Introduction to Robotics for cognitive science

Dr. Andrej Lúčny

KAI FMFI UK

lucny@fmph.uniba.sk

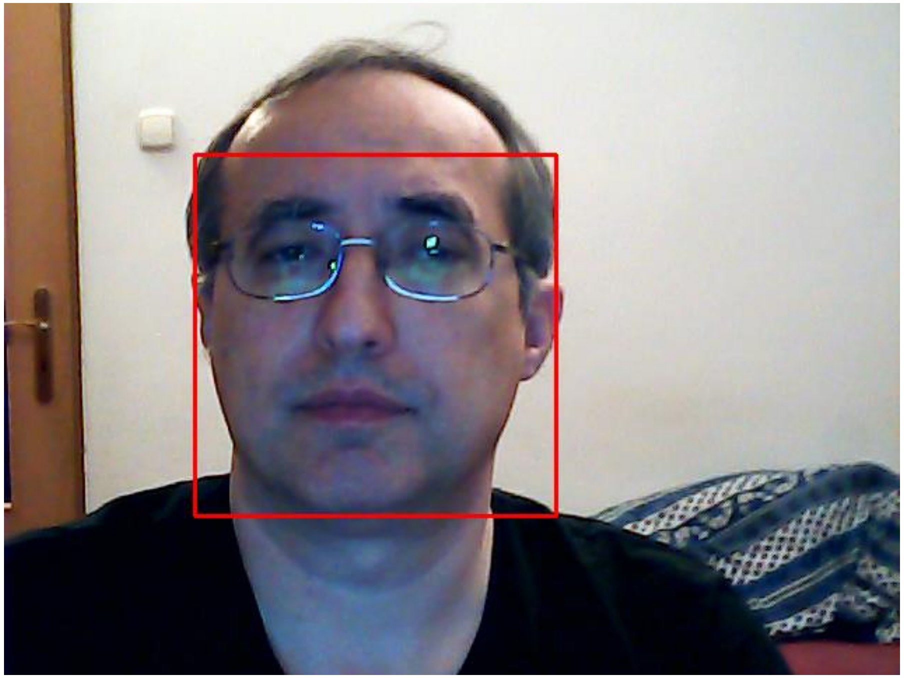
Web page of the subject

www.agentspace.org/kv

More general object detectors

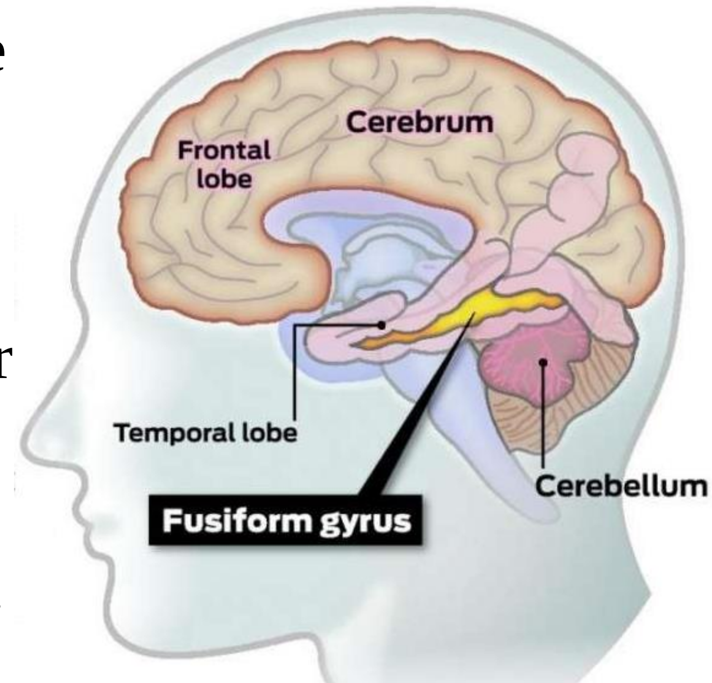
So far, we have been detecting and tracking very particular objects, e.g. part of your face

Now, we would like to detect a general object, e.g. the face (of any human regardless of age, gender or race)



Face detection in human brain

- The ability to recognize faces is strongly stored in our genes, and we do not need to learn it (unlike walking on two legs)
- In the brain, there is an anatomic structure responsible particularly for face recognition
- Its malfunction causes no other incapability if it happens to an adult but has profound effects if it is congenital, including no empathy for other people



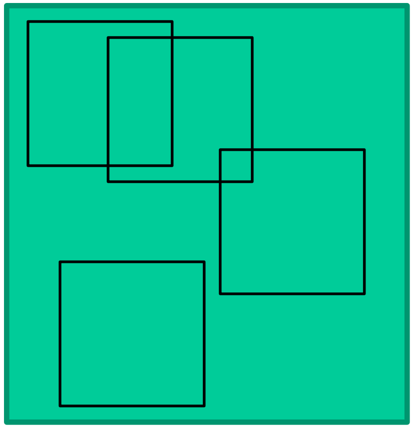
Machine Learning

- In robotics, we have to put into robot a model (instead of the human genes)
- The model is created from a dataset using a supervised machine learning method
- Data in the dataset have to be annotated
- The typical model is a classifier
- Classifier tells category (face: yes or no, animal: cat, dog, elephant, monkey, other)

Dataset

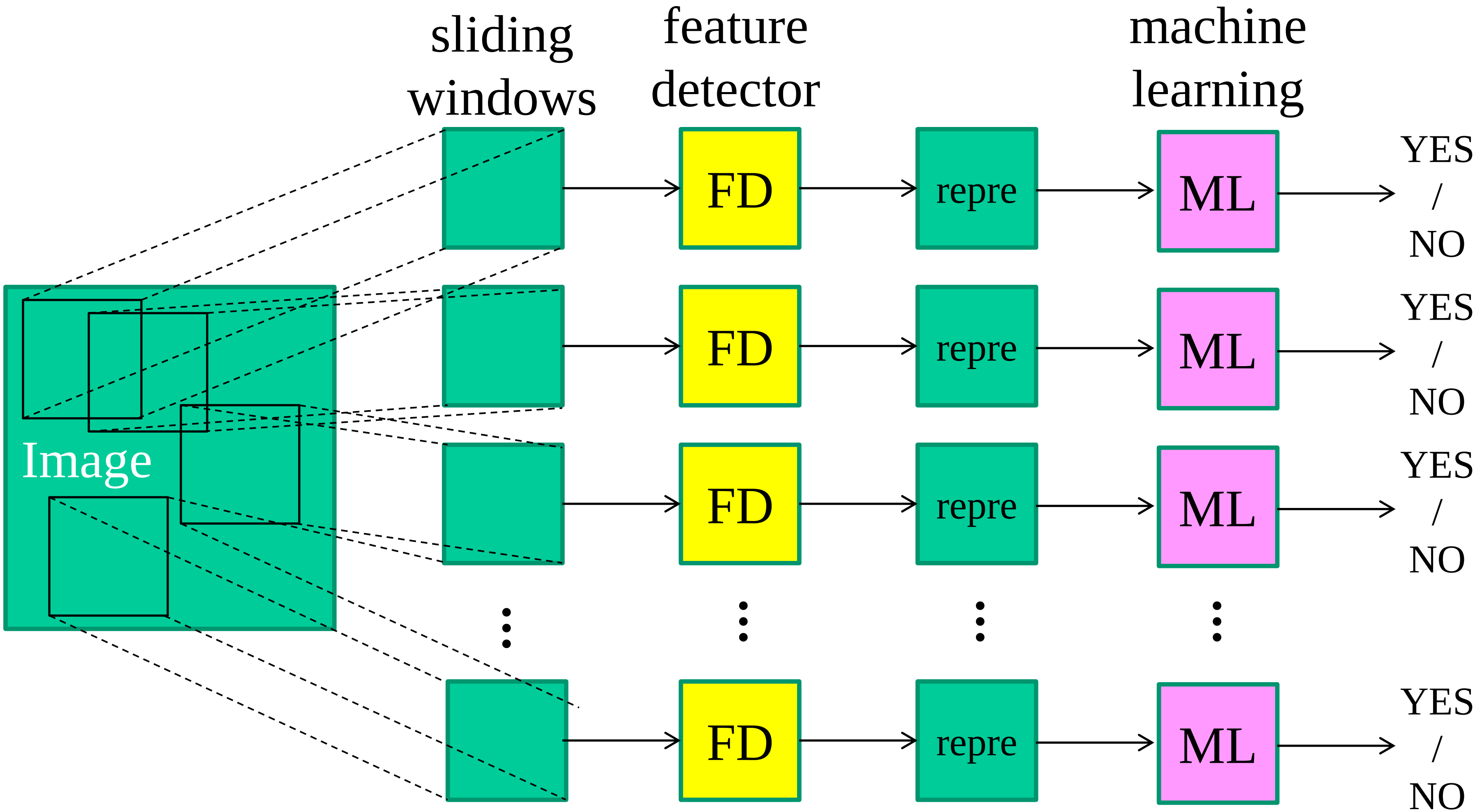
It contains positive and also negative samples

Sliding window



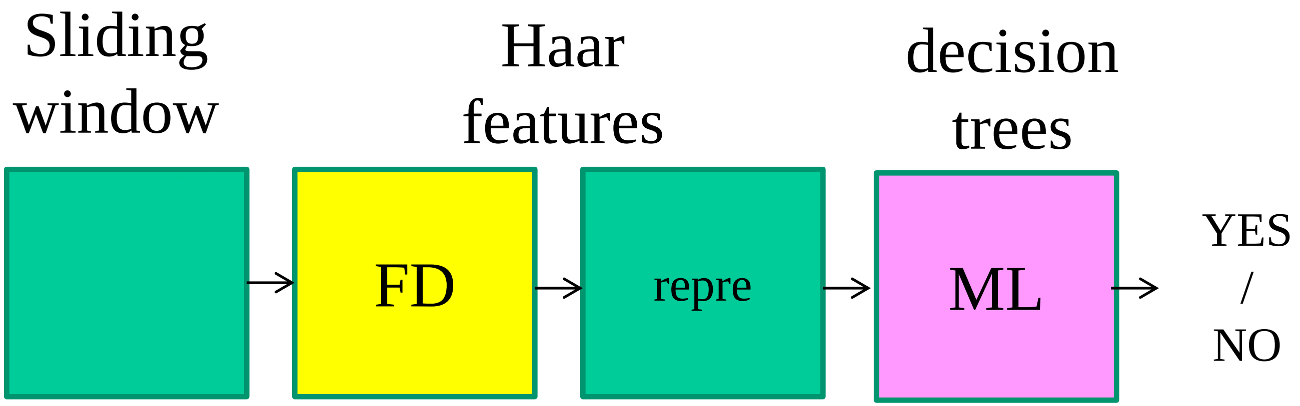
- The classifier processes an image and returns YES or NO.
- The detector processes an image and returns parameters of rectangles and confidences of the detected objects.
- The sliding window is a universal algorithm that turns the classifier into the detector using brute force.

General schema of classifier-based detector



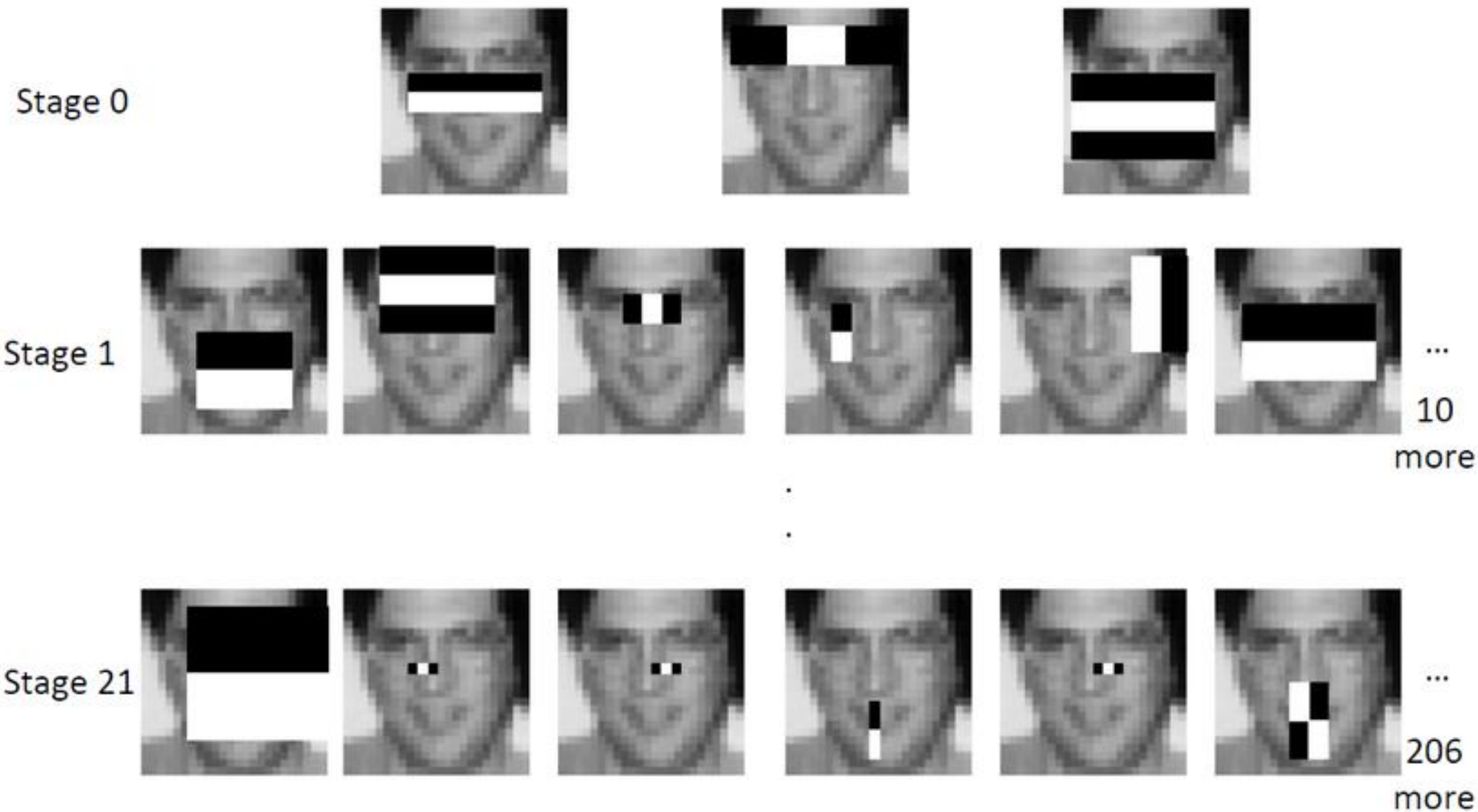
Viola Jones Algorithm

The first operational face detector [2001]

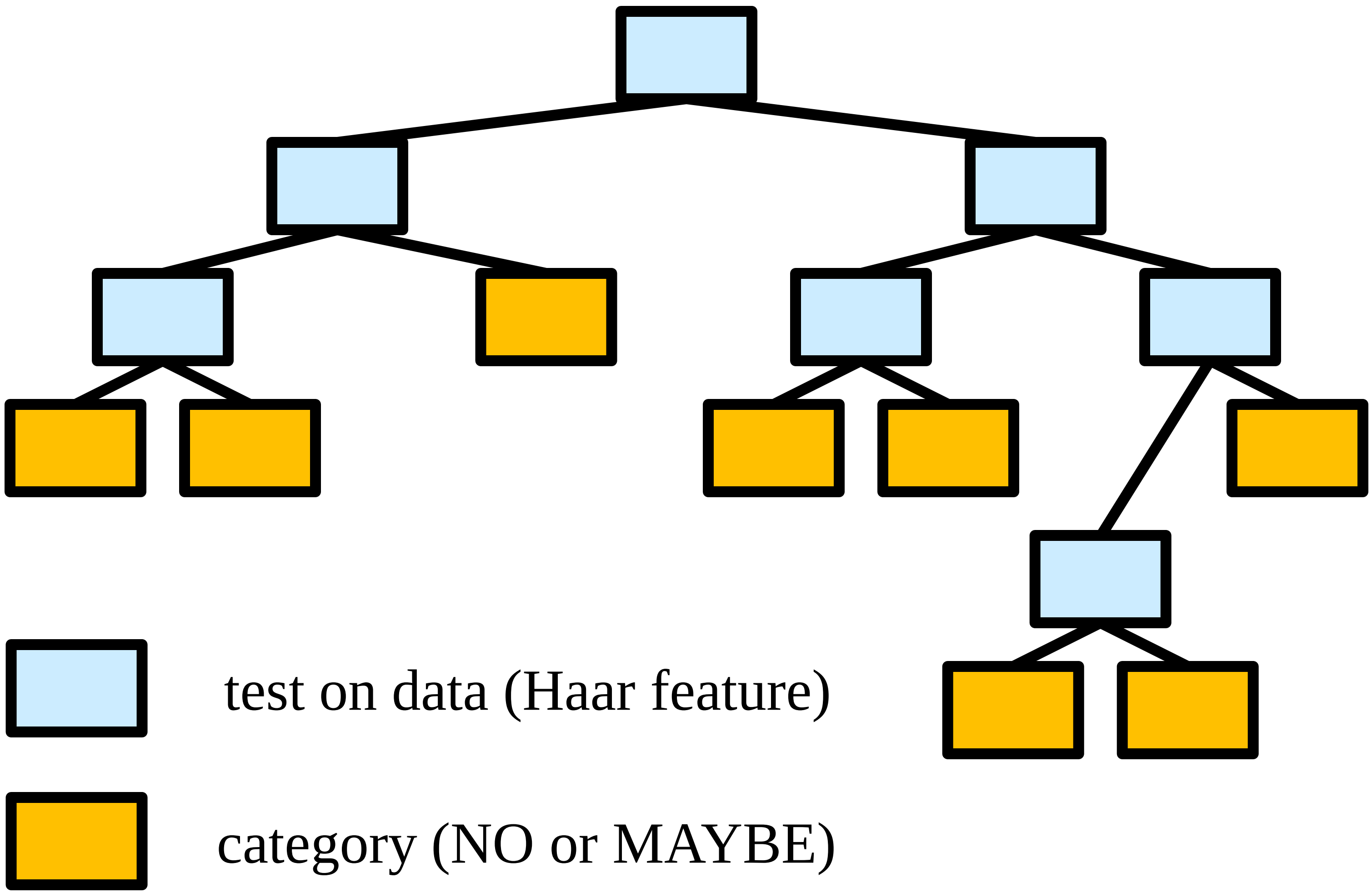


Haar Features

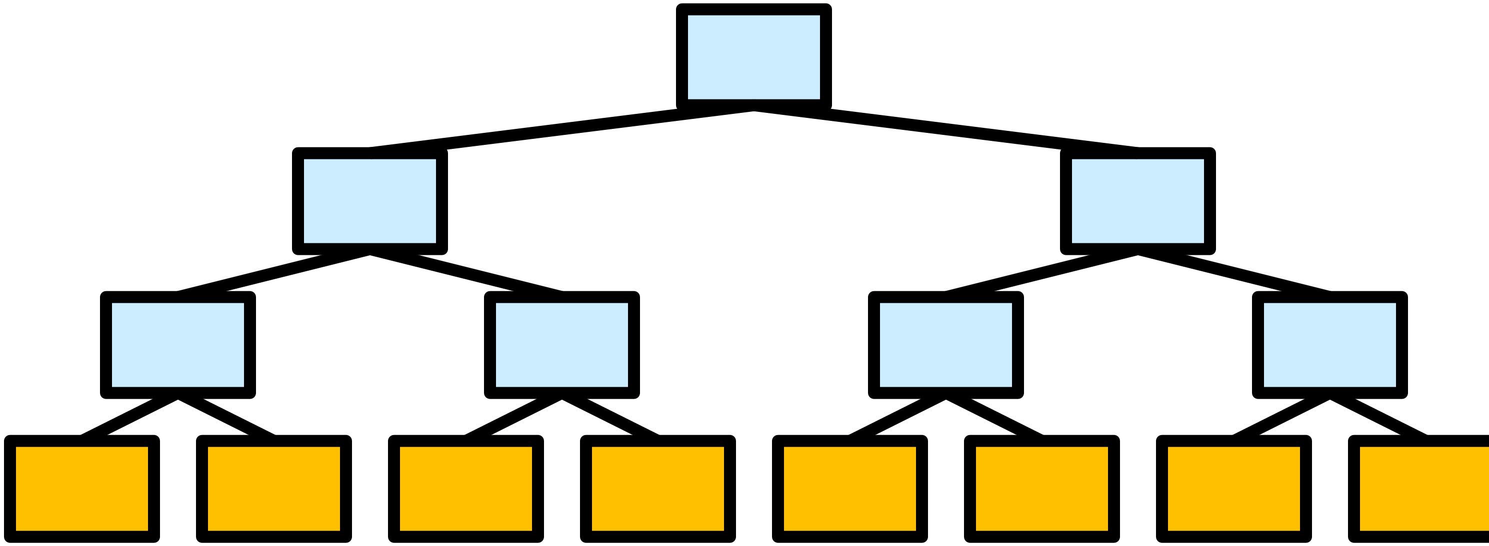
$$\sum c.\,\square < \sum c.\,\blacksquare \;?$$



Decision tree

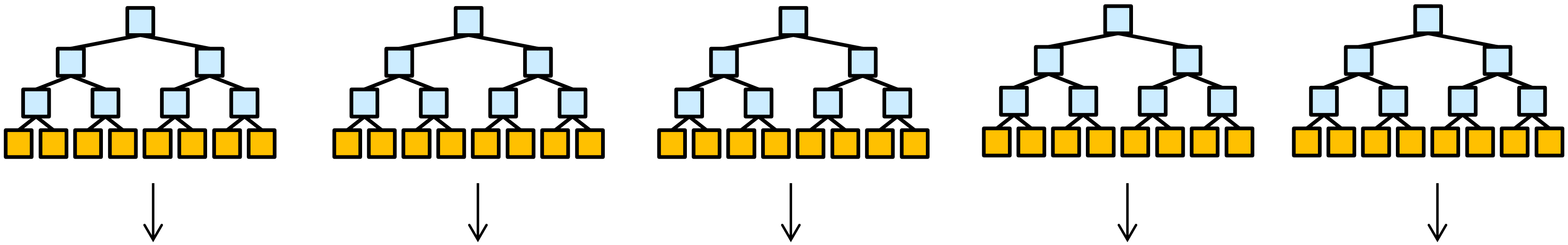


How to get a good decision tree?



- We generate candidate features and select one that best distinguishes faces from non-faces

Cascade classifier



NO or MAY-BE

If MAYBE, then NO or MAY-BE

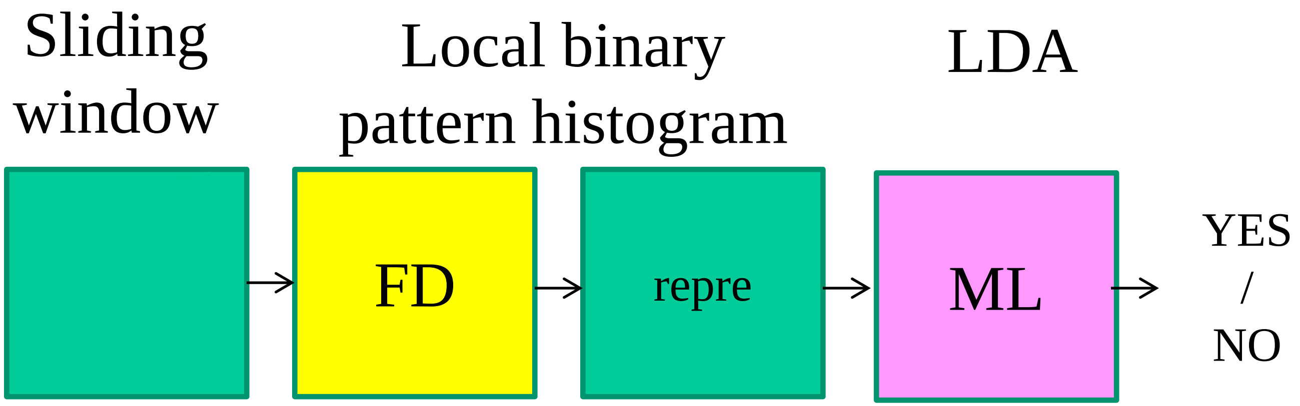
If MAYBE & MAYBE, then NO or MAY-BE

…

NO or YES

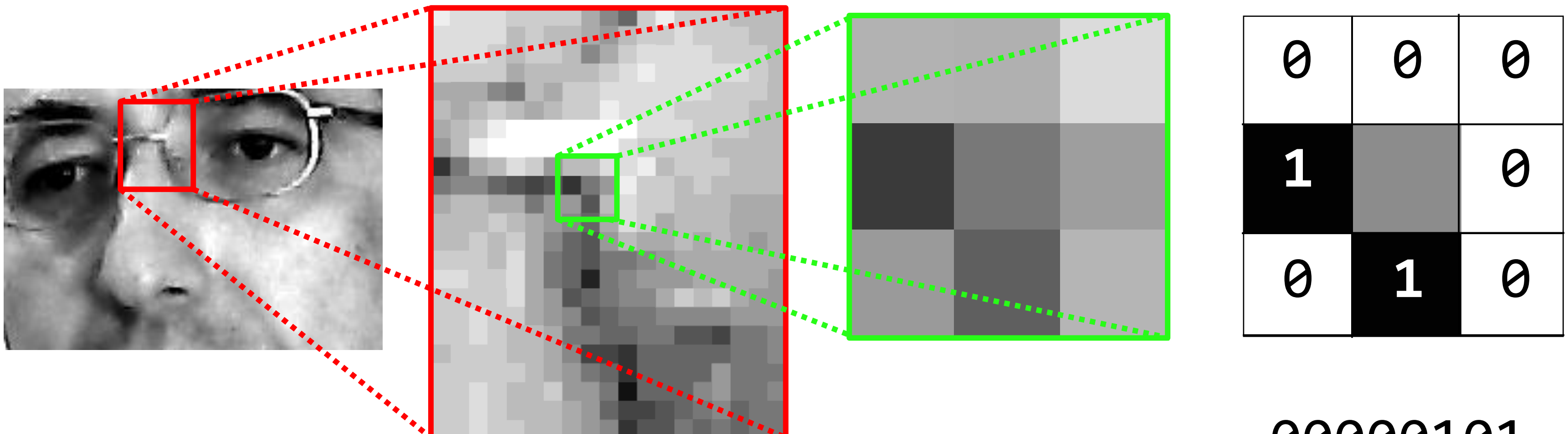
Viola-Jones algoritmus: Haar features + cascade classifier

LBPH approach



[T. Ojala, M. Pietikäinen, and D. Harwood 1994]

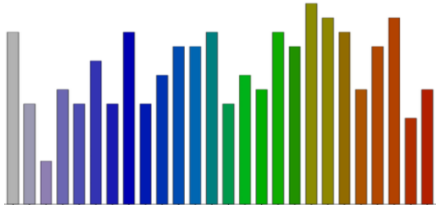
LBPH Features



00000101

5

- Each pixel has its LBP
  - 0 .. 255

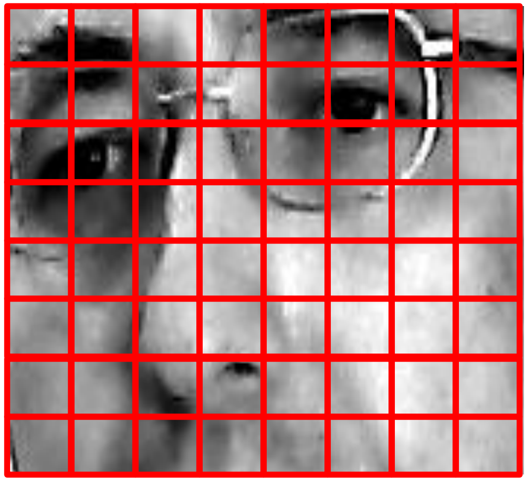


LBPH Features

`3F 2E 01 ... FF`

256 bytes

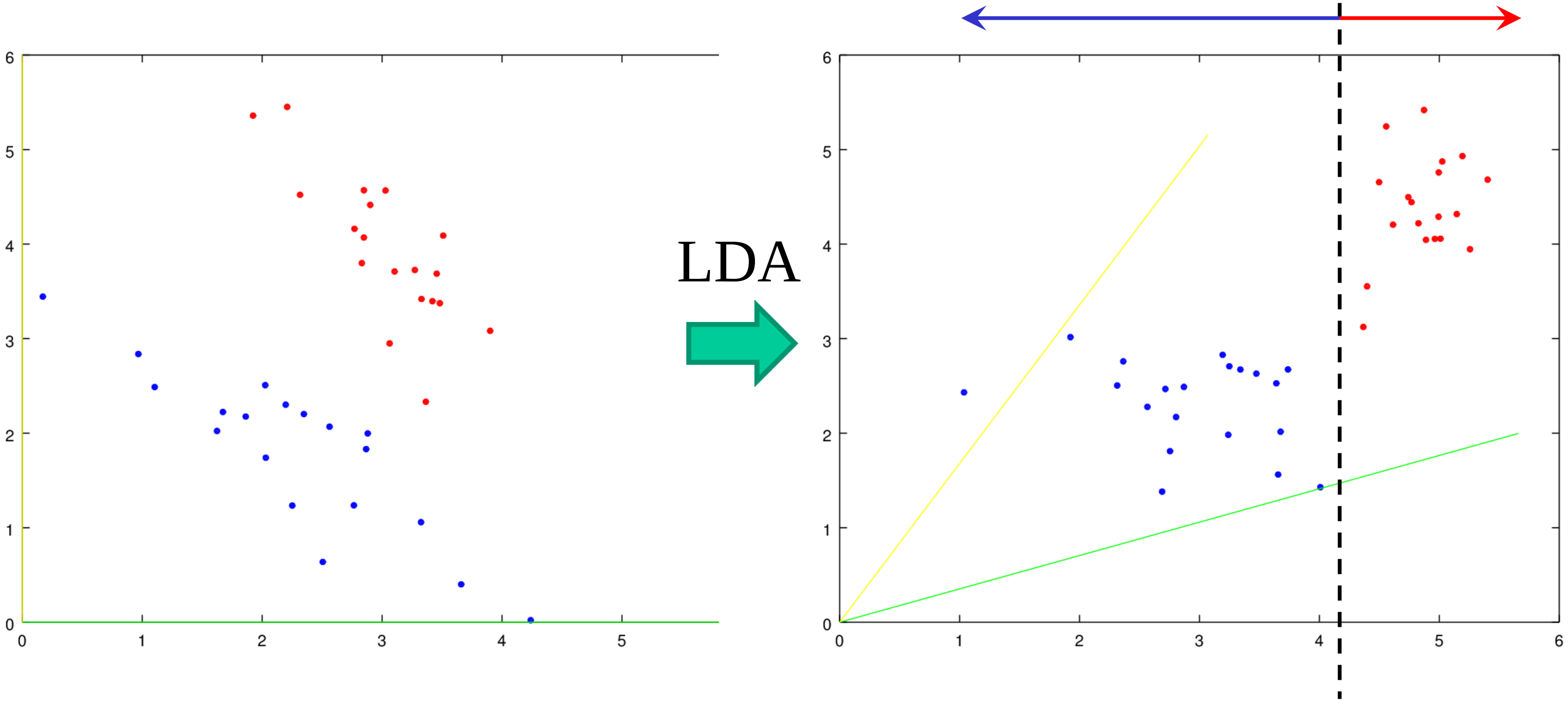
- Each region can be associated with histogram of LBP



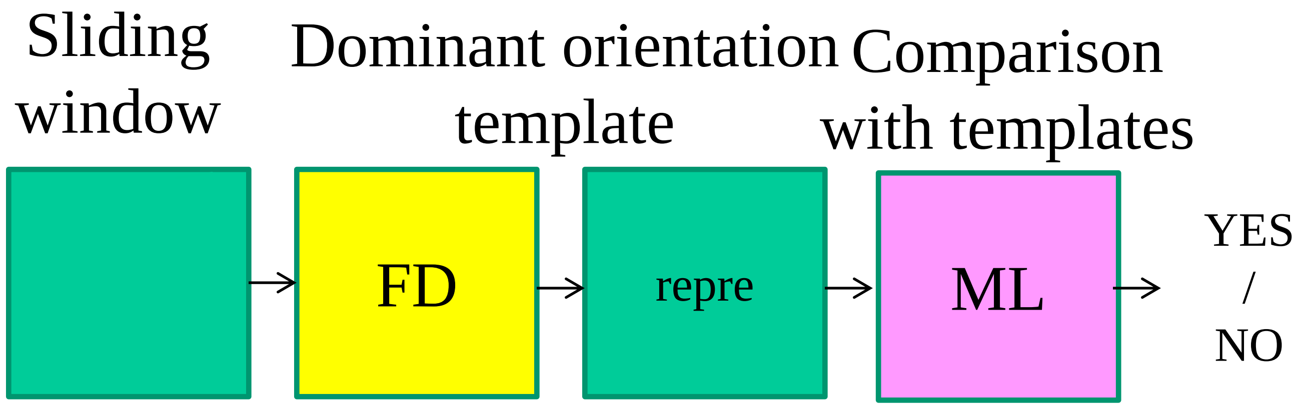
- Object is represented by set of LPBH

Linear Discriminant Analysis

- Data are represented as points in multidimensional space (fixed number of dimensions)
- The space is reduced and transformed to easy distinguish e.g. data categories



Dominant Orientation Templates



[Hinterstoisser, 2010]

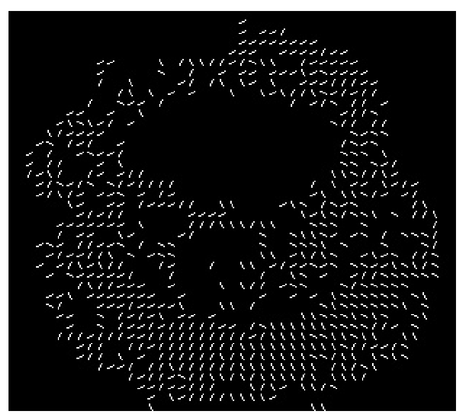
DOT Features

obraz

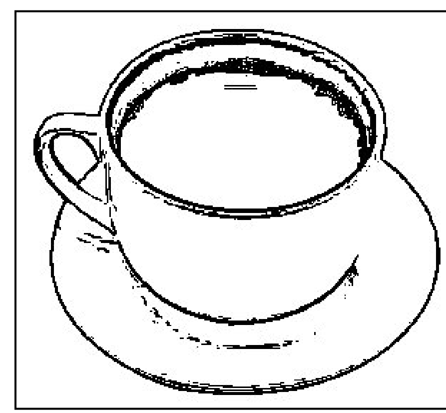
dx

dy

|gradient|

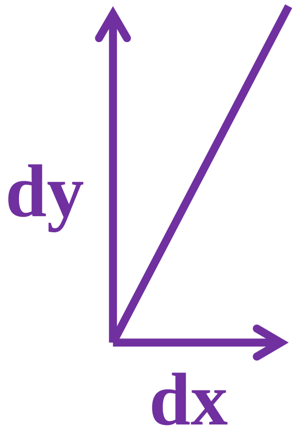


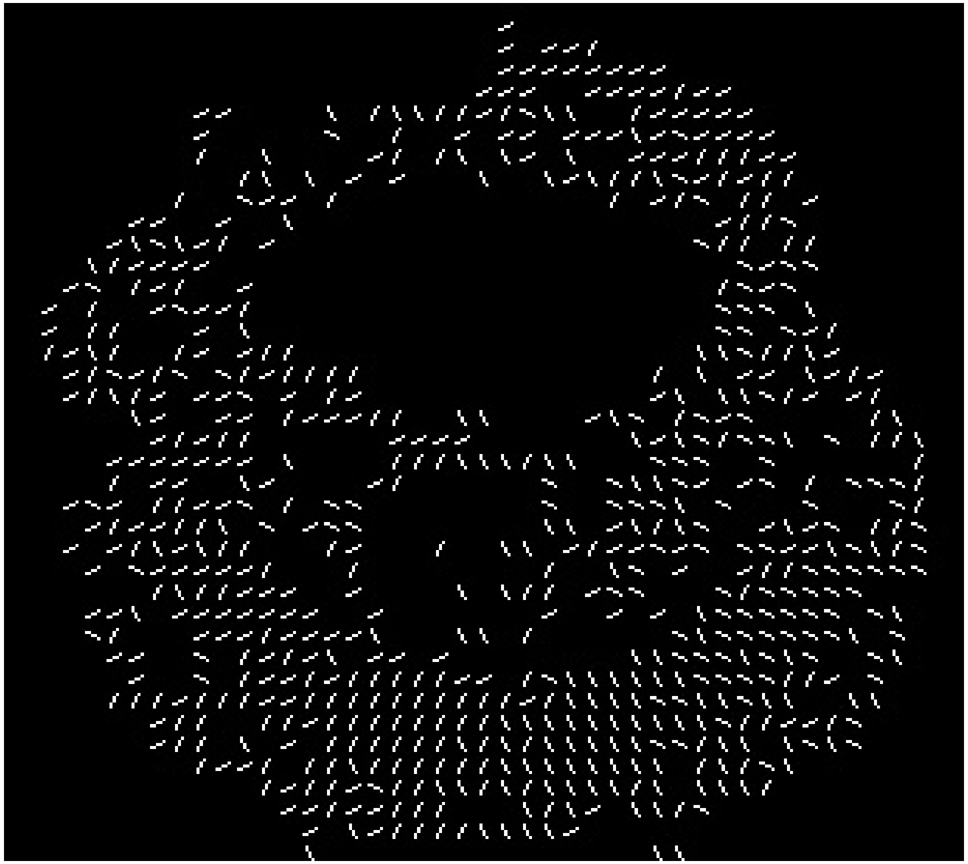
slope

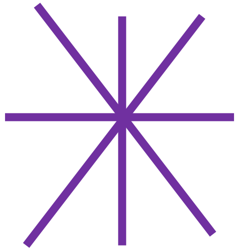


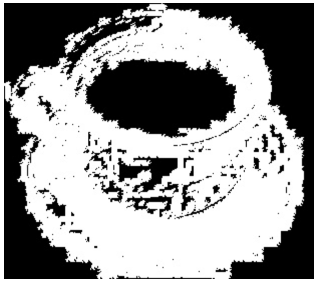
edges

Edge orientations (=slopes)







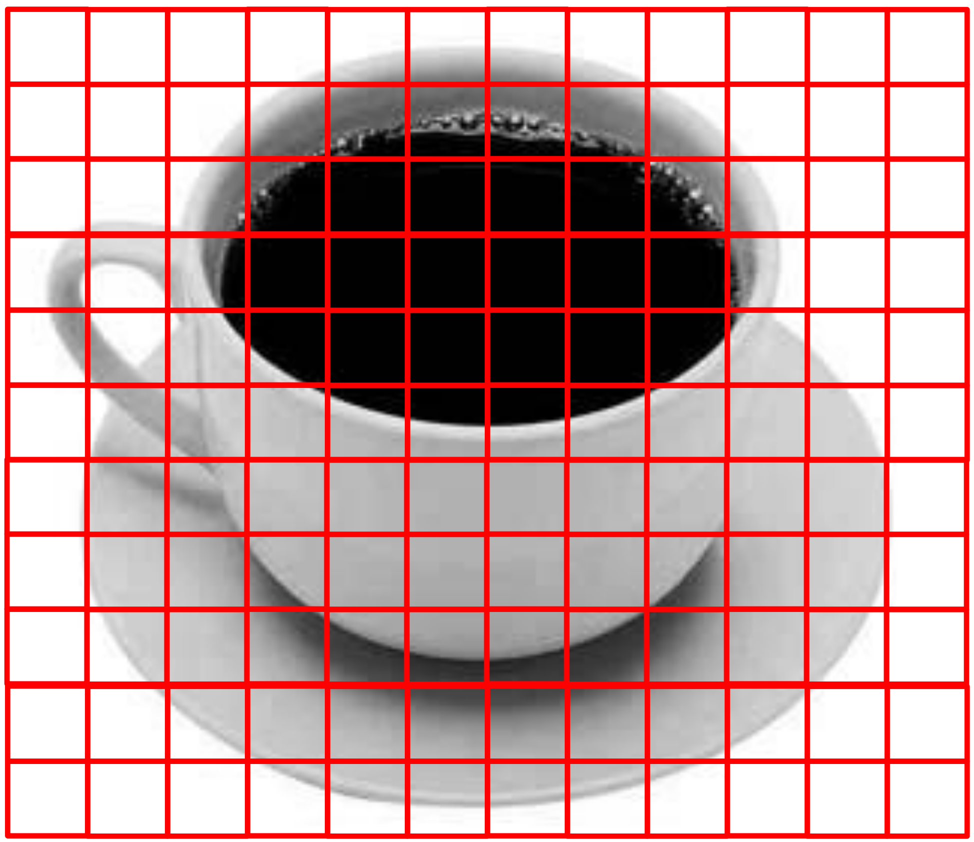


DOT Templates

- Just clustered dominant orientations will represent the object

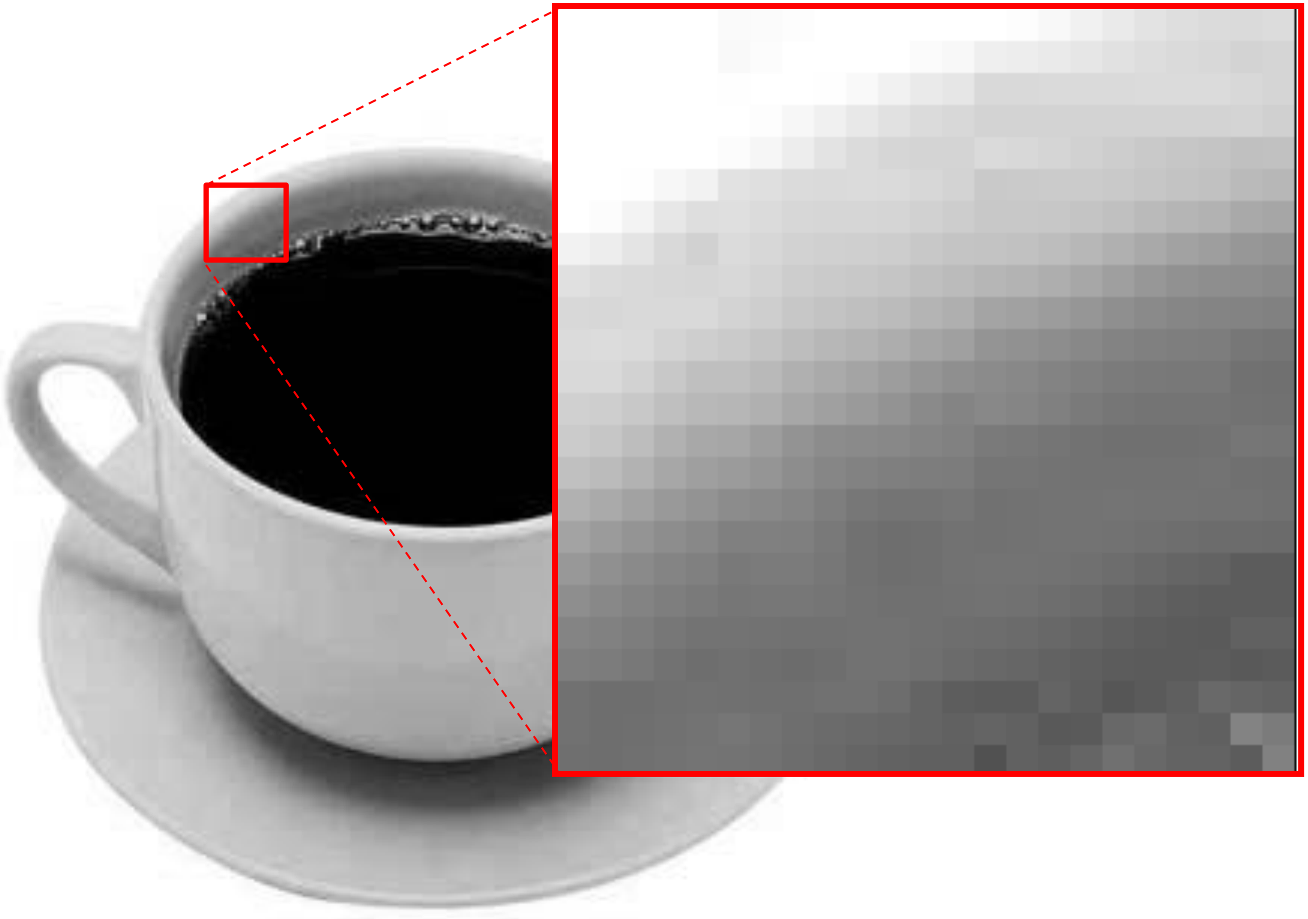
DOT Templates

- We cover object with set of non-over lapping regions



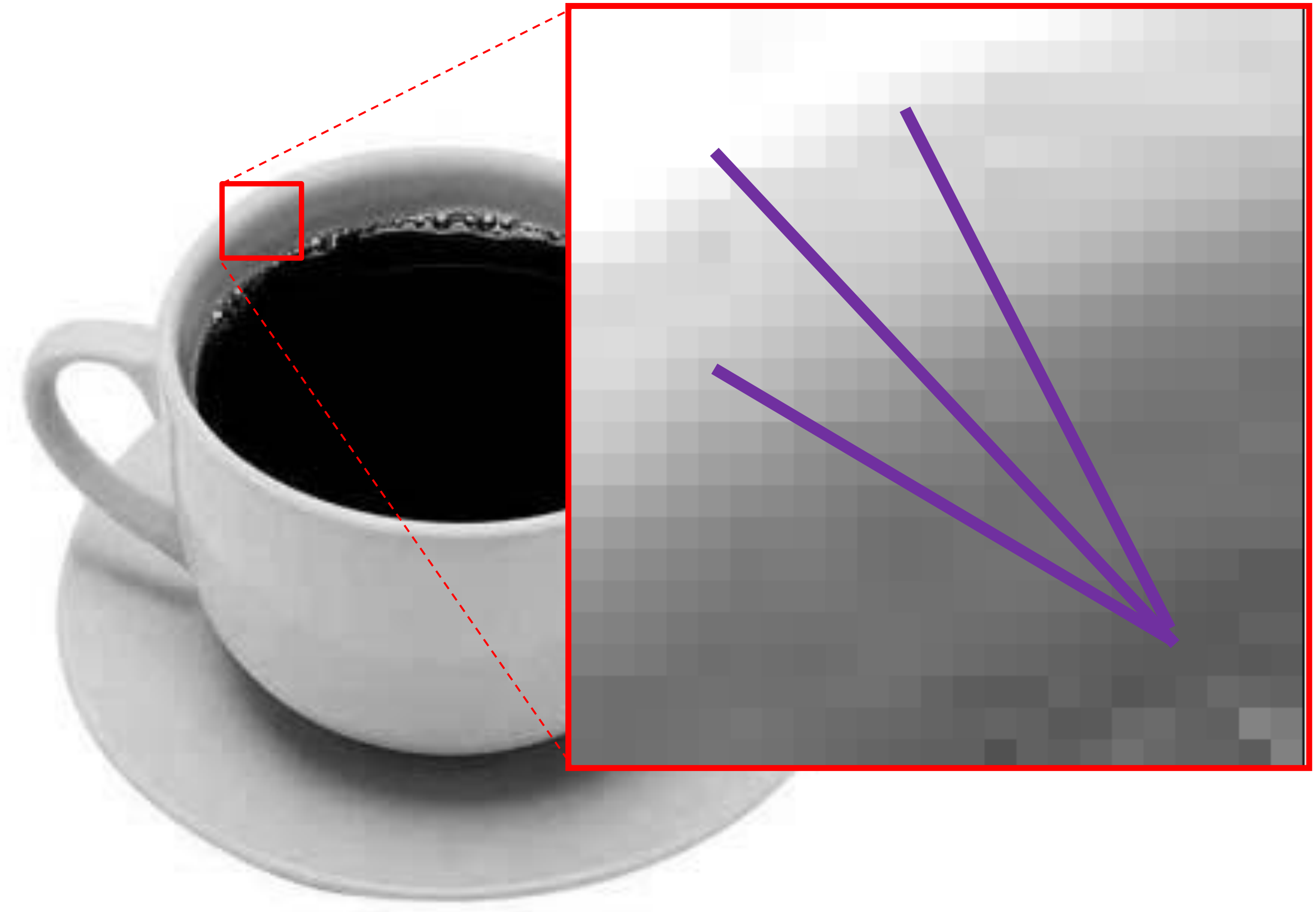
DOT Templates

- For each pixel we calculate edge slope



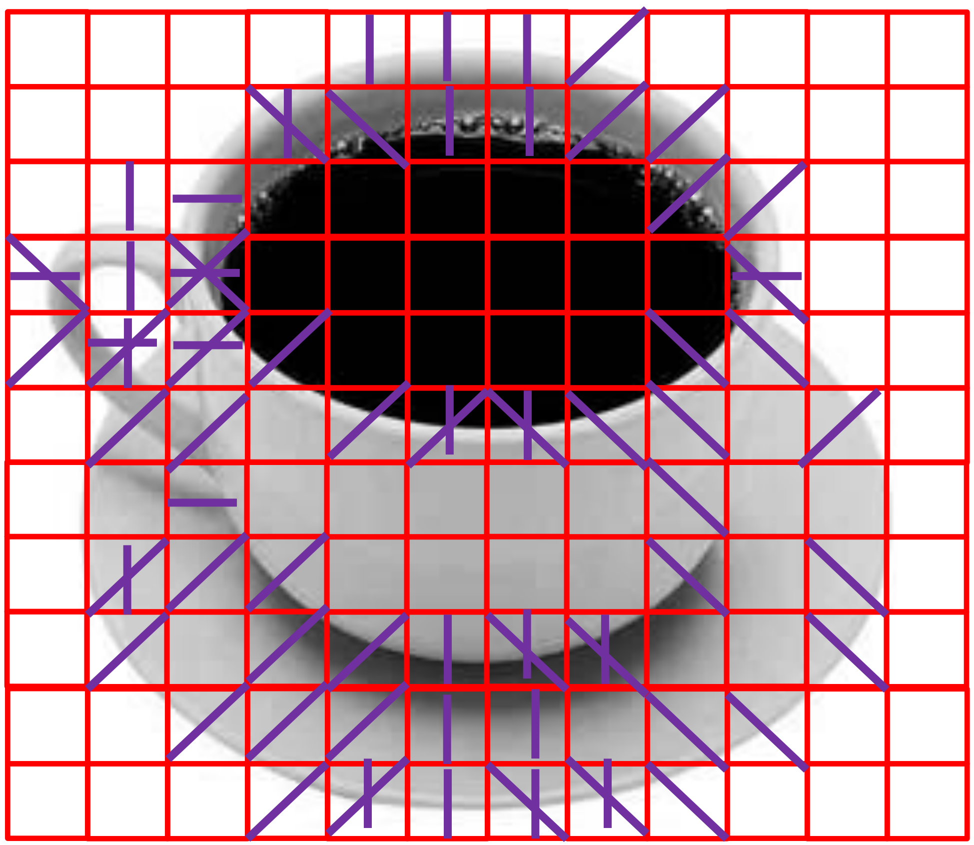
DOT Templates

- We select prevailing slopes



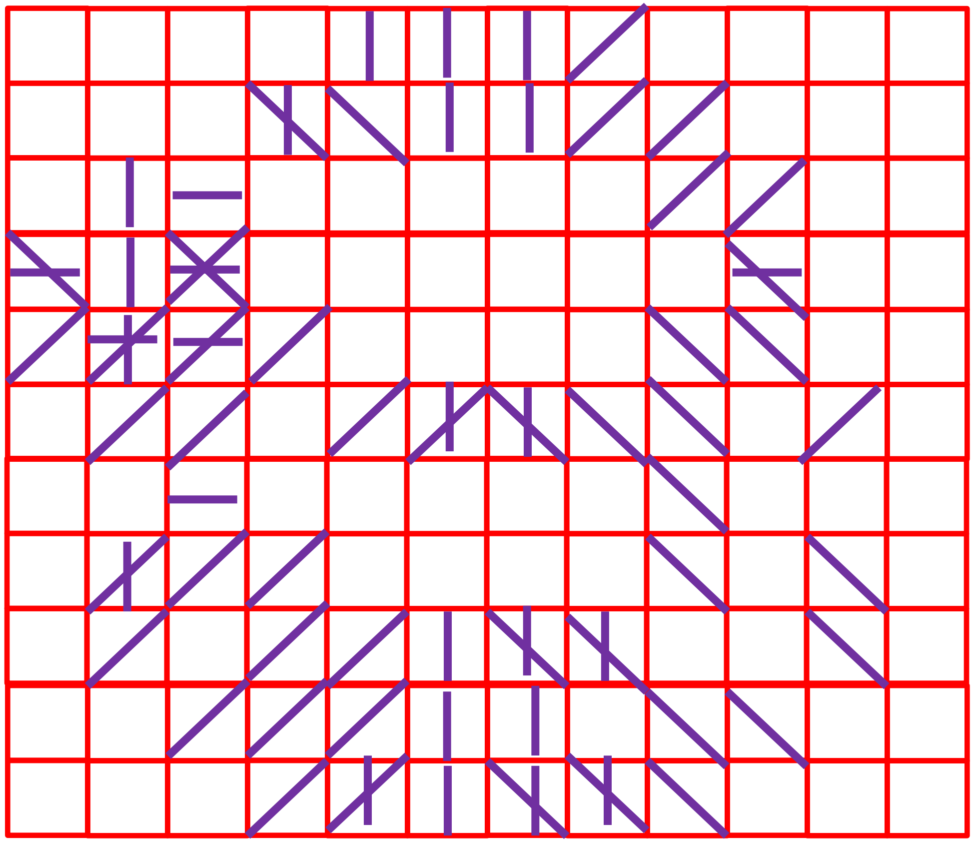
DOT Templates

- And this our template …



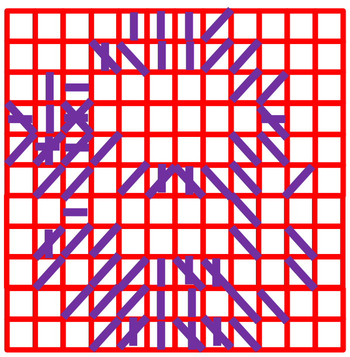
DOT Templates

- … which represent the object

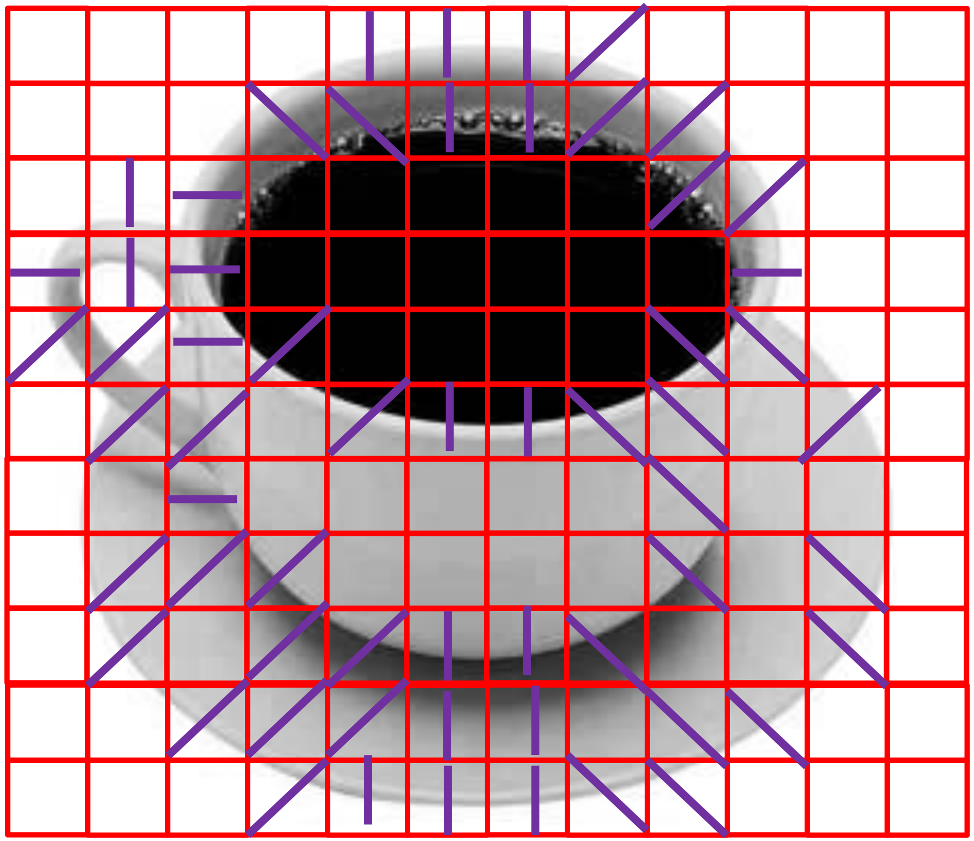


Searching for object

- For each place we calculate one dominant orientation for each region



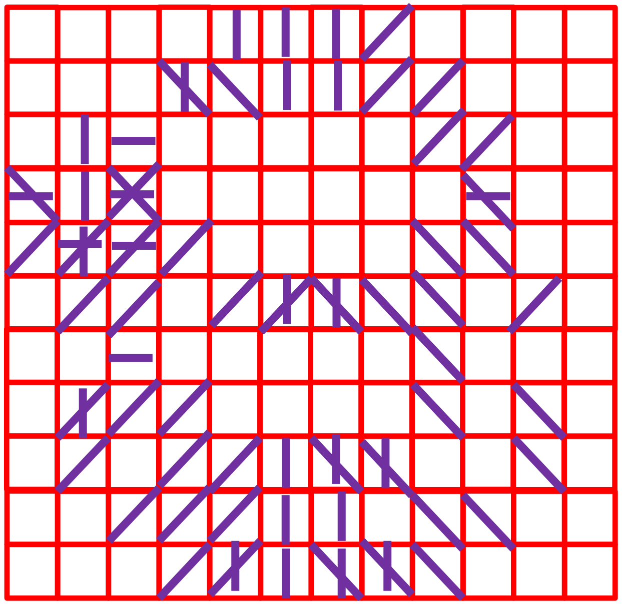
template



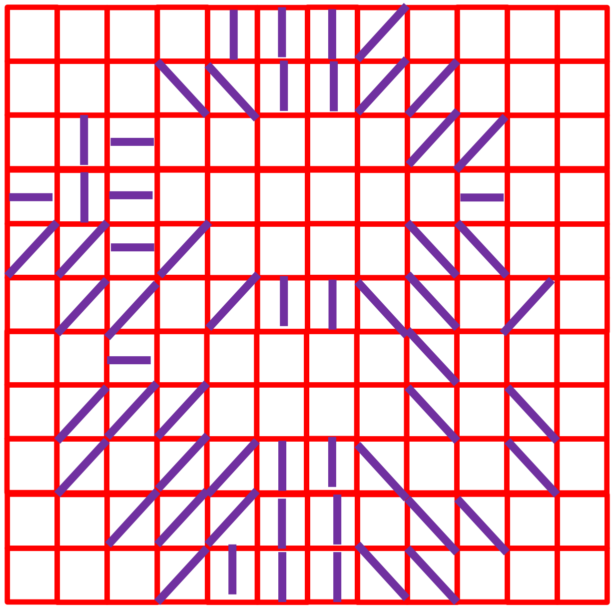
image

Comparison with templates

- If the image matches template, we have found the object



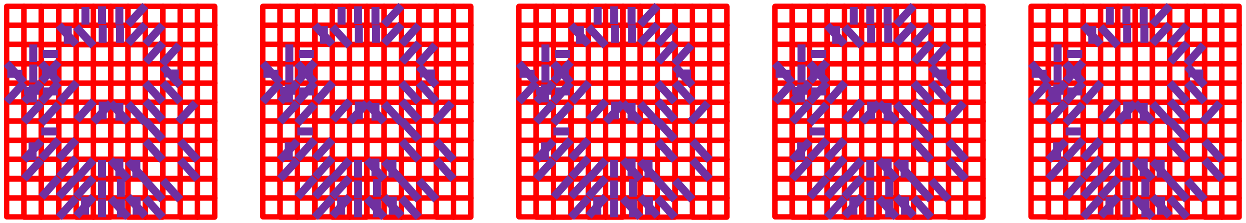
template

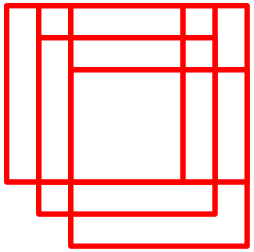


image

What about translation of regions?

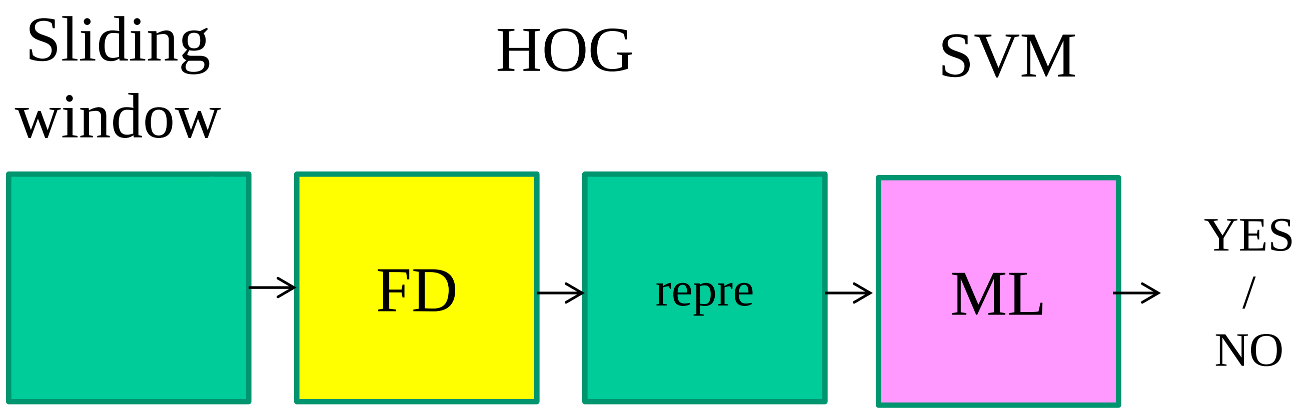
- No problem we prepare more templates for various translations, even various viewpoints, but same size

templates 

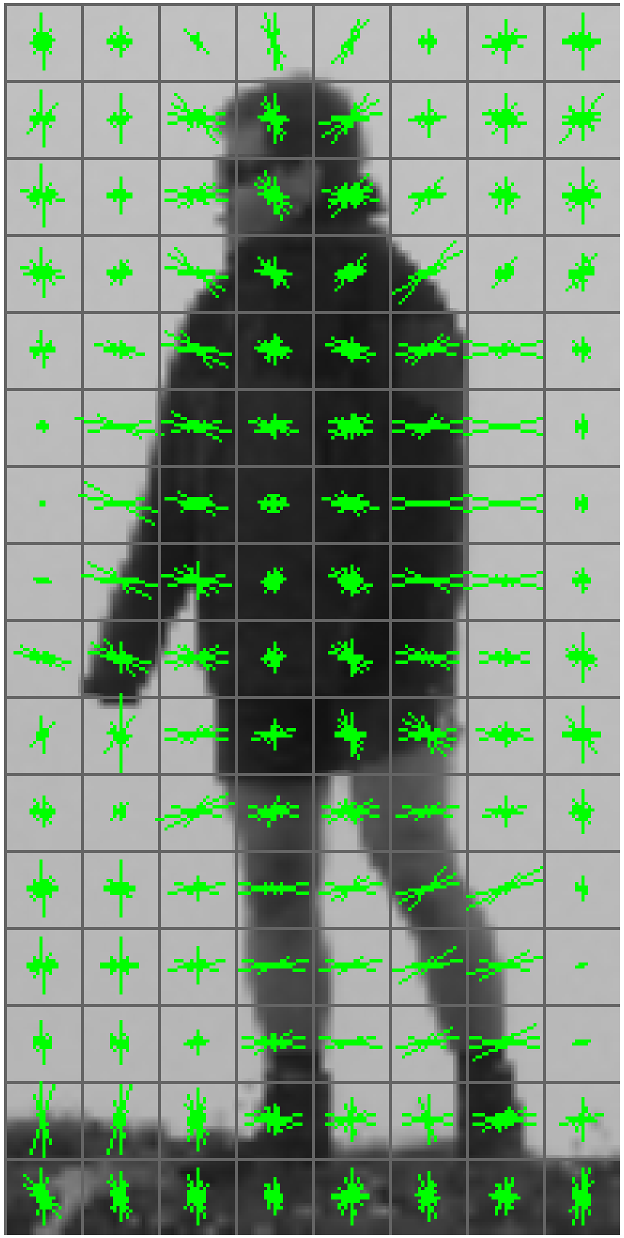


- Then we try all parts of image with variable size

Histogram of Gradients



[Dalal, Triggs, 2005]

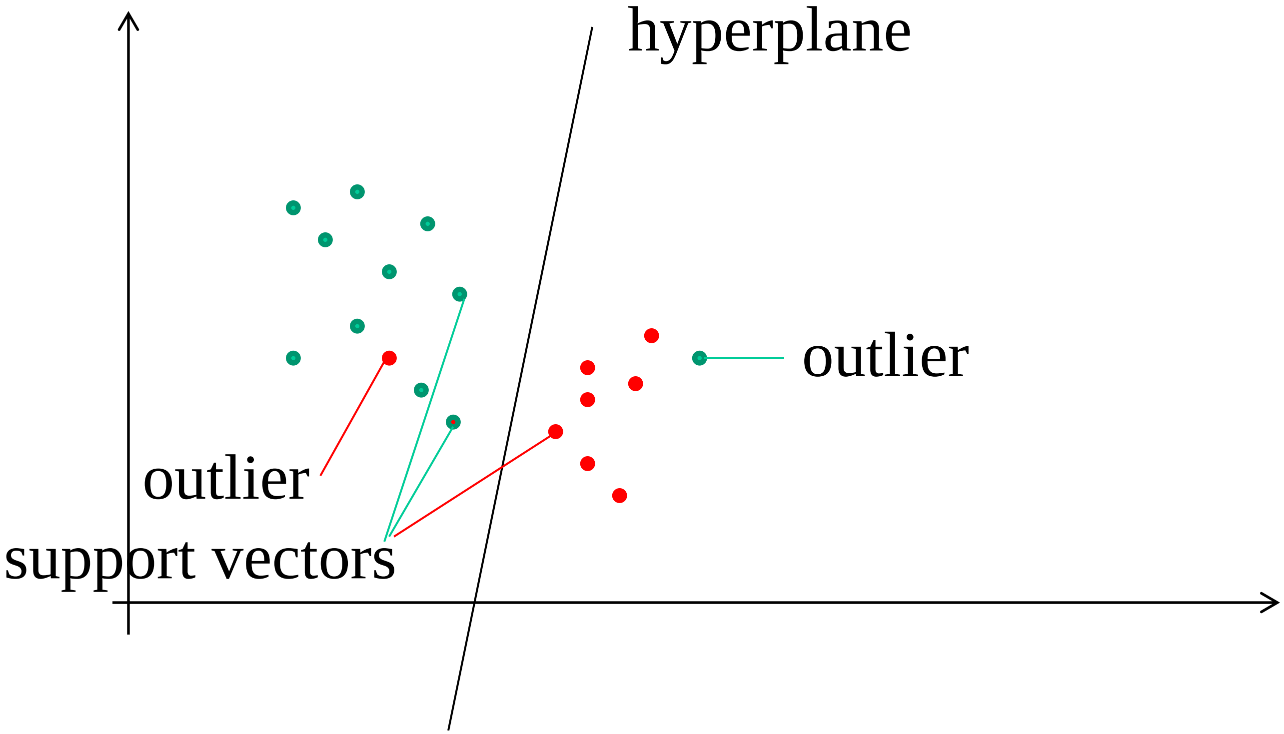


HOG features

- Instead of few dominant slopes we take their histogram

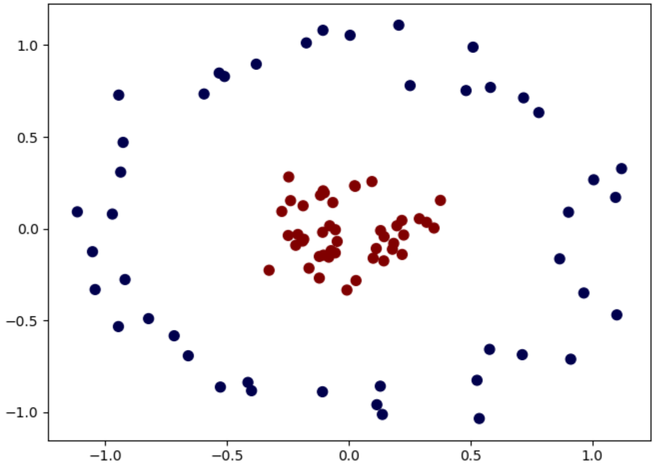
(Linear) Support Vector Machine

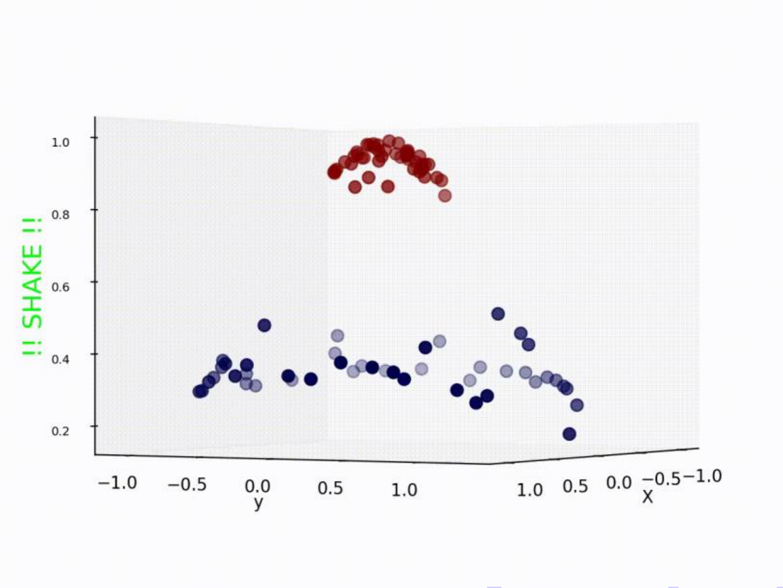
- Fast and good method which can handle outliers by maximalization of so called soft margin



SVM Kernel trick

- SVM expand dimension of data by application of a kernel to enable separation by hyperplane



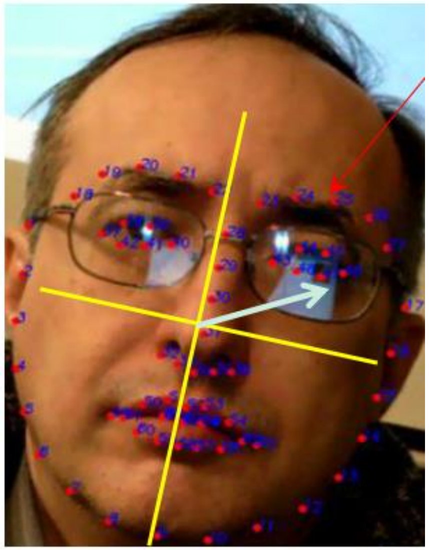


[more details]

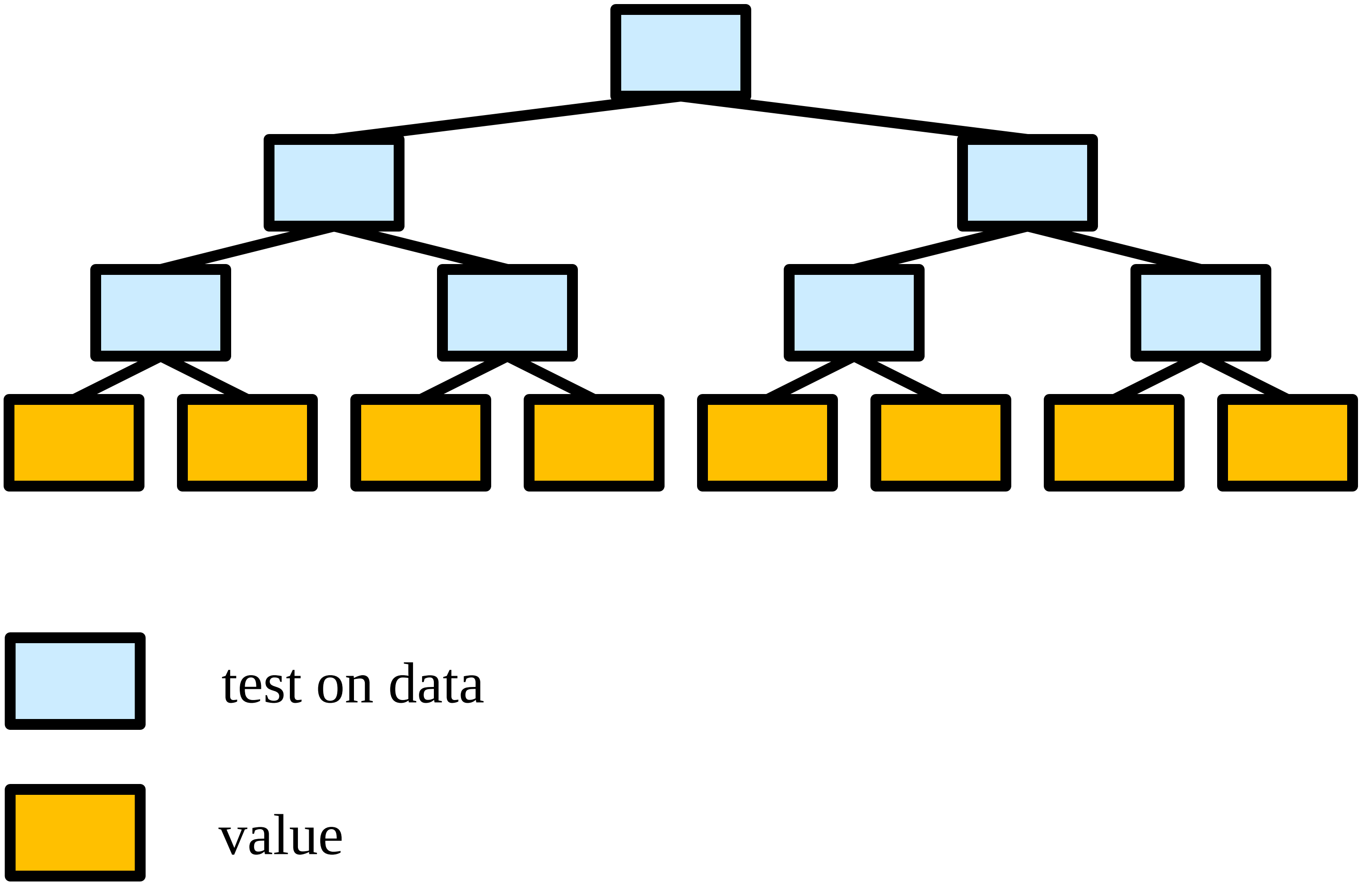
- kernels are selected so that distance in the expanded space can be calculated directly in the original space

Regression

- Analogically to classification, we can have model of regression, which does not provide category but value.
- E.g., Kazemi facial landmark detector employs a cascade regressor based on decision trees (features are based on the intensities comparison of two pixels positioned relative to average face landmarks). So, we put average face features on the image, and the regressor tells us how to move them to the correct position.



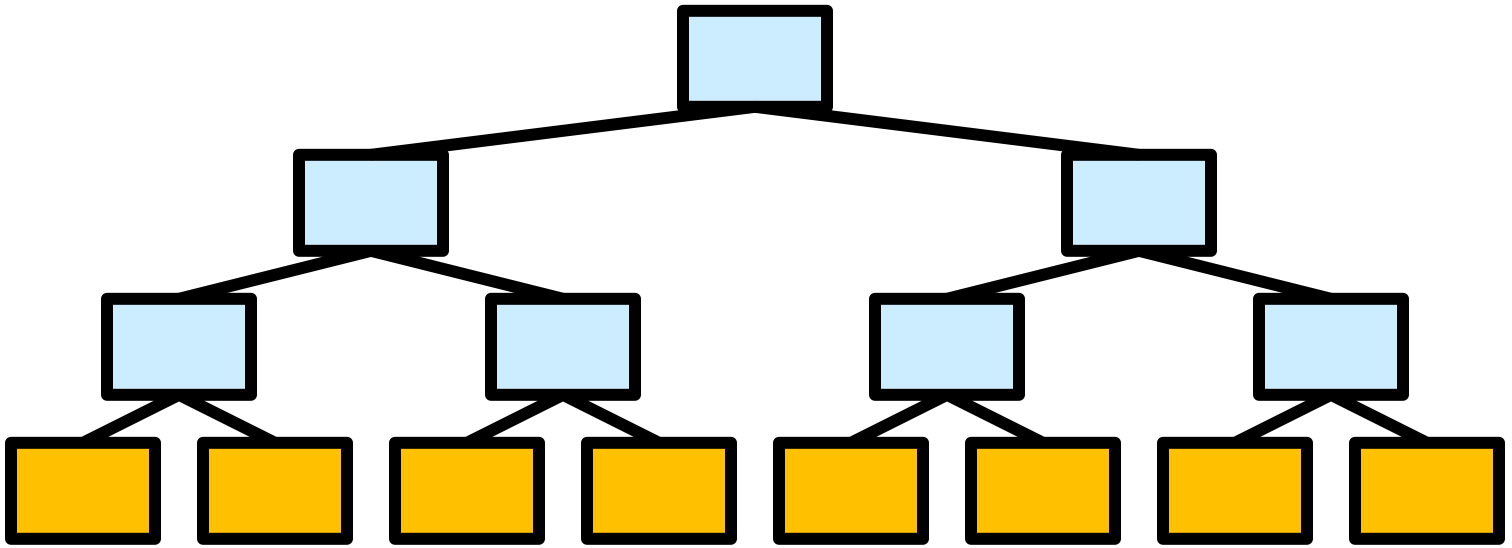
Weak Regressor



test on data

value

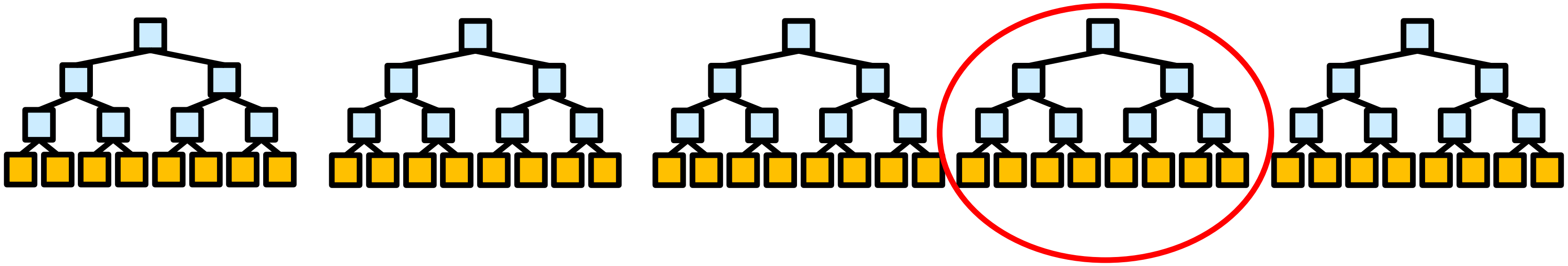
How to get a regression tree?



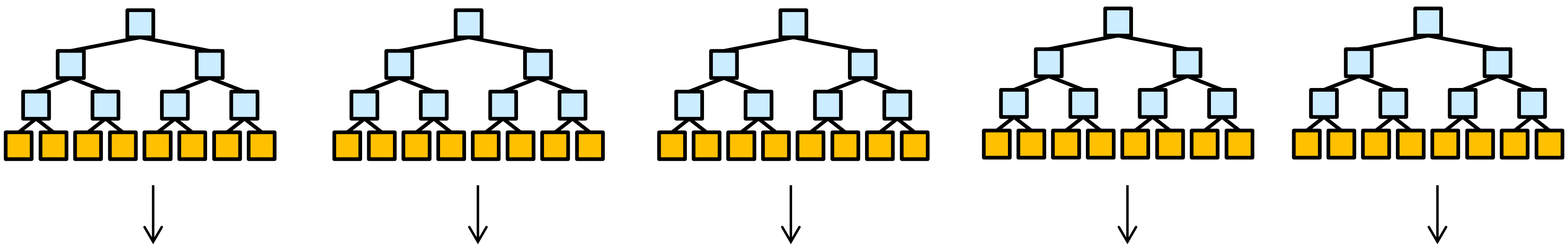
- We specify depth, e.g. 3
- In the inner nodes we test a feature (successful test – left, failed test – right)
- In the list we report: average value from samples in dataset which falls in that list

How to find a good tree?

1. Generating splitting tests we generate randomly more features and select such a one which splits the samples to sets with significant size and lowest variation in each set
2. We generate more trees and select the lowest error of classification



Gradient boosting

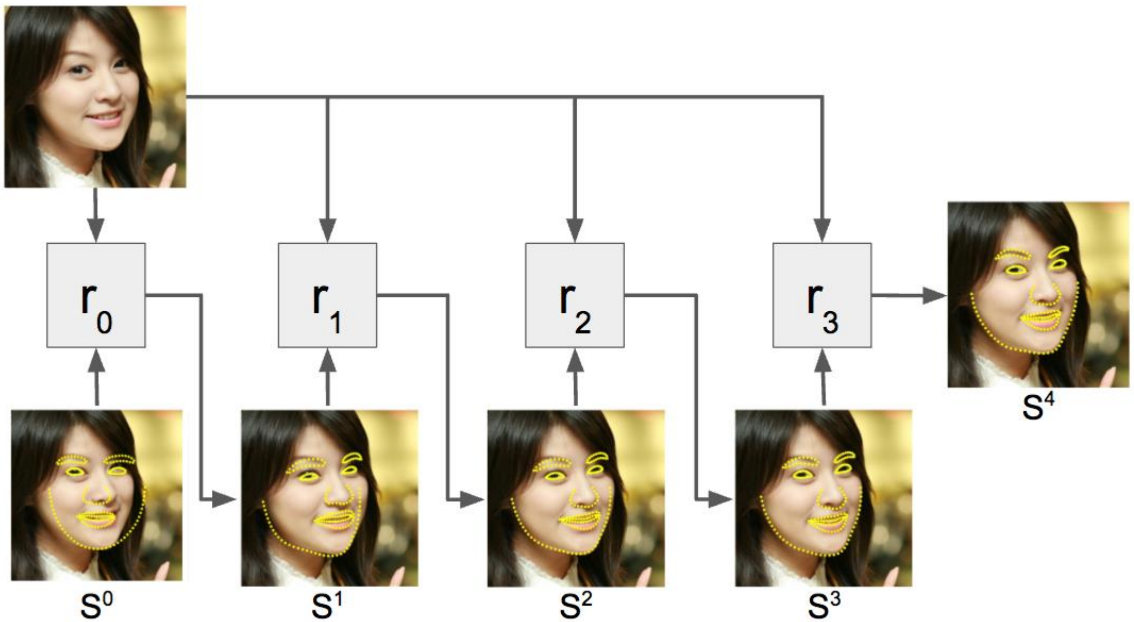


estimated value + estimated error + estimated error of estimated error + estimated ... + estimated ...

We can join more weak regressors by gradient boosting method

Cascade regressor

Kazemi regressor:



- 10 cascades
- 500 regression trees
- features based on comparison of two pixels from 400 pixels selected from image 128x128
- 20 trials for choosing features