INFO - H - 501

Pattern recognition and image analysis

2 - recognition

Object recognition

- How do we recognize objects ?

Object recognition

- How do we recognize objects ?
  - object defined by sample
  - by usage
  - by definition
  - by context
  - ...

Object recognition

- direct approach
  - e.g. pattern matching
  - face recognition using PCA
- feature based approach
  - supervised classification: recall
  - corner-based: bag-of-visual word
  - edge histogram
  - Viola & Jones

Pattern matching

- find a pattern *h(i,j)* in the *f(i,j)* image
- matching criteria

$$C_1(u,v) = \frac{1}{\max_{(i,j)\in V} |f(i+u, j+v) - h(i,j)|}$$

$$C_2(u,v) = \frac{1}{\sum_{(i,j)\in V} |f(i+u, j+v) - h(i,j)|}$$

$$C_3(u,v) = \frac{1}{\sum_{(i,j)\in V} [f(i+u, j+v) - h(i,j)]^2}$$

Pattern matching

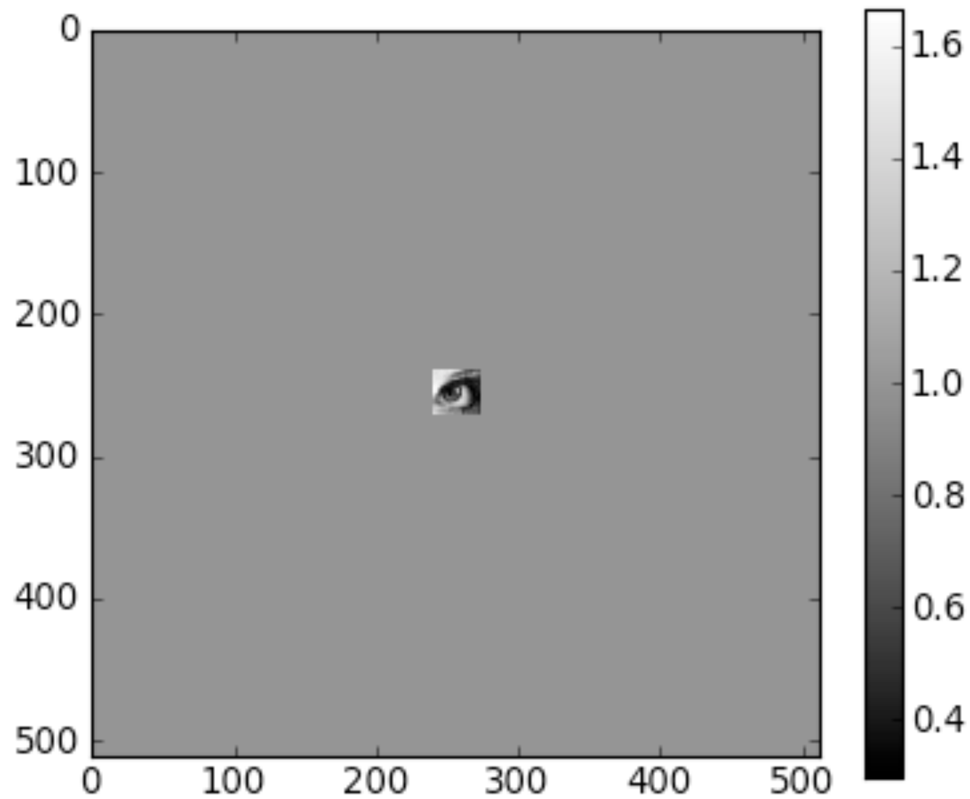
- correlation

$$f(x) \circ g(x) = \int_{-\infty}^{+\infty} f^*(\alpha) g(x+\alpha)\, d\alpha$$

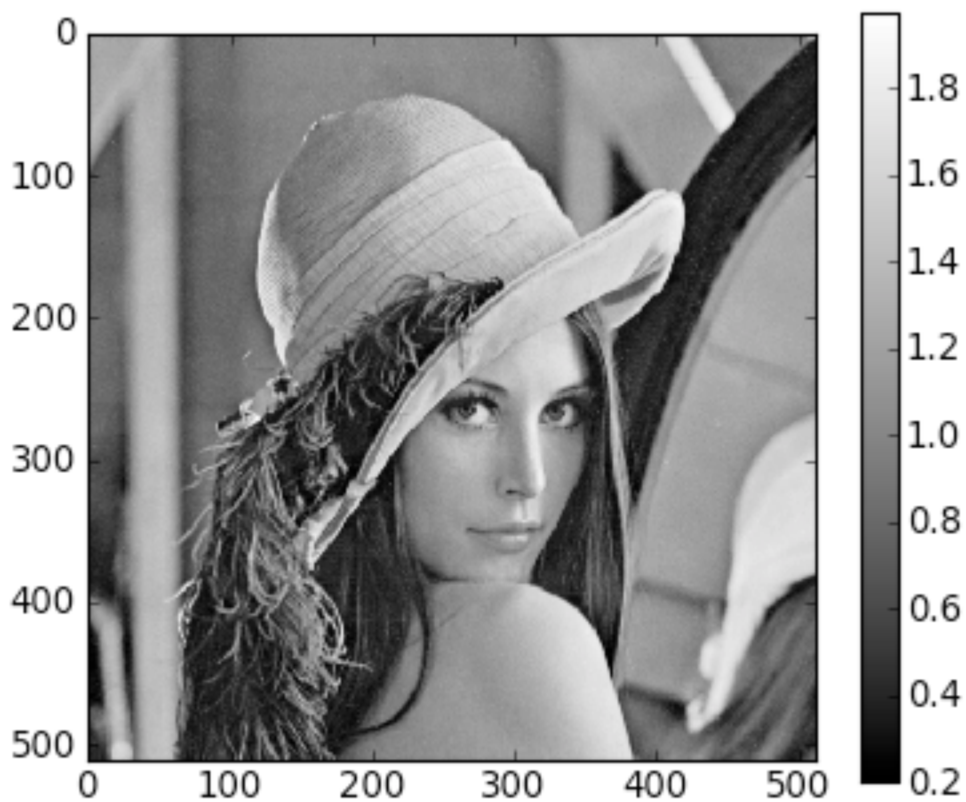
$$f(x) \circ g(x) = \mathcal{F}^{-1}(F^*(u,v)G(u,v))$$

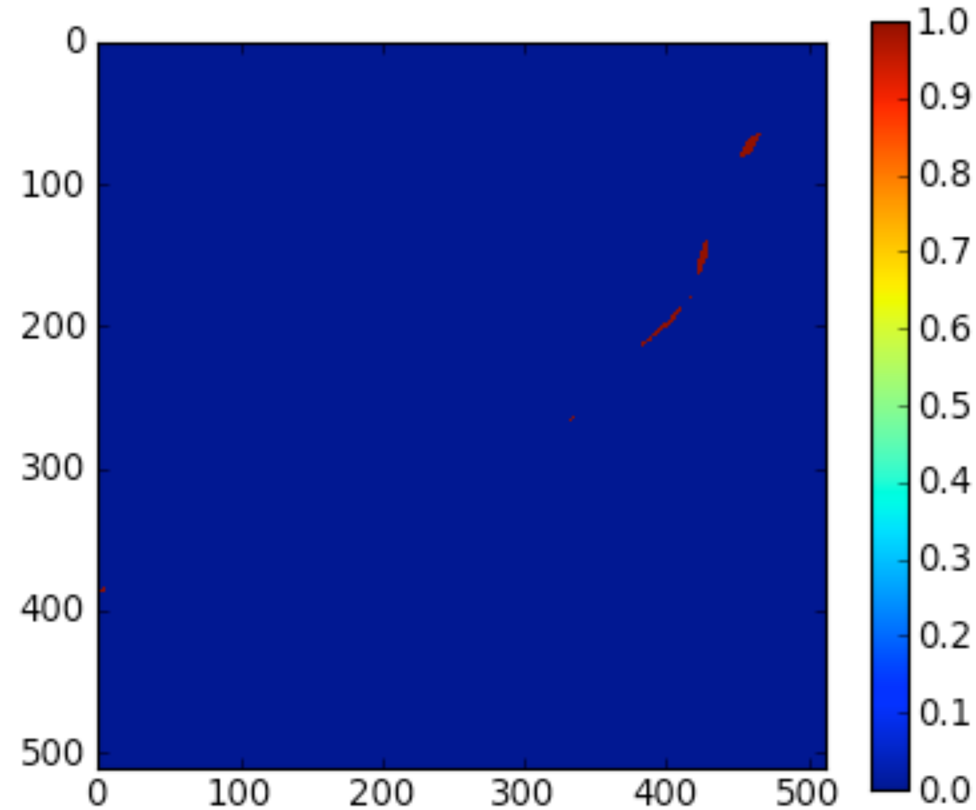
Pattern matching

Pattern matching

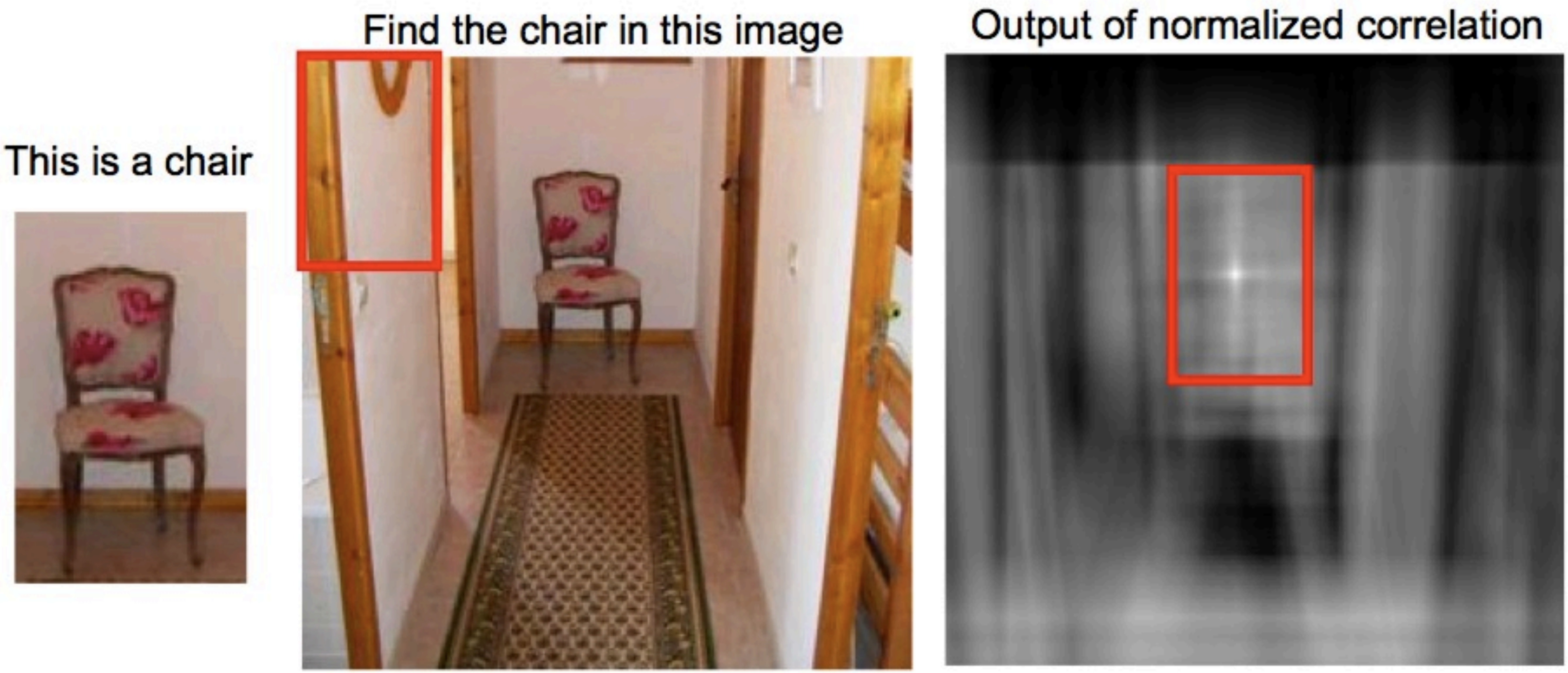


Pattern matching





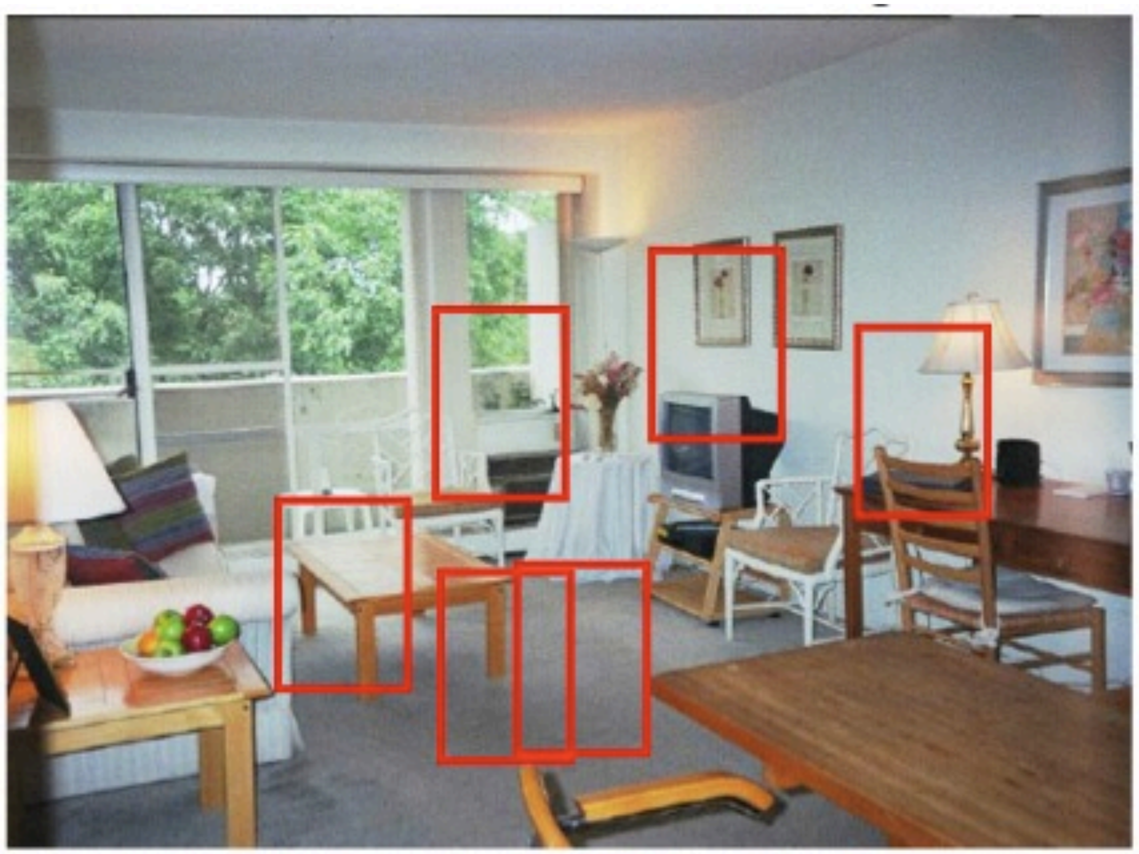
Pattern matching

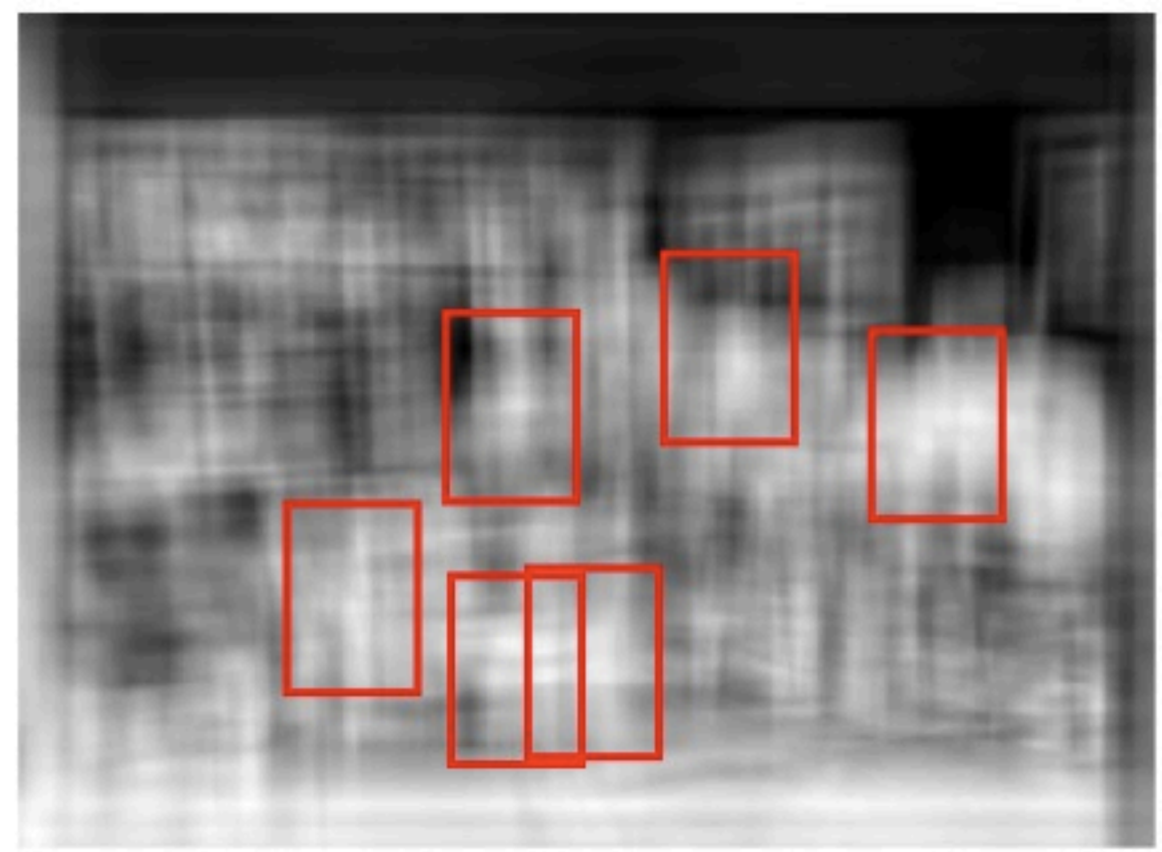


http://people.csail.mit.edu/torralba/shortCourseRLOC/index.html

Pattern matching

- ?





Face recognition using PCA

- Eigen Faces
- How to recognise a face from an image database ?

Eigen faces

- 1 pixel = 1 feature
- e.g. 64x64 = 4096
- gray level
- centered face
- same resolution
- averaged

Eigen faces

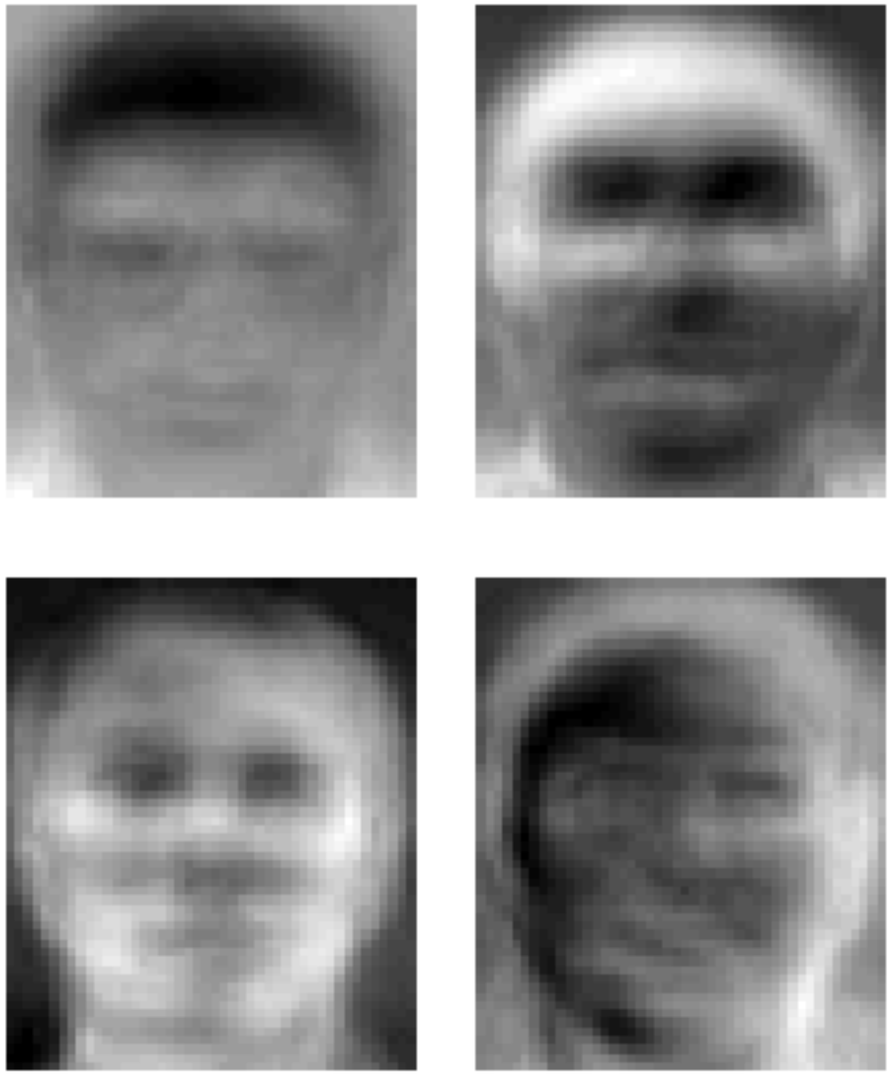
- 1 pixel = 1 feature
- e.g. 64x64 = 4096
- gray level
- centered face
- same resolution
- averaged

$$\mathbf{X} = [g_{00}\ g_{01}\ g_{02} \cdots g_{63\,63}]^T$$

- variable change
- PCA (principal Component Analysis)

Eigen faces

- after variable change, one axe is a principal face
- one face = linear combination of variables



wikipedia

PCA (Principal Component Analysis)

- eigen values of the covariance matrix
- eigen vectors sorted by increasing eigen valus
- N first eigen values
  - signature
  - compression
- recognition :
  - the new face is projected into new axes
  - euclidian distance

Eigen faces

- Advantages
  - fast
  - easy (naive)
- Limitations
  - sensitive to exposition
  - variable to view pose
  - other approaches more robust
  - limitation due to the number of dimensions

Eigen faces

- How to compute eigen values for large dimensions ?
- example 64x64 = 4096 dim.
- covariance matrix = 4096x4096 !

$$\mathbf{X} = \left[g_{00}\ g_{01}\ g_{02} \cdots g_{63\,63}\right]^T$$

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]$$

$$\mu_i = \mathrm{E}(X_i)$$

Eigen faces

- matrix rank = number of images
- if N examples, N-1 eigen values <> 0

$$\mathbf{S}\mathbf{v}_i = \mathbf{T}^T\mathbf{T}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

$$\mathbf{T}\mathbf{T}^T\mathbf{u}_i = \lambda_i\mathbf{u}_i$$

$$\mathbf{T}^T\mathbf{T}\mathbf{T}^T\mathbf{u}_i = \lambda_i\mathbf{T}^T\mathbf{u}_i$$

- then $\mathbf{v}_i = \mathbf{T}^T\mathbf{u}_i$ is eigen vector of S

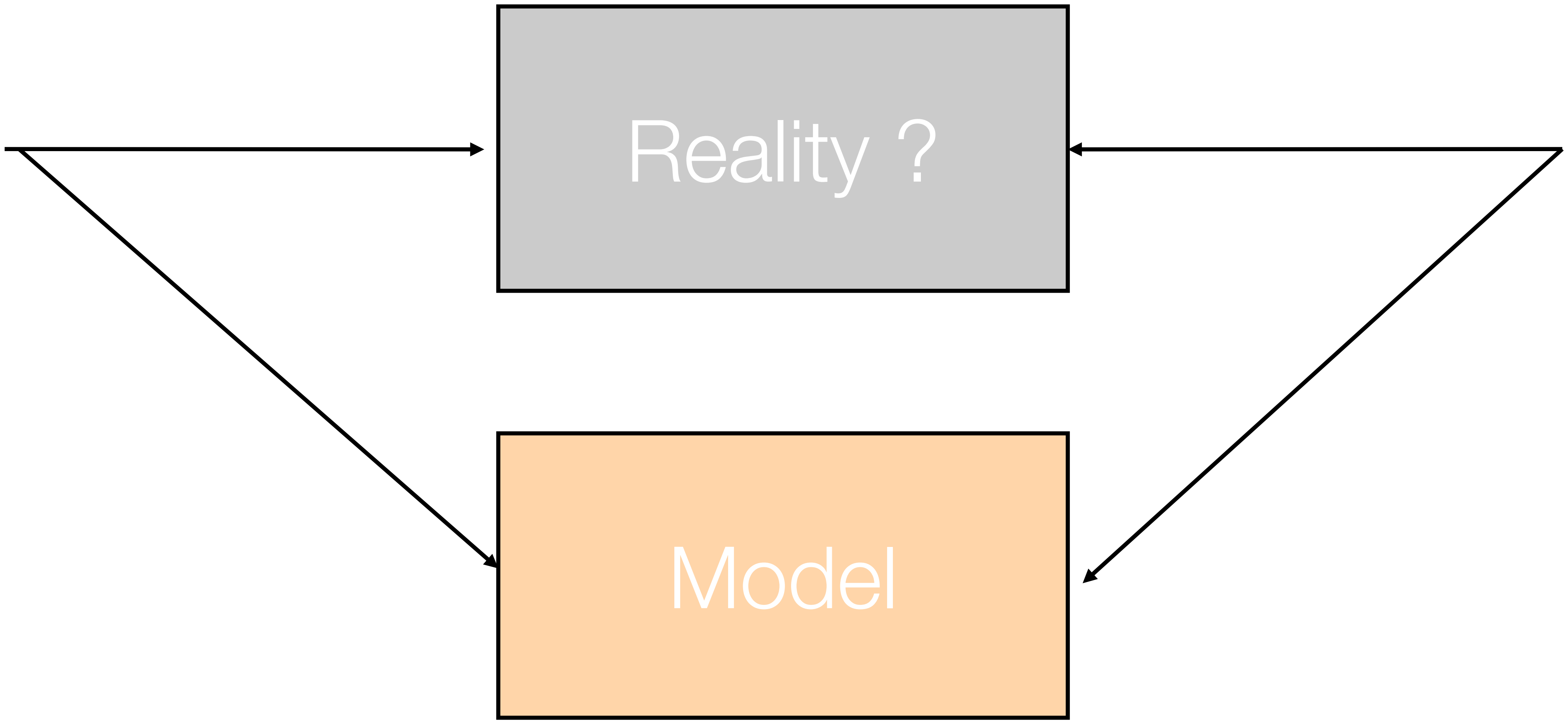
Supervised classification

- recall

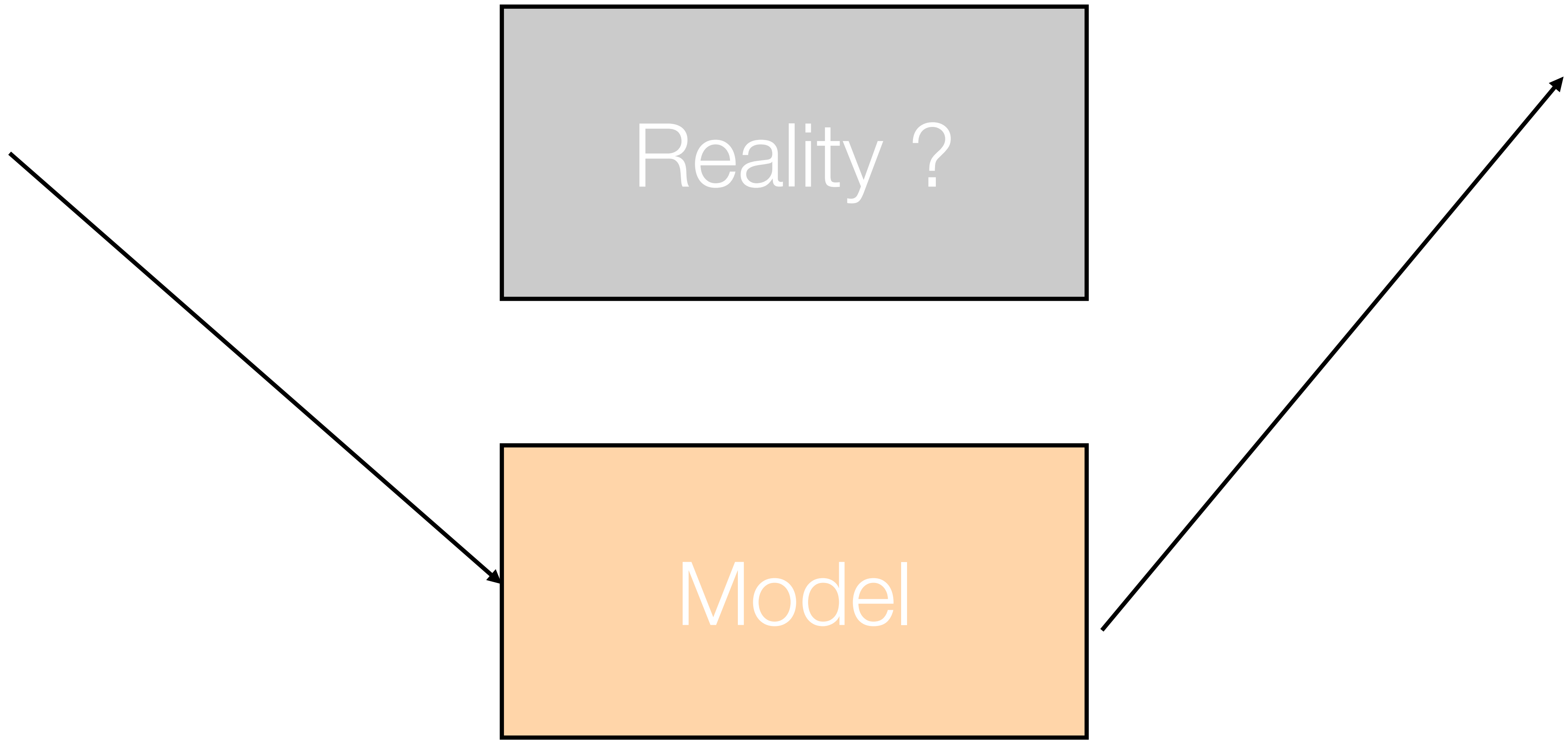
Machine learning

- Unknown underlying real world
- Model
- Data
  - input
  - output
- aims:
  - Inference (value)
  - Prediction (class)

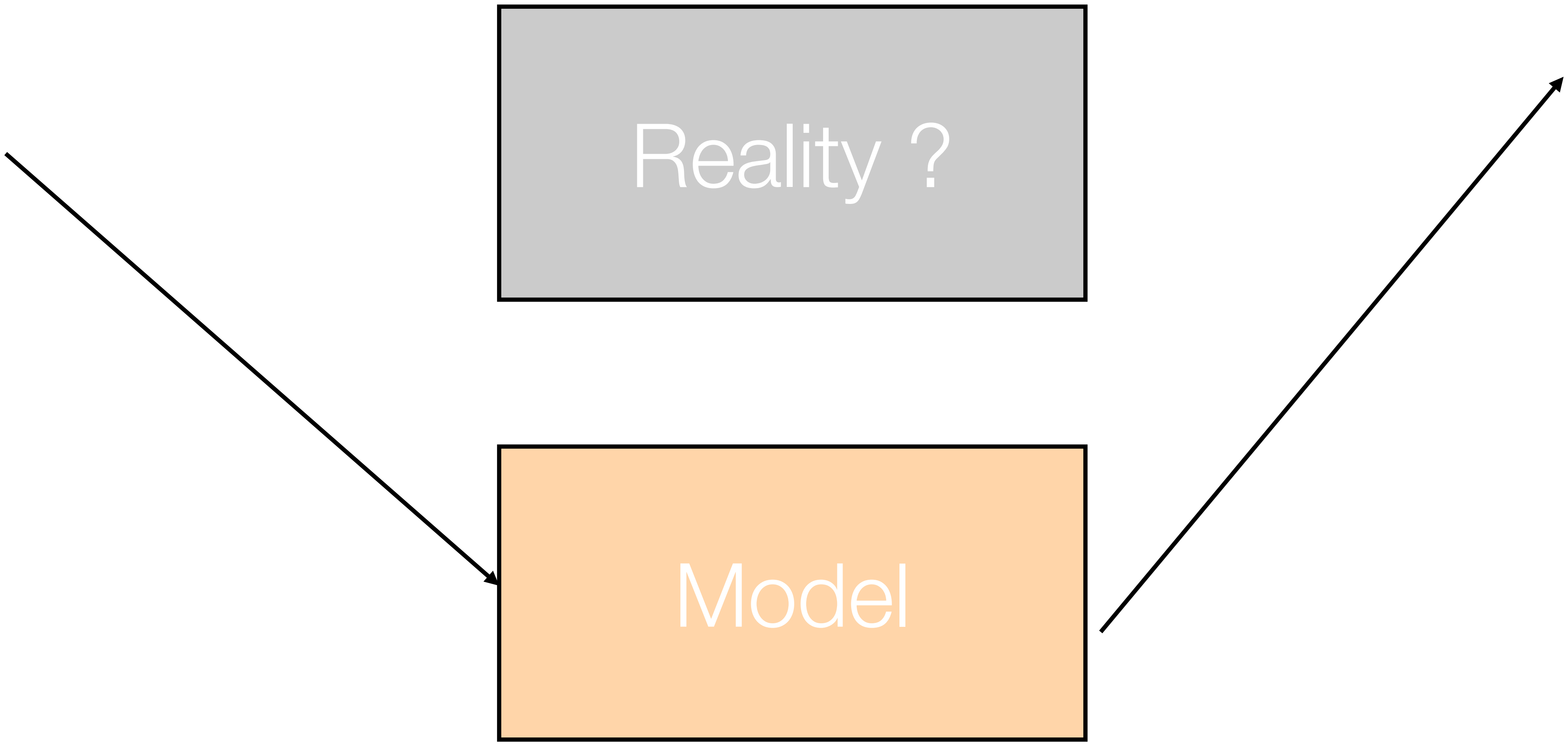
Modelisation



Inférence

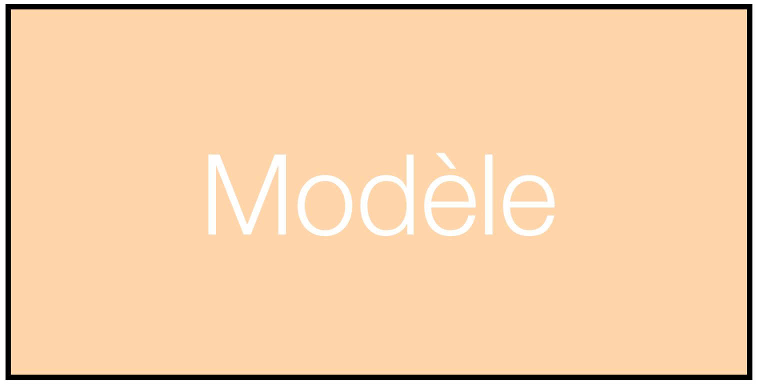


Classification



Model

- Variable complexity
- Linear / non linear
- output = value
- output = class
- ! stationarity
- Various methods, Bayes, neural network, decision tree, SVM,…



Methods

- number of available data
- number of features
- type of features
- feature space
- distance definition
- discriminent function

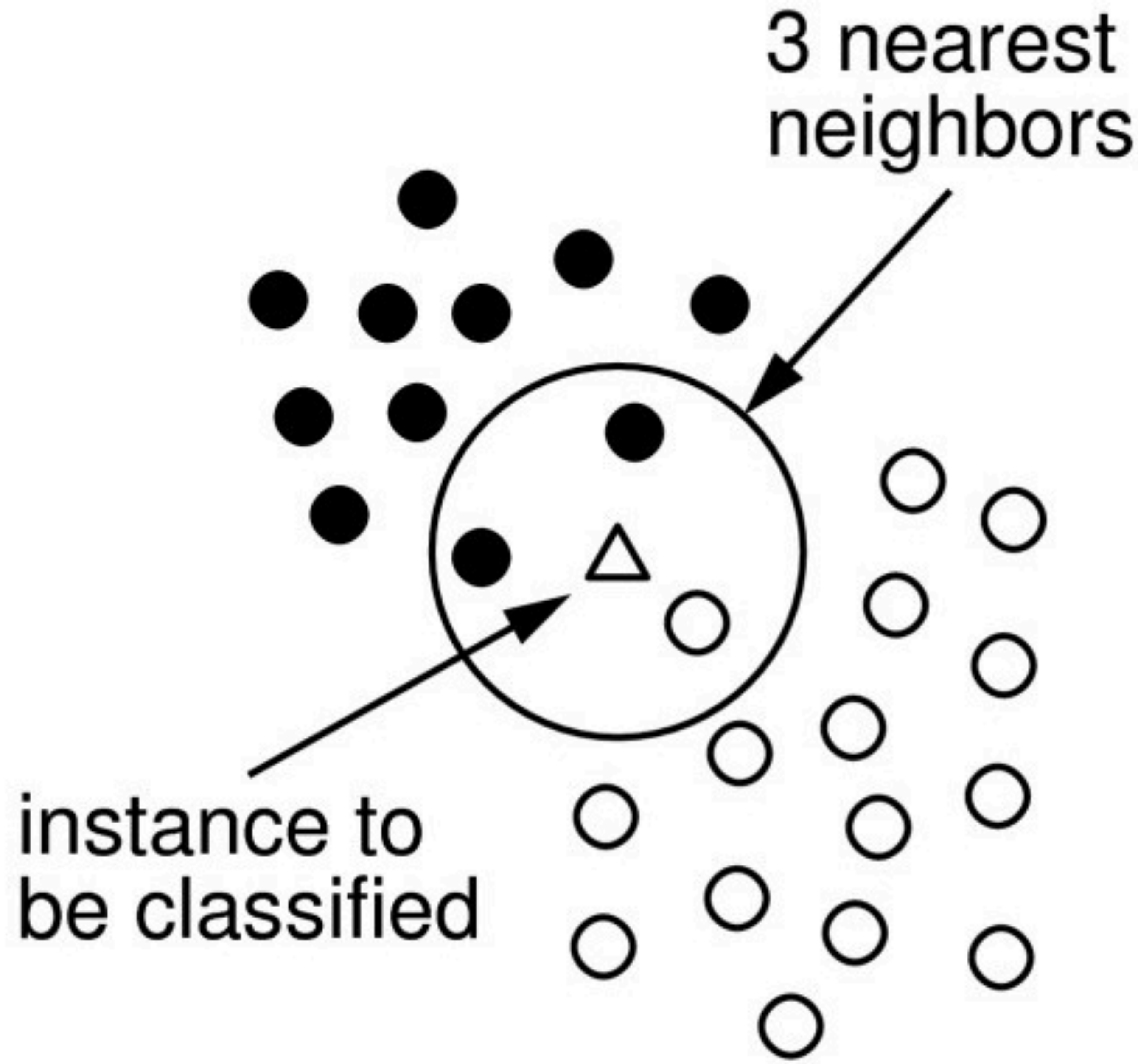
Supervised classification

- Matching
- nearest neighbor
- Bayes
- discriminant analysis
- neural network
- decision tree
- ...

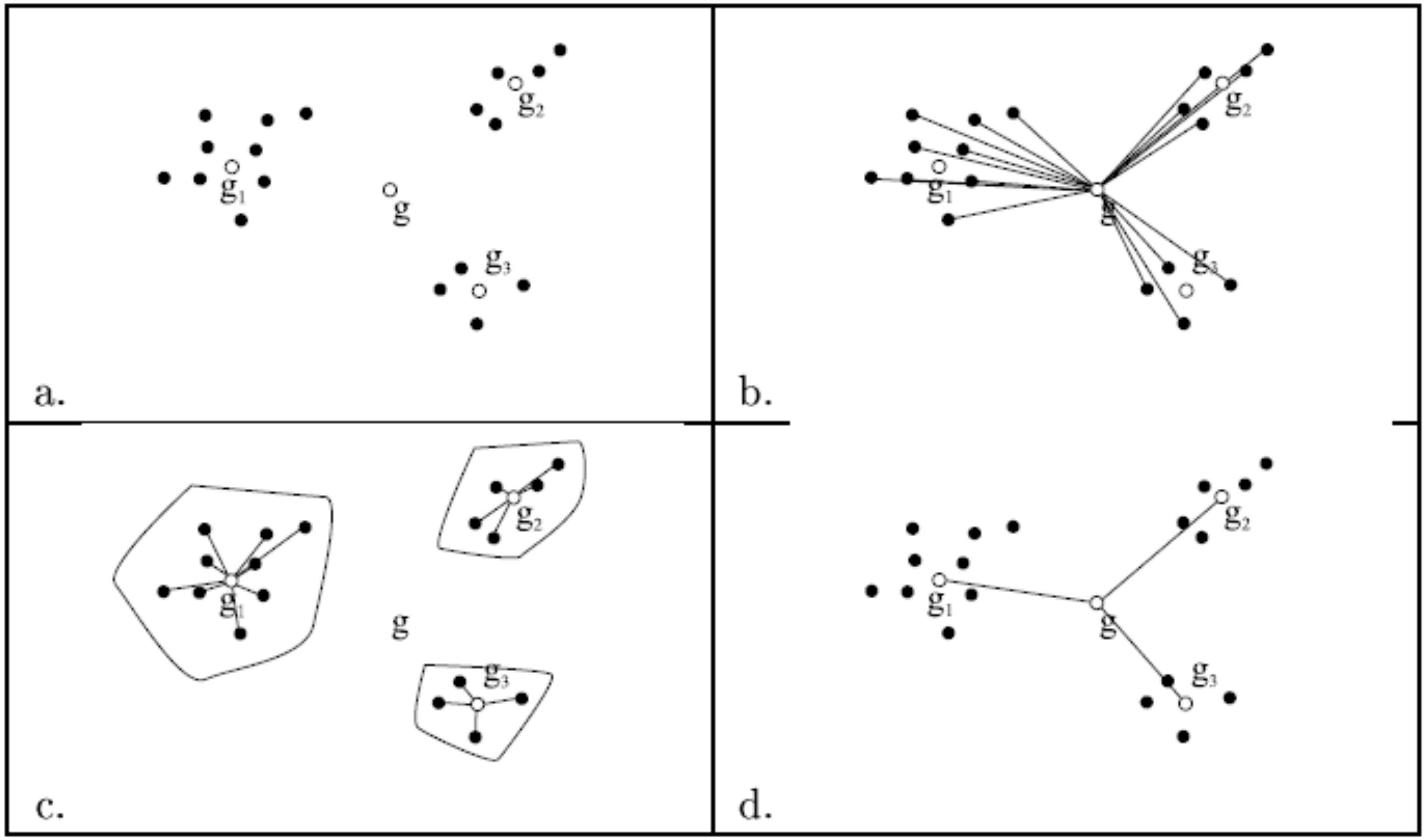
Nearest neighbor

- distance
- normalisation

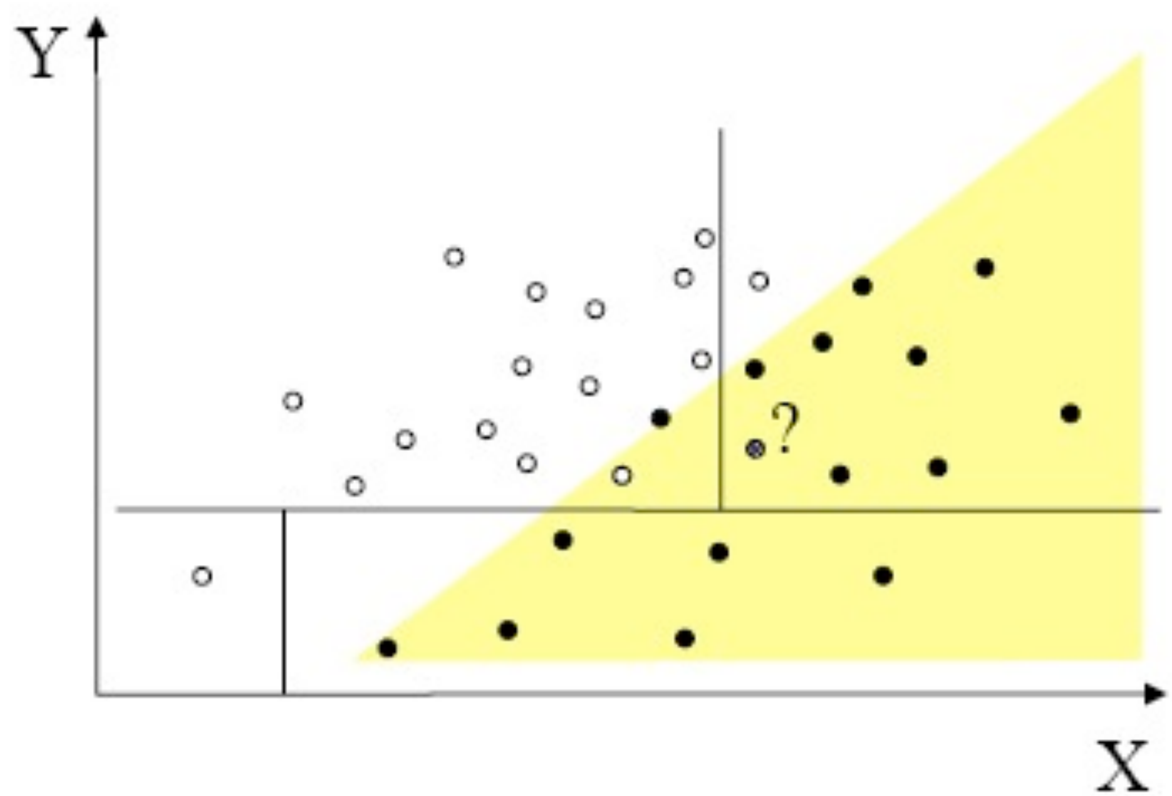
Nearest neighbor



Discriminent analysis



Decision tree



Decision tree

- entropy measure

$$entropy(T) = -\sum_{k=1}^{q} p_k \log(p_k)$$

$$\sum_{k=1}^{q} p_k = 1$$

- entropy

$$info(T) = -\sum_{k=1}^{q} \frac{\#(C_k,T)}{\#(T)} \log_2 \left(\frac{\#(C_k,T)}{\#(T)} \right)$$

- entropy gain

$$info_X(T) = \sum_{i=1}^{n} \frac{\#(T_i)}{\#(T)} \quad info(T_i)$$

$$gain(X) = info(T) - info_X(T)$$

Decision tree

- in order to limit small leaves: split info

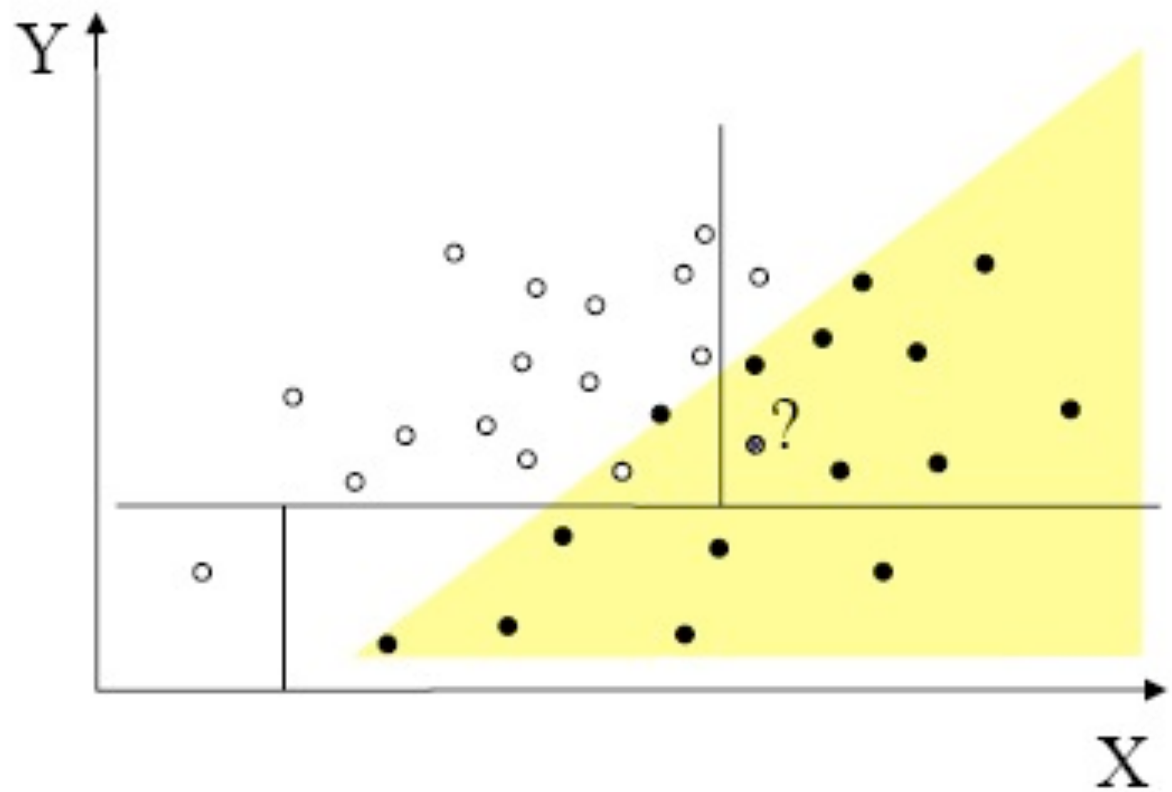
$$split\ info(X) = -\sum_{i=1}^{n} \frac{\#(T_i)}{\#(T)} \log_2 \left(\frac{\#(T_i)}{\#(T)} \right)$$

$$gain\ ratio(X) = \frac{gain(X)}{split\quad info(X)}$$

- process applied recursively until number of case in leave are all of the same class or # = 2

Decision tree

- Pruning

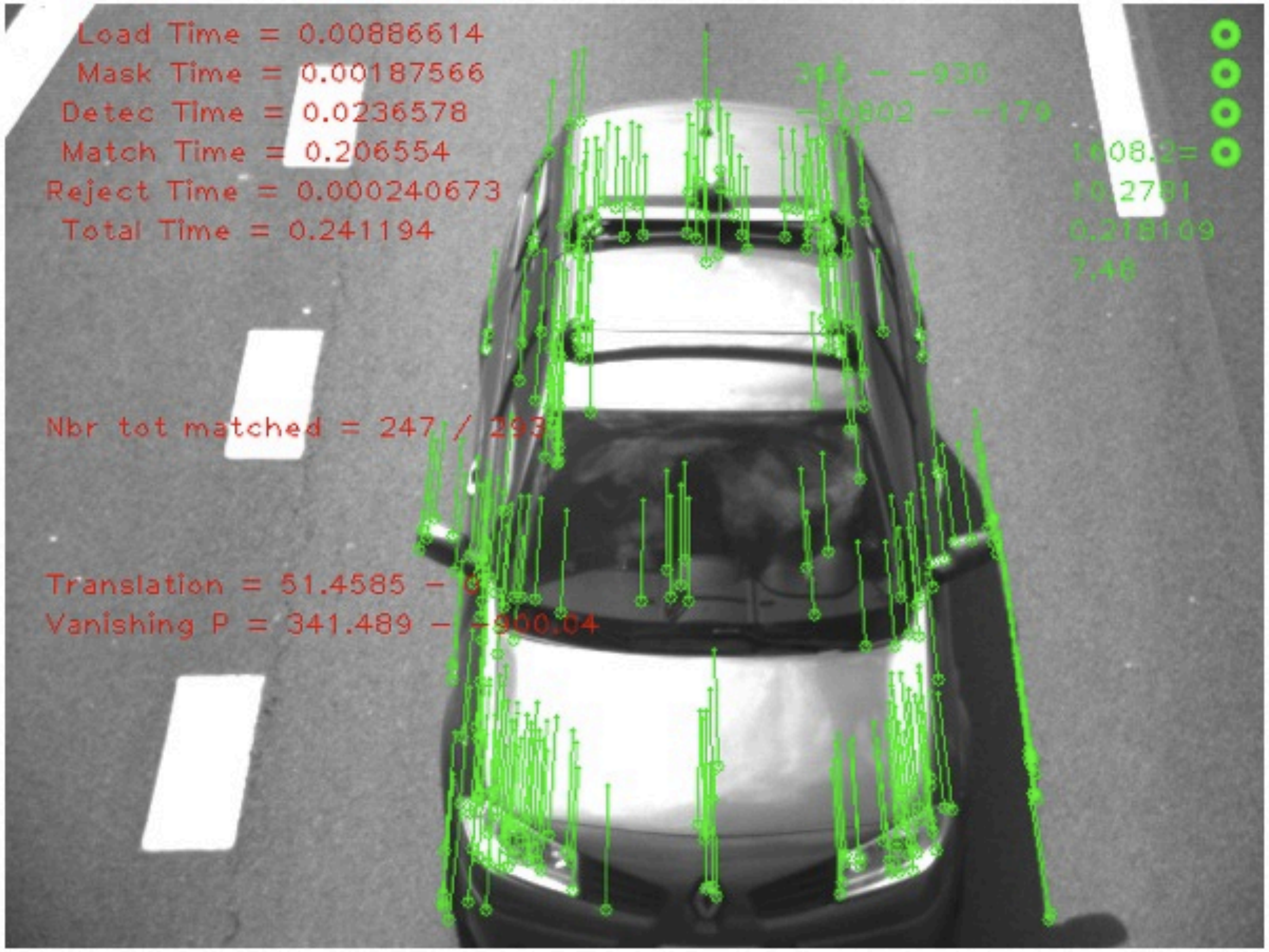


Object detection/recognition: Bag of visual word

- initial usage: bag of word
  - text recognition
- General idea:
  - image contains remarkable points
  - distribution of these points is a signature
  - machine learning algorithm allows recognition
  - no segmentation

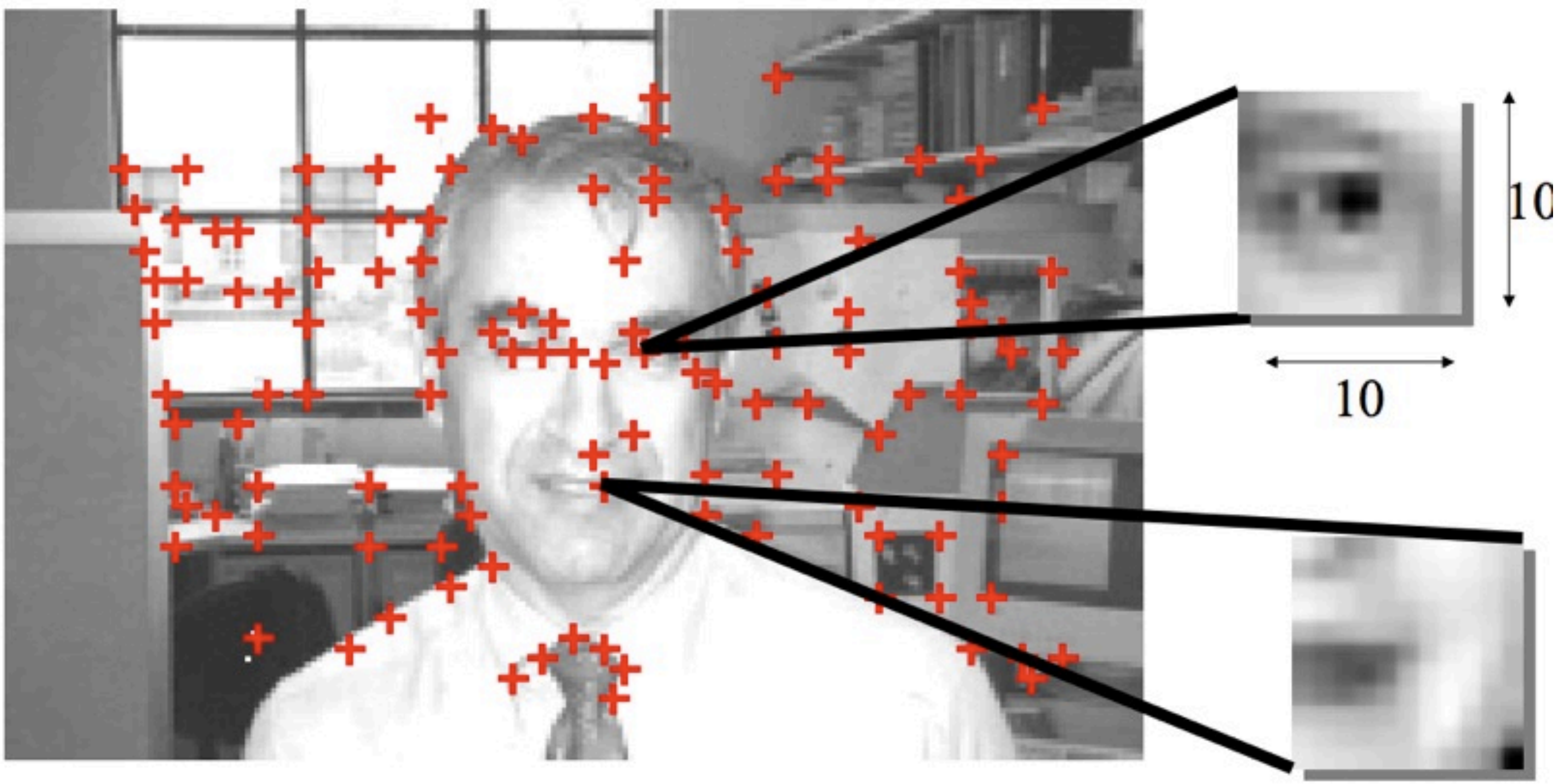
Bag of visual word

- find robust remarkable points
  - Harris corners
  - Sift, surf, fast
- normalized patches
- build vocabulary
  e.g. k-means



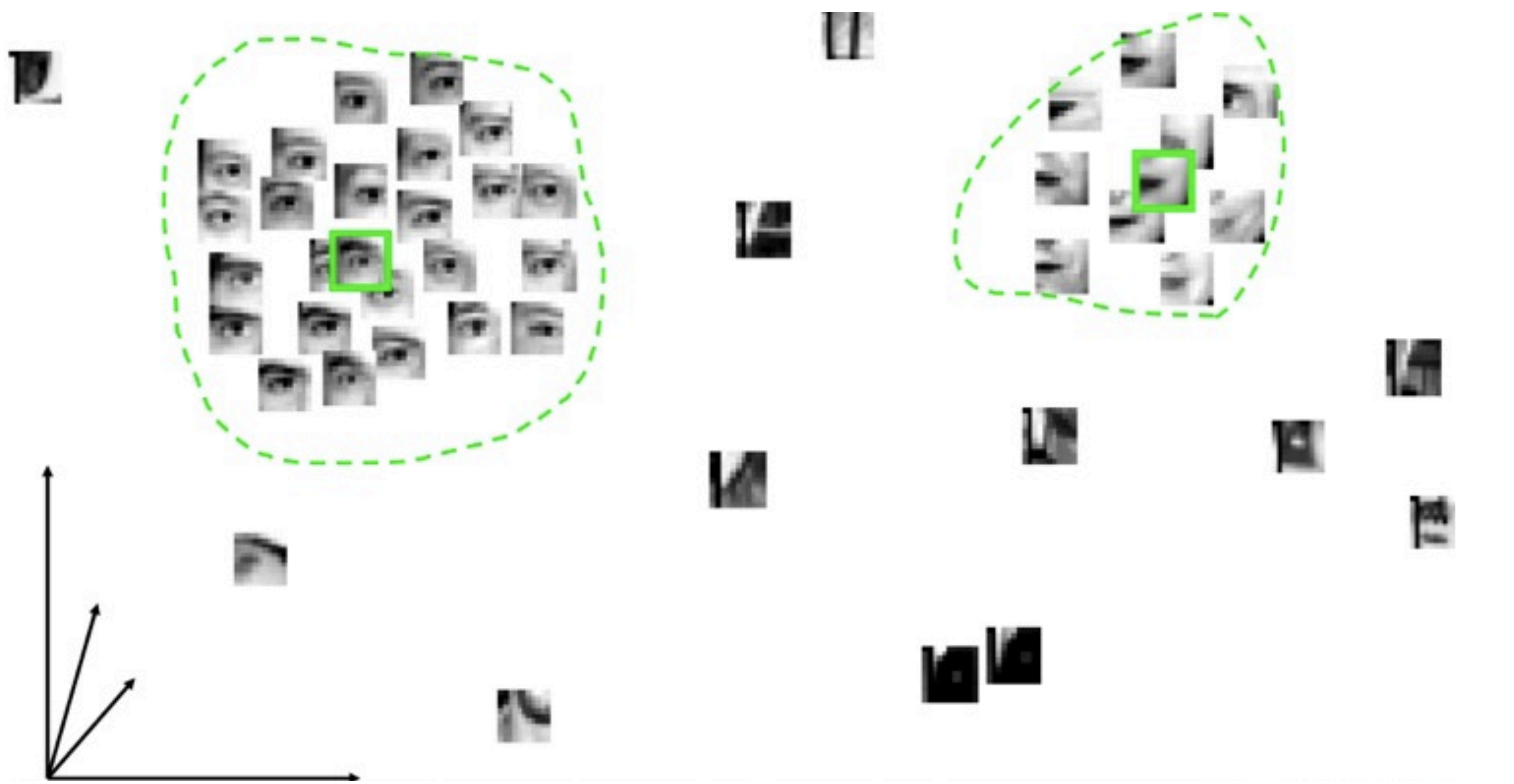
Ch. Kaes

Bag of visual word



www.robots.ox.ac.uk/~az/**icvss08_az_bow**.pdf

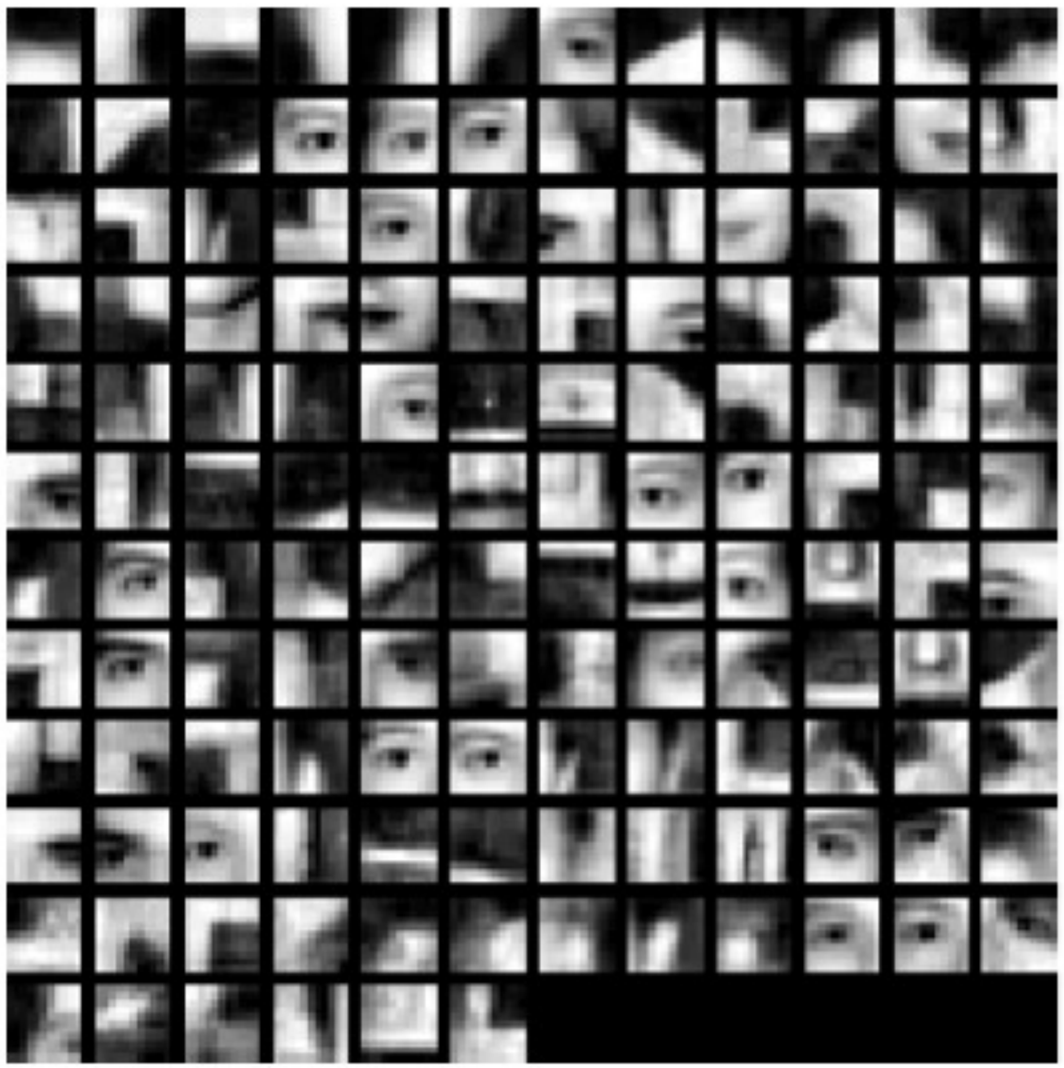
Bag of visual word



www.robots.ox.ac.uk/~az/**icvss08_az_bow**.pdf

Bag of visual word

100-1000 images



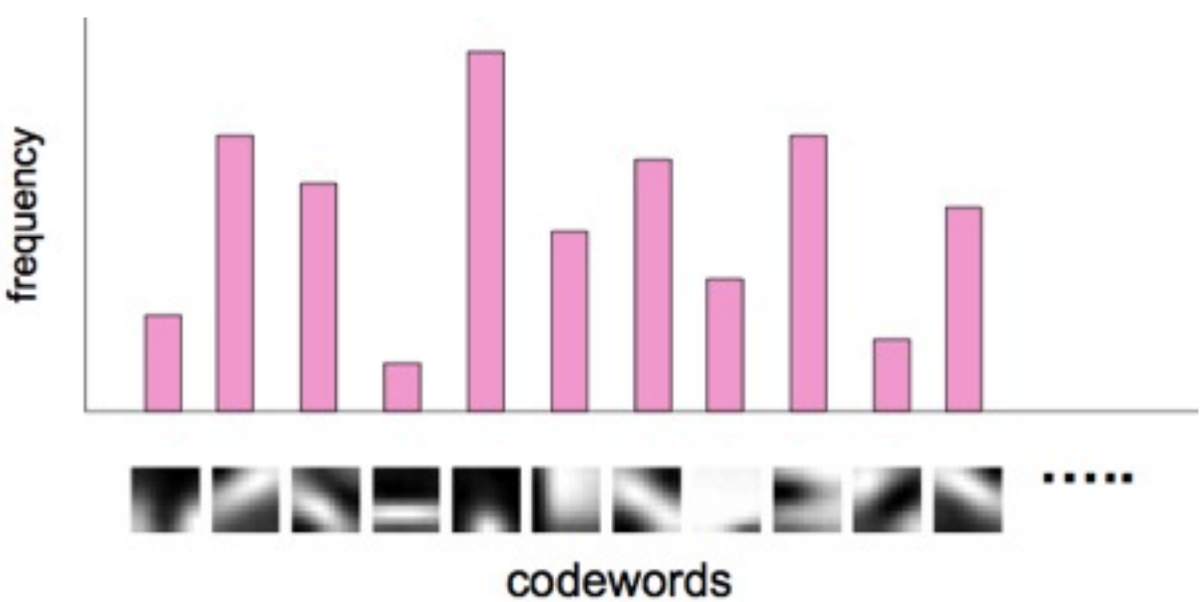
~100 visual words

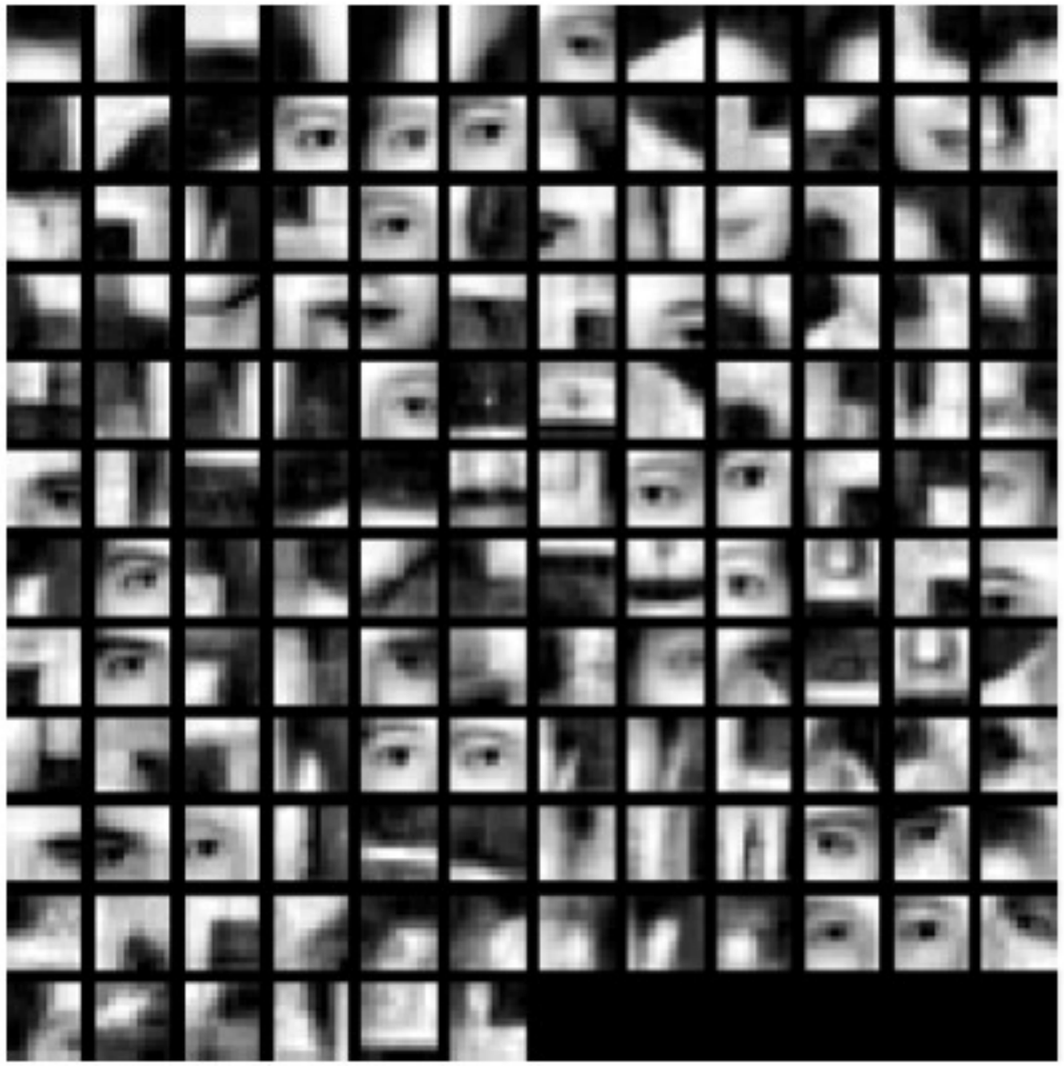
www.robots.ox.ac.uk/~az/**icvss08_az_bow**.pdf

Bag of visual word

- Visual word book

- histogram of word is the signature for an image

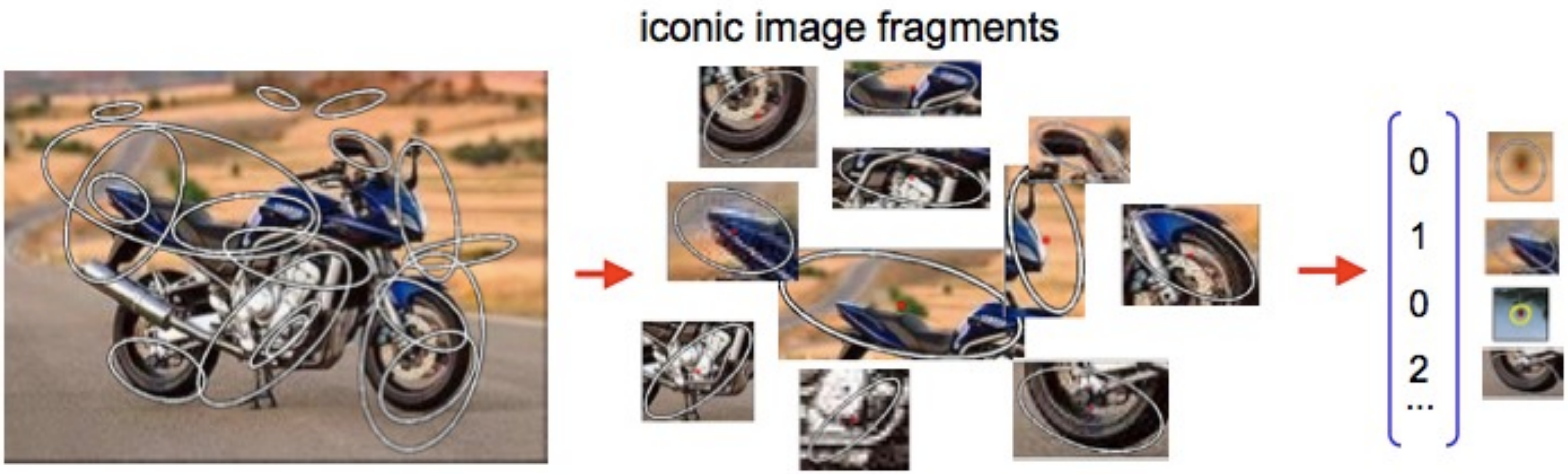




~100 visual words

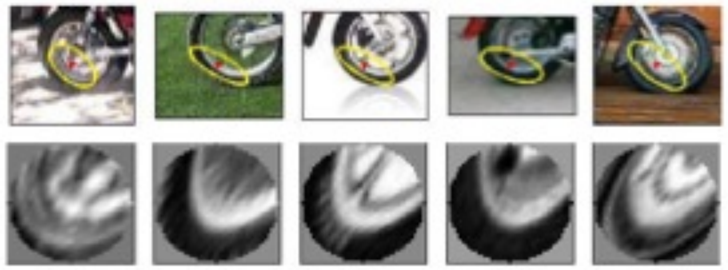
www.robots.ox.ac.uk/~az/**icvss08_az_bow**.pdf

Bag of visual word



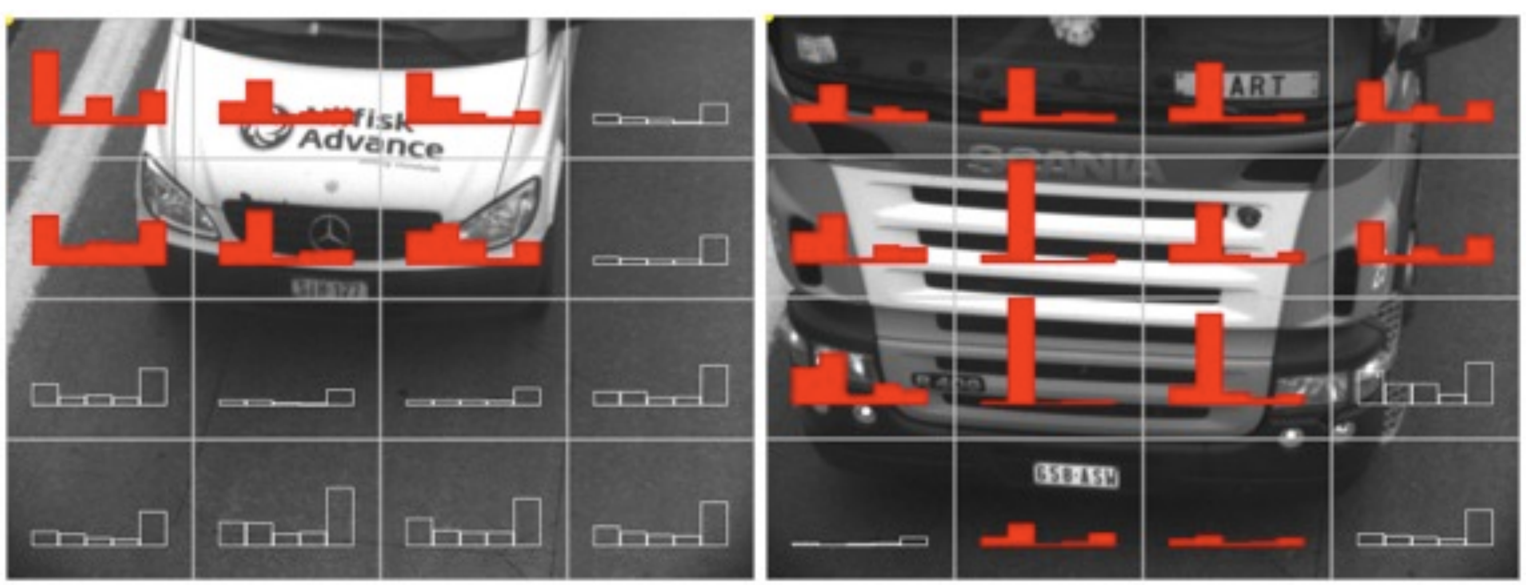
www.robots.ox.ac.uk/~az/**icvss08_az_bow**.pdf

- detection of robust points
- extraction of point descriptors
- find the closest match in the word book
- compare signature with database
- no segmentation
- no localization



Edge histogram

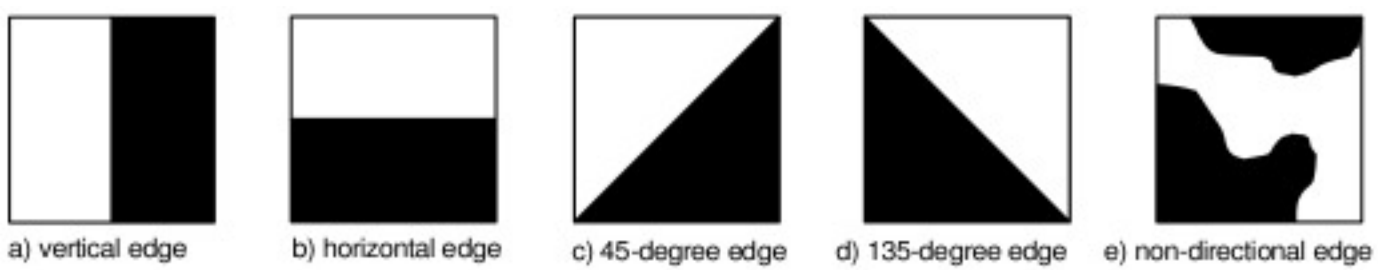
- basically a texture descriptor
- edge histogram computed on image parts
    - edge detection
    - grouped by image part
- histogram are then combined in one unique signature



Ch. Kaes

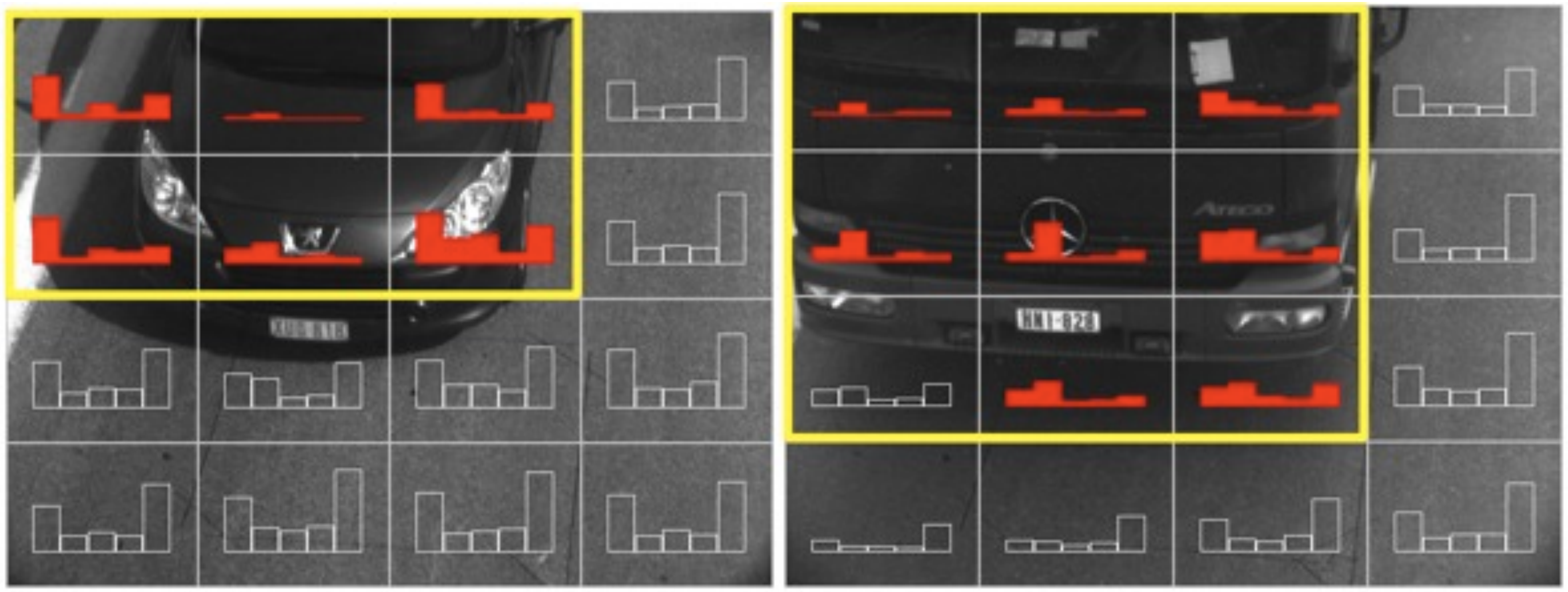
Edge histogram

- different edge type measures (directional / un-directional)

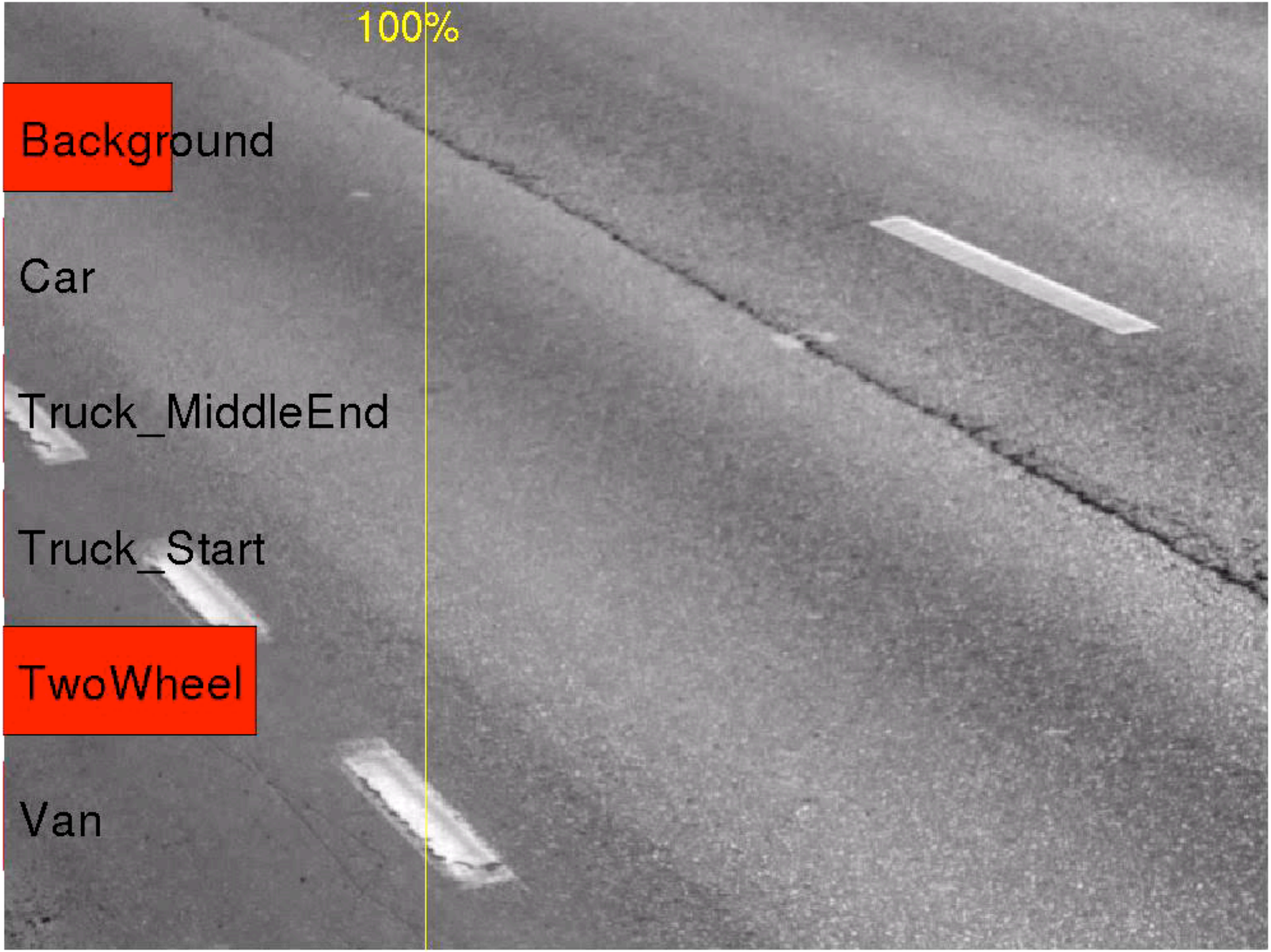


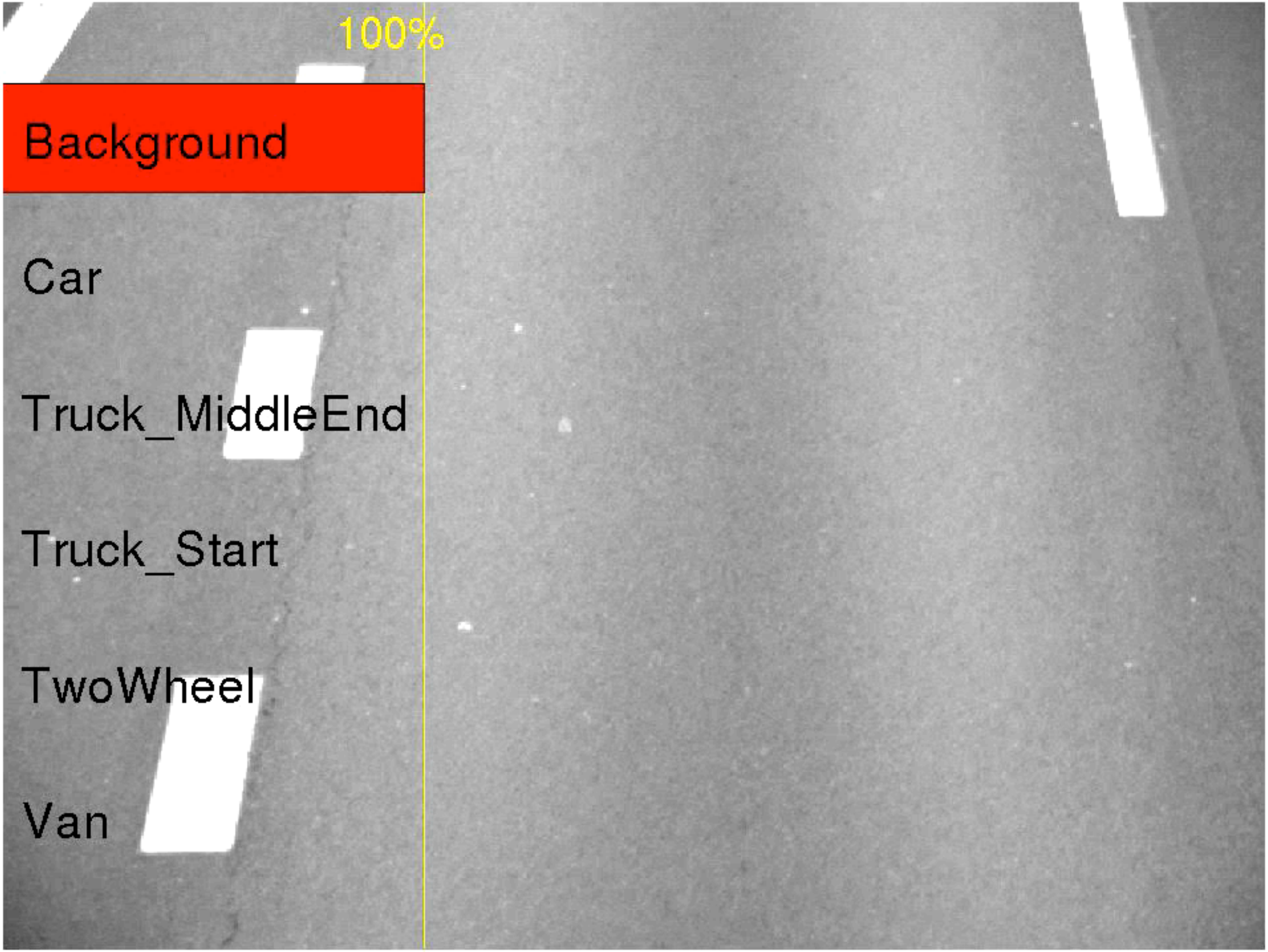
Edge histogram

- spatial localization:
  - adaptive approach
  - only histogram different to background are grouped into signature
  - signature are compared using supervised classification
    e.g. SVM



Ch. Kaes


100%
Background
Car
Truck_MiddleEnd
Truck_Start
TwoWheel
Van


100%
Background
Car
Truck_MiddleEnd
Truck_Start
TwoWheel
Van