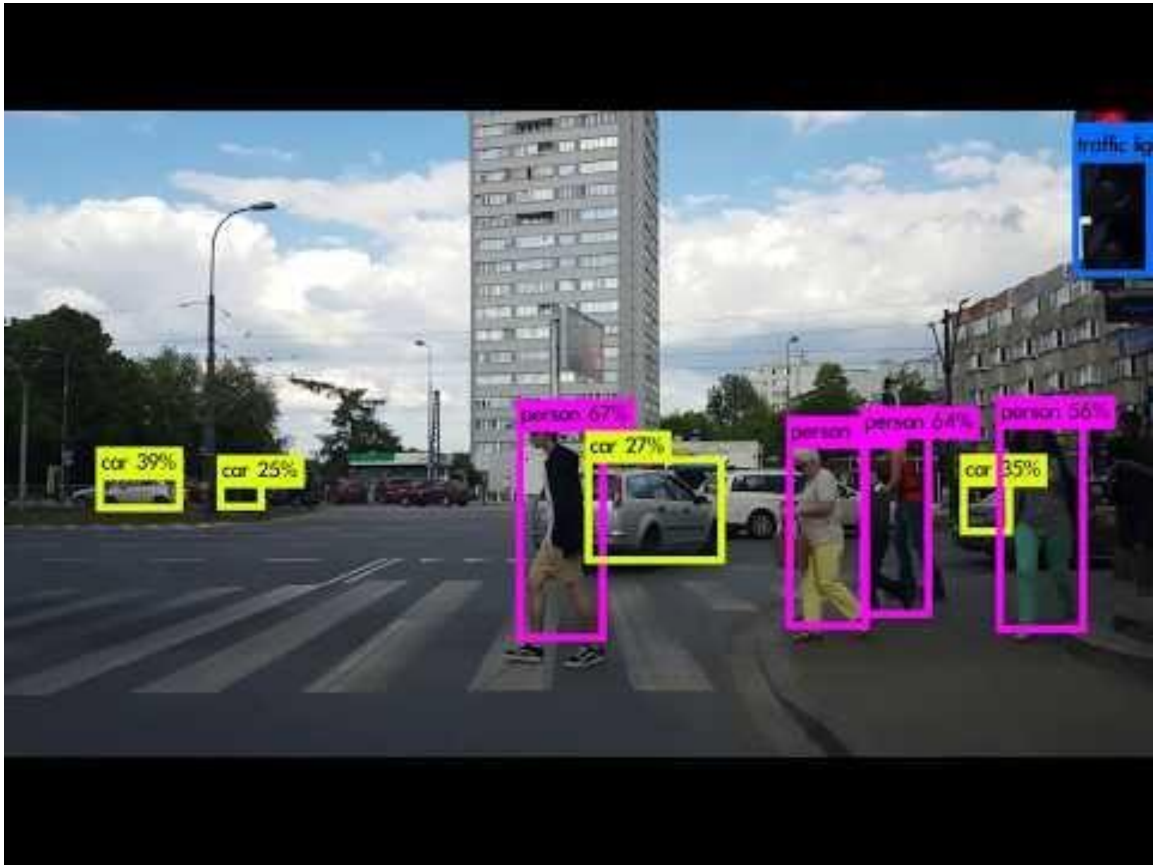
Contents of the course

10 Object detection and recognition



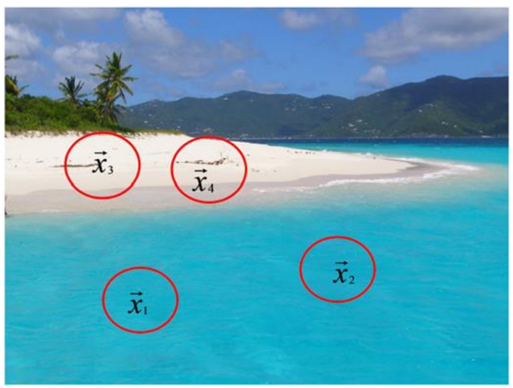
Contents of the course

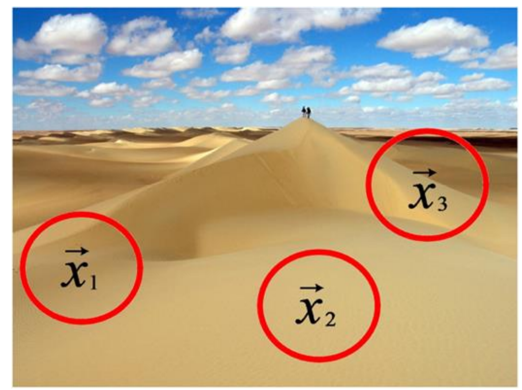
10 Object detection and recognition



Contents of the course

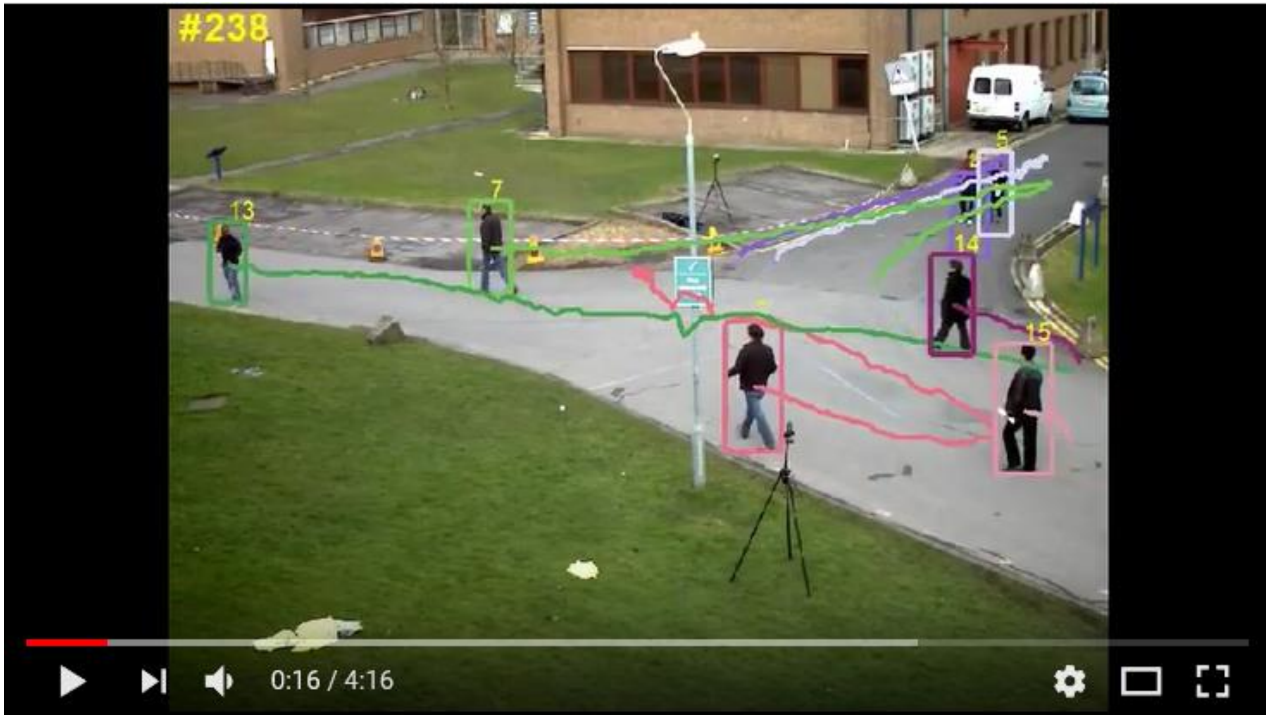
11 Scene classification (Multiple Instance Learning)





Contents of the course

12 Visual Tracking



Contents of the course

13 Salient Object segmentation

Contents of the course

14 Human Action Recognition