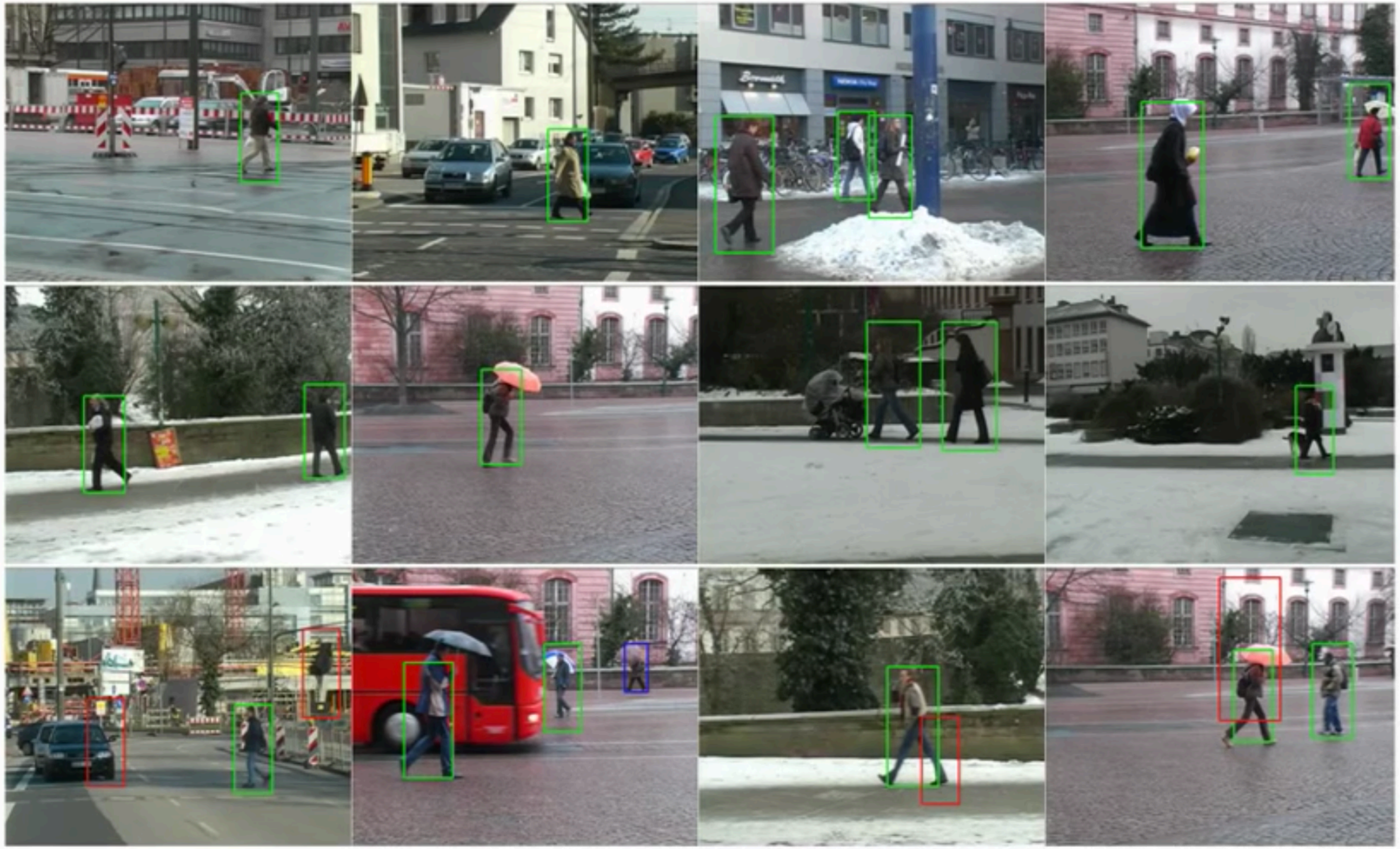
Random Forest

Fabricio Vargas Matos

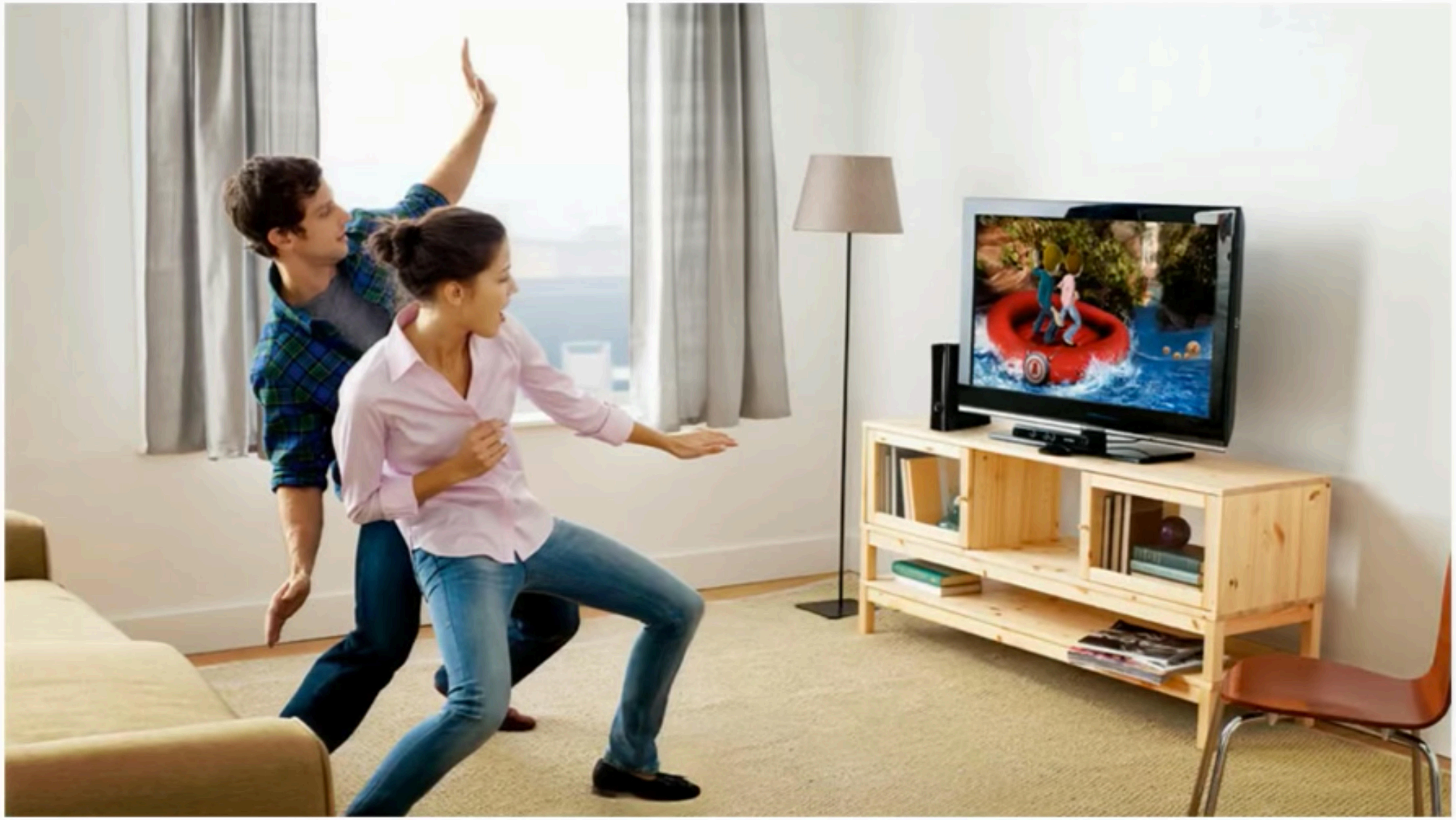
Motivation example 1: object detection



CPSC 540
Nando De Freitas

Feb 12 2013
12:48

Motivation example 2: Kinect

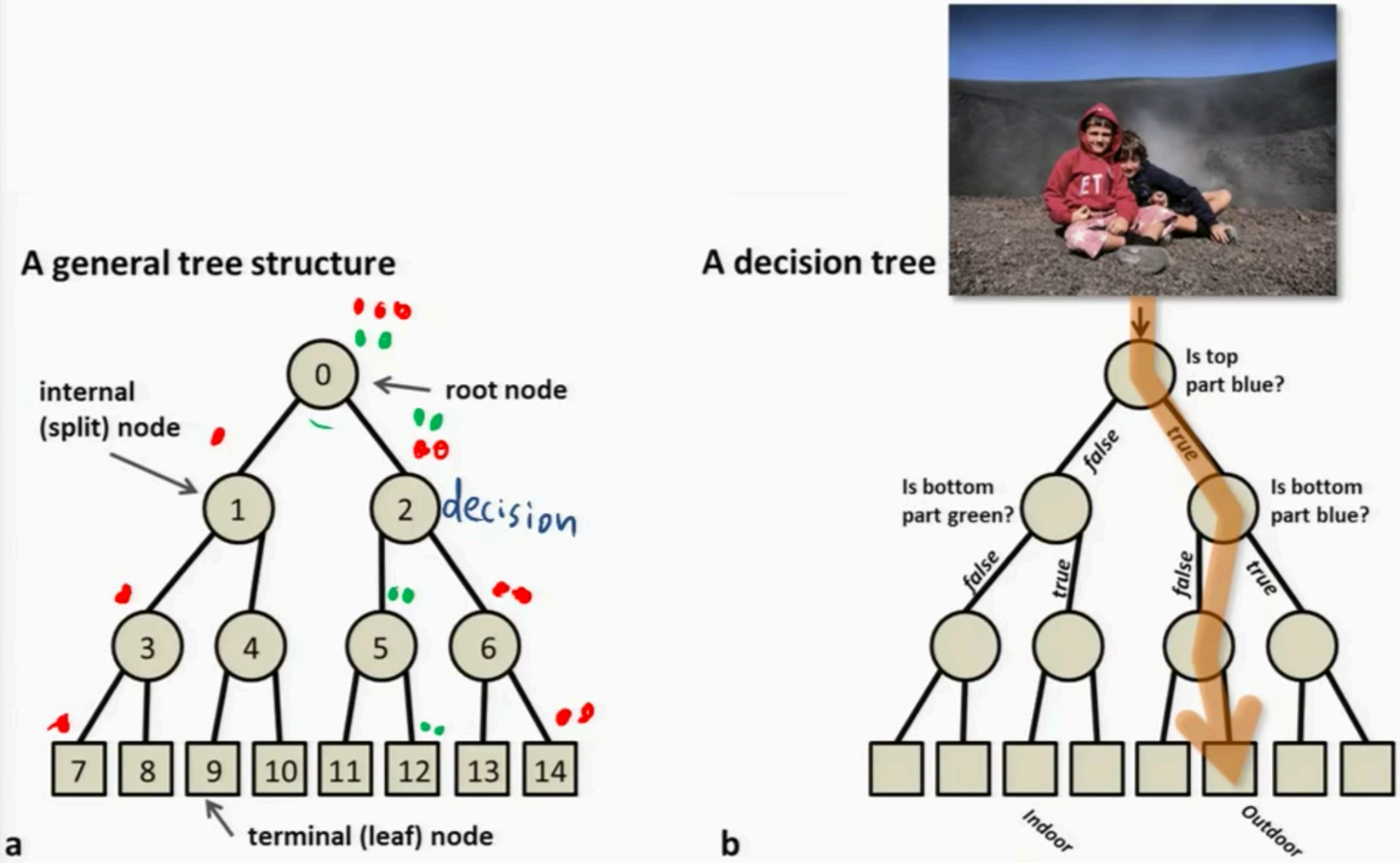


Computer Science

CPSC 540
Nando De Freitas

Feb 12 2013
12:49

Image classification example



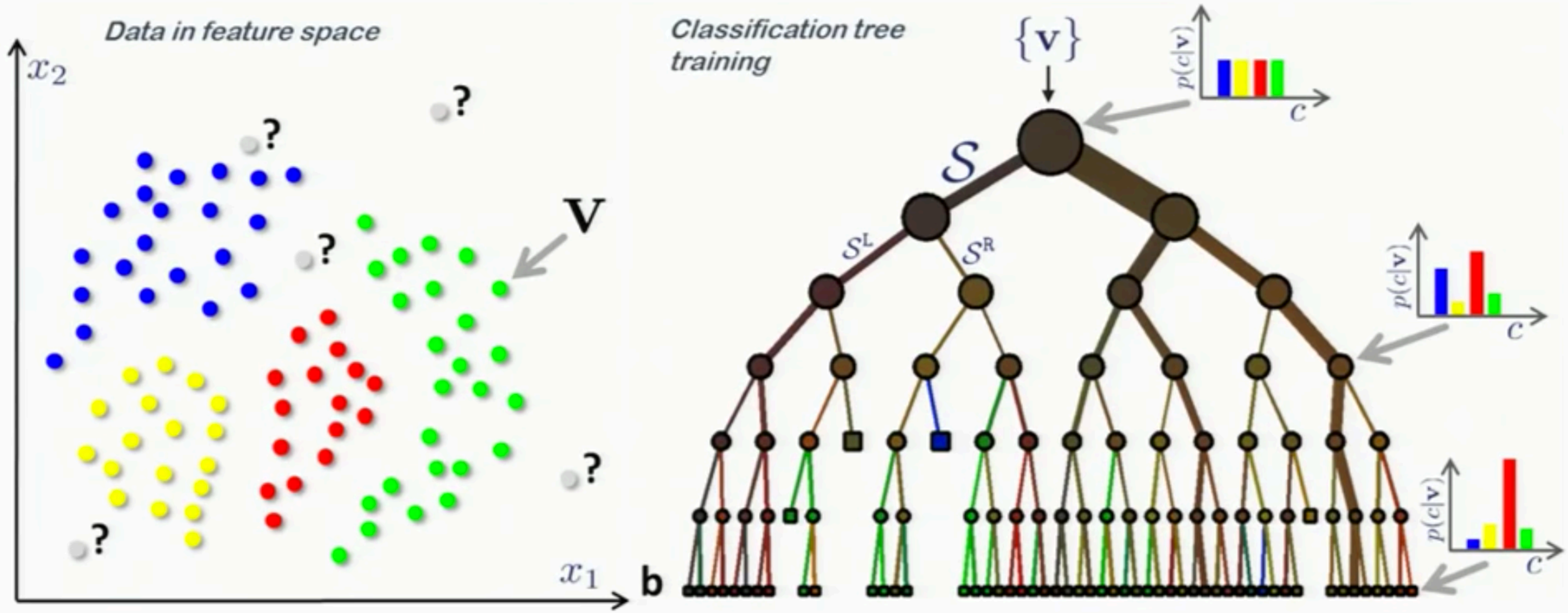
[MSR Tutorial on decision forests by Criminisi et al. 2011]

Computer Science

CPSC 540
Nando De Freitas

Feb 12 2013
12:51

Classification tree



A generic data point is denoted by a vector $\mathbf{v} = (x_1, x_2, \cdots, x_d)$

$$\mathcal{S}_j = \mathcal{S}_j^{\mathrm{L}} \cup \mathcal{S}_j^{\mathrm{R}}$$

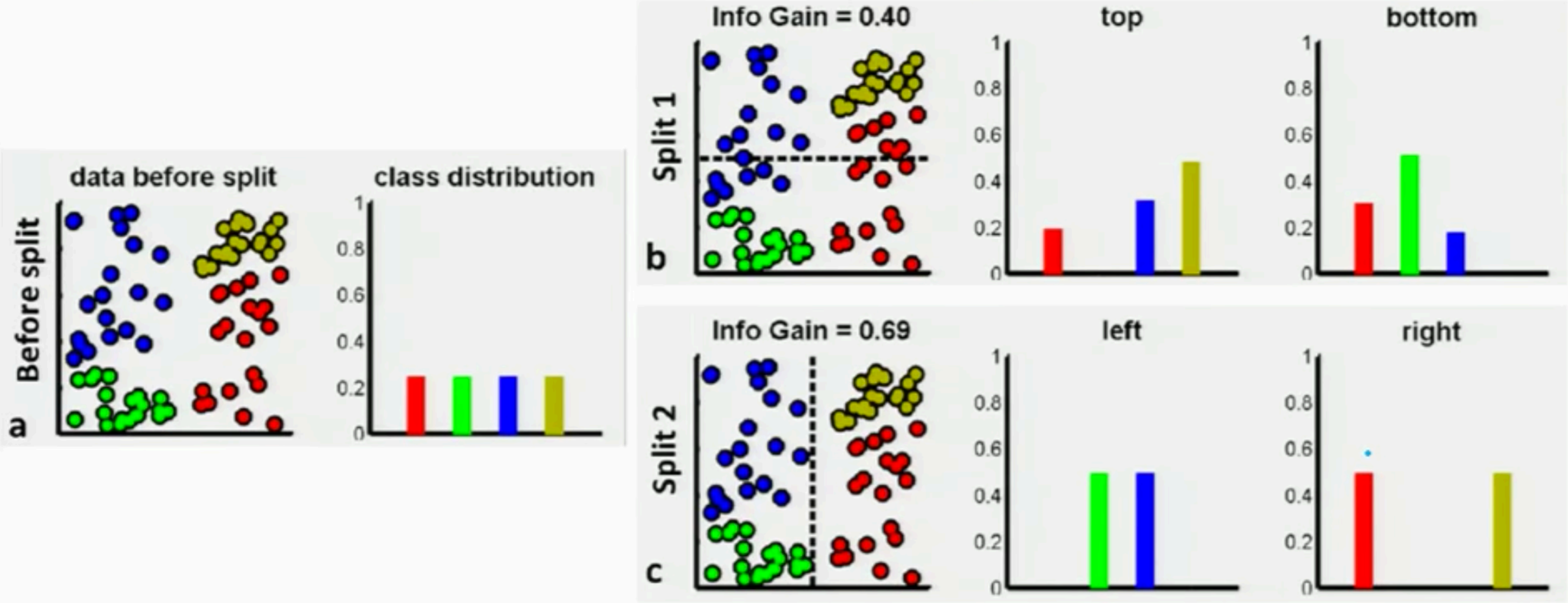
[Criminisi et al, 2011]

Computer Science

CPSC 540
Nando De Freitas

Feb 14 2013
12:38

Use information gain to decide splits



$$I_j = H(\mathcal{S}_j) - \sum_{i \in \{L,R\}} \frac{|\mathcal{S}_j^i|}{|\mathcal{S}_j|} H(\mathcal{S}_j^i)$$

[Criminisi et al, 2011]

Computer Science

CPSC 540
Nando De Freitas

Feb 14 2013
12:43

Random Forests algorithm

1. For $b = 1$ to B:

   (a) Draw a bootstrap sample $\mathbf{Z}^*$ of size N from the training data.

   (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.

      i. Select m variables at random from the p variables.

      ii. Pick the best variable/split-point among the m.

      iii. Split the node into two daughter nodes.

2. Output the ensemble of trees $\{T_b\}_1^B$.

[From the book of Hastie, Friedman and Tibshirani]

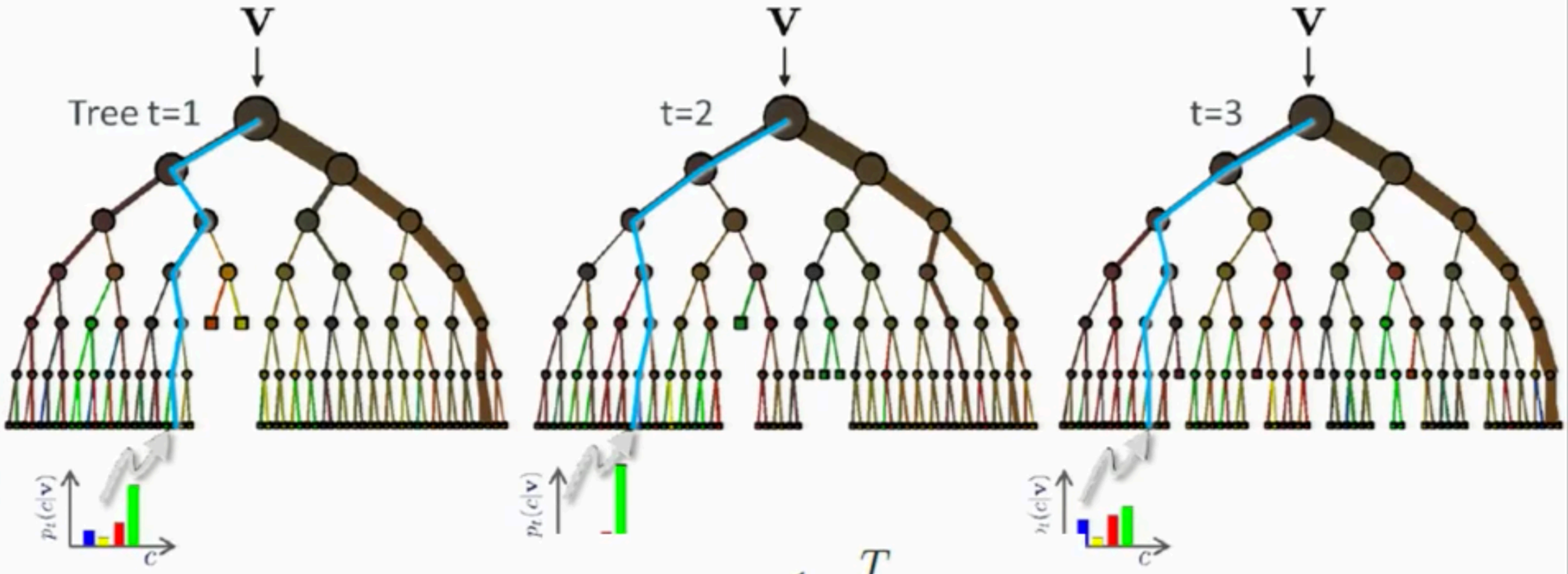
Computer Science

CPSC 540
Nando De Freitas

Feb 14 2013
12:58

Building a forest (ensemble)

In a forest with T trees we have $t \in \{1, \cdots, T\}$. All trees are trained independently (and possibly in parallel). During testing, each test point $\mathbf{v}$ is simultaneously pushed through all trees (starting at the root) until it reaches the corresponding leaves.



$$p(c|\mathbf{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\mathbf{v})$$

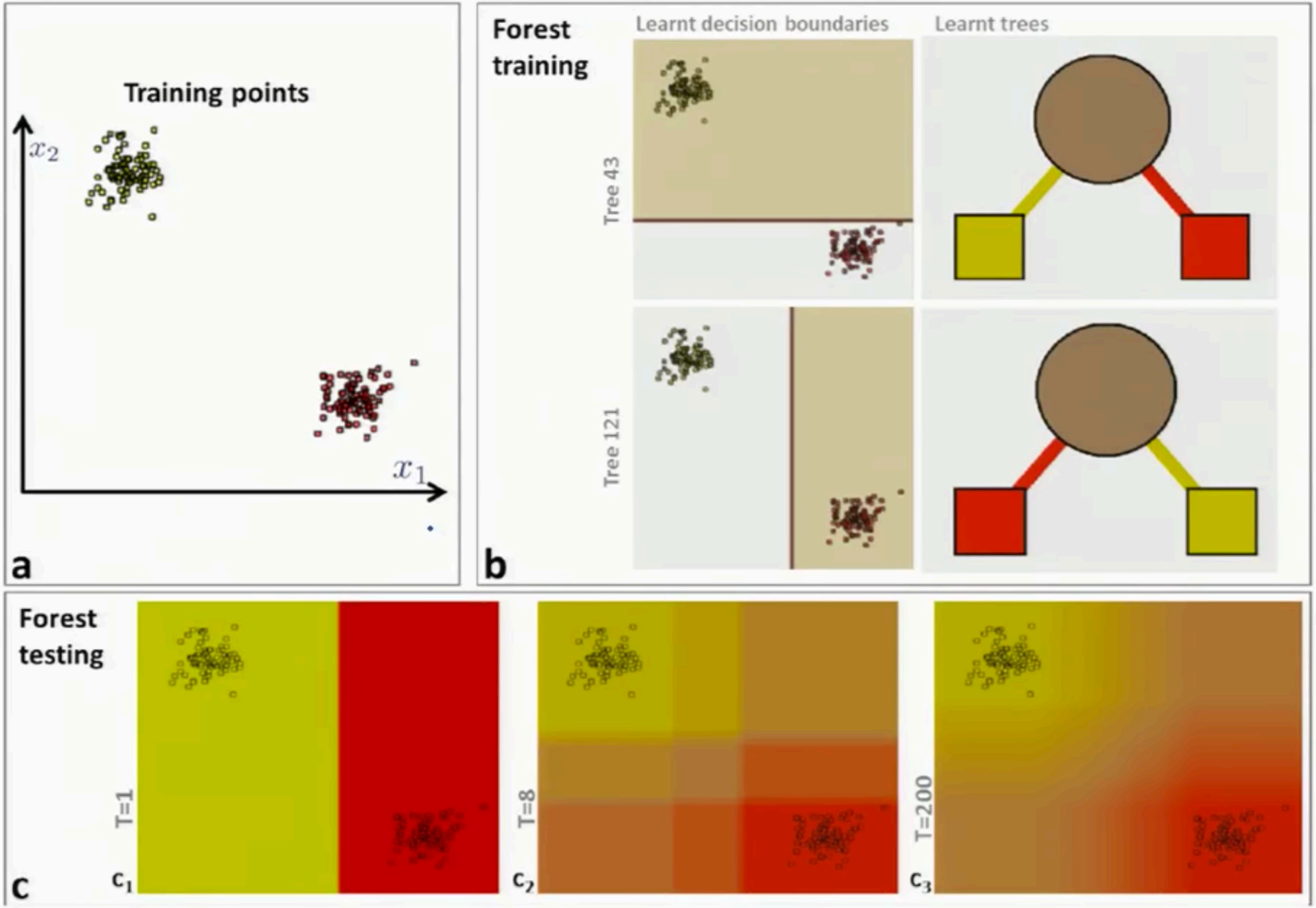
[Criminisi et al, 2011]

Computer Science

CPSC 540
Nando De Freitas

Feb 14 2013
13:06

Effect of forest size

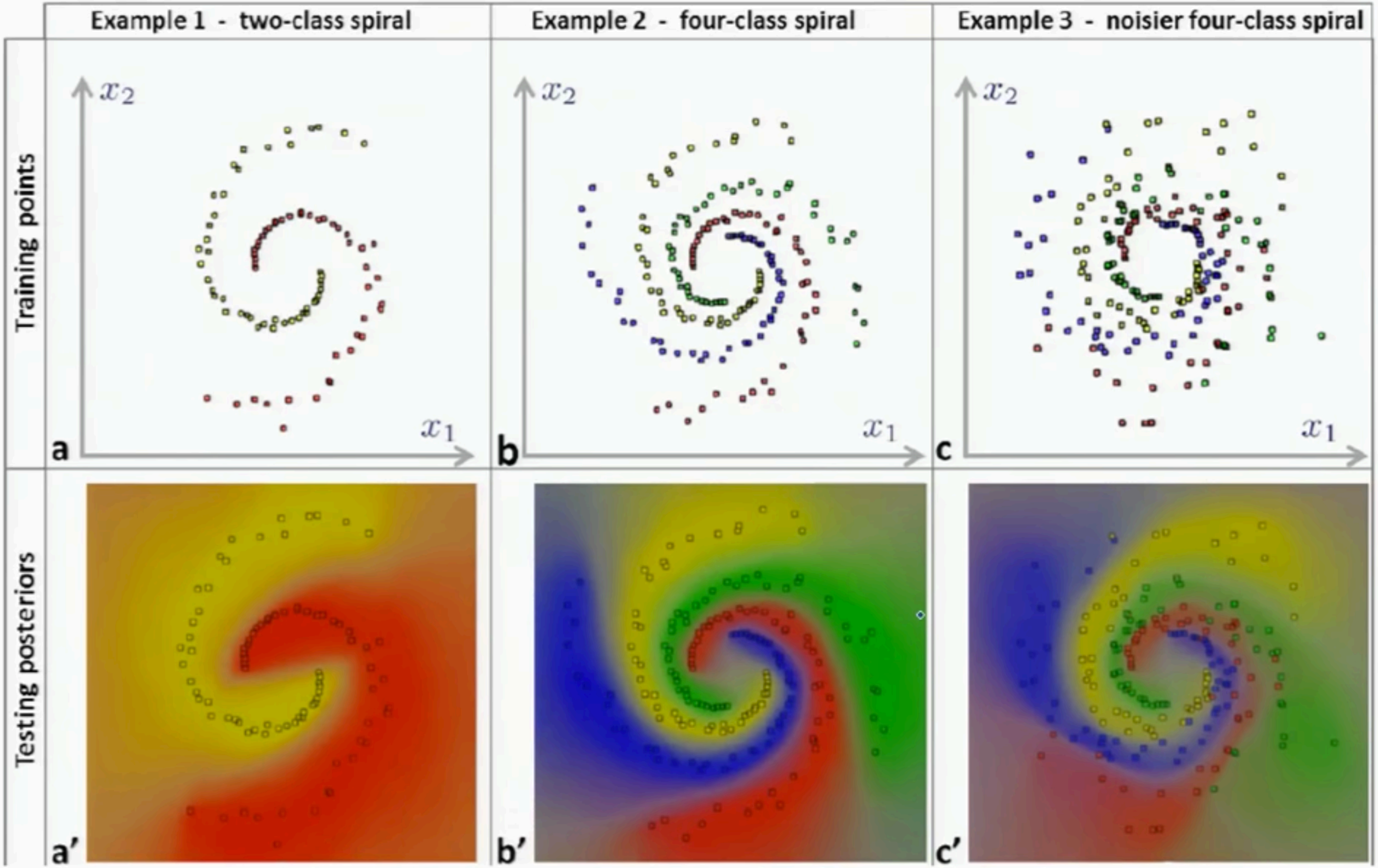


Computer Science

CPSC 540
Nando De Freitas

Feb 14 2013
13:39

Effect of more classes and noise

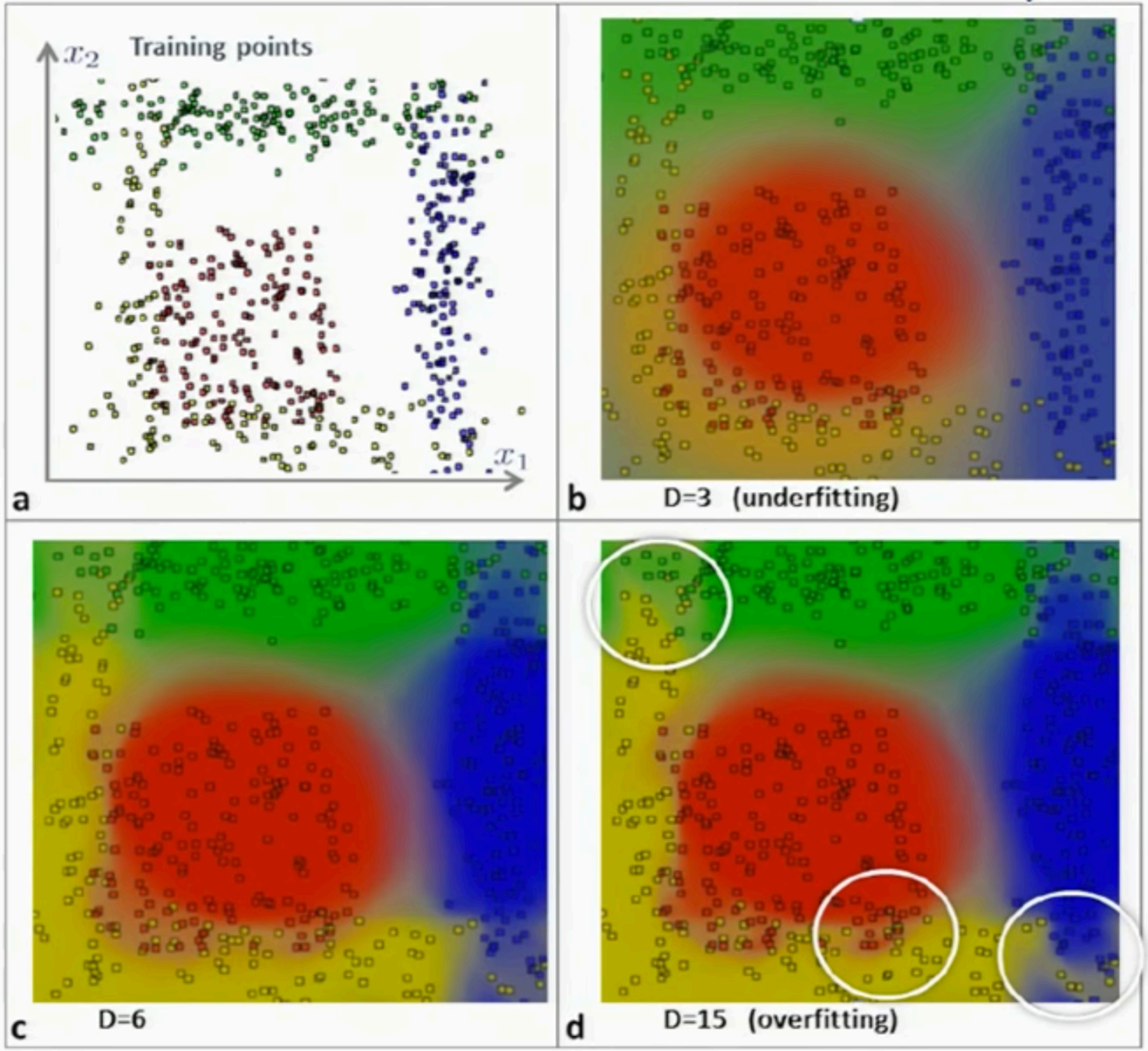


[Criminisi et al, 2011]

CPSC 540
Nando De Freitas

Feb 14 2013
13:41

Effect of tree depth (D)



[Criminisi et al, 2011]

CPSC 540
Nando De Freitas

Feb 14 2013
13:42

Application to face detection

Training Data

- 5000 faces
  - All frontal
- 300 million non faces
  - 9400 non-face images

[Viola and Jones, 2001]

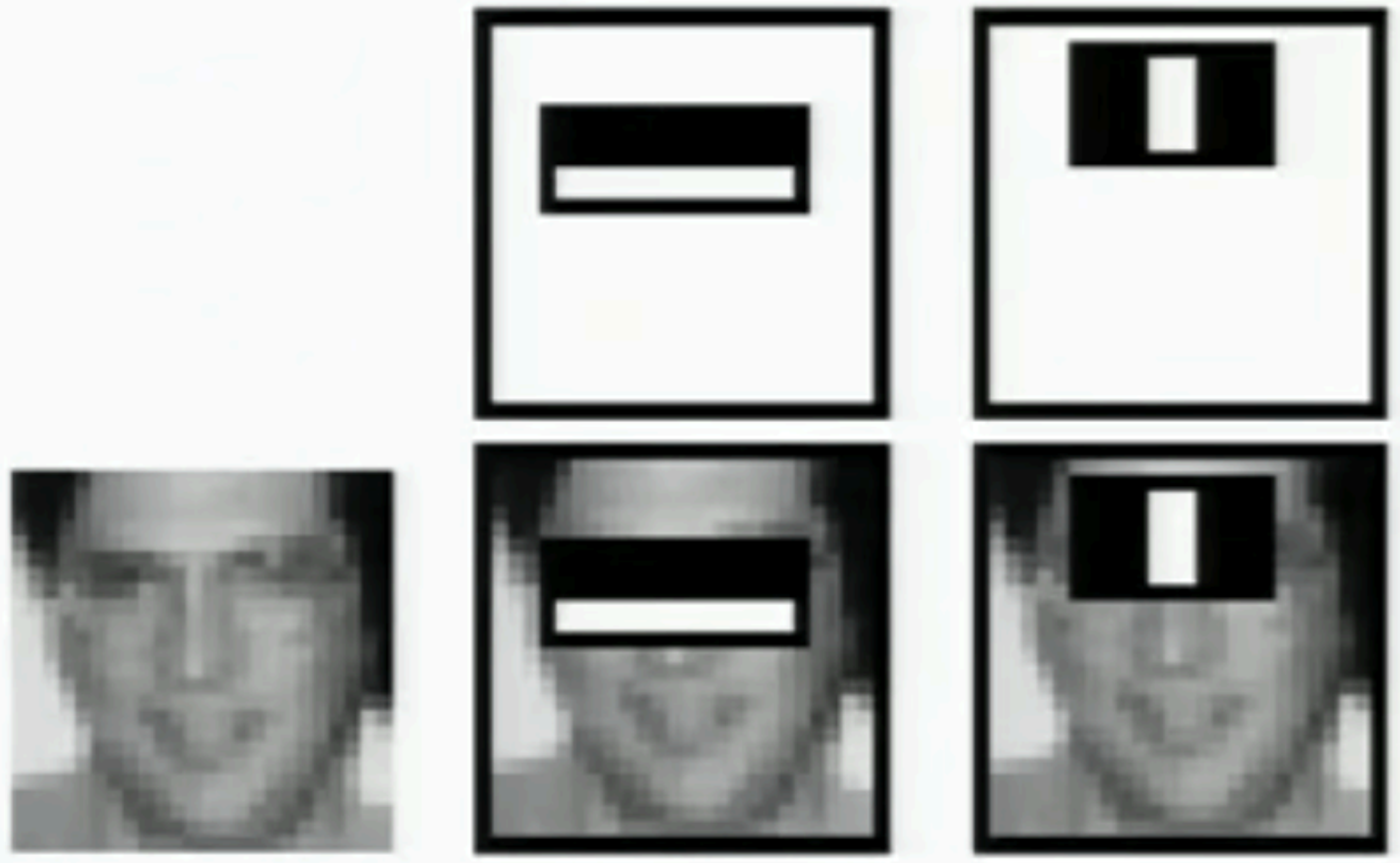
Computer Science

CPSC 540
Nando De Freitas

Feb 26 2013
12:40

Object detection

Idea: Extract simple features from all 24 by 24 pixel patches x_i. E.g., the value of a *two-rectangle feature is the difference between the sum* of the pixels within two rectangular regions. Then compare the level of activation (value of the feature f) with respect to a threshold (theta).



$$h_t(x_i) = \begin{cases} 1 & \text{if } f_t(x_i) > \theta_t \\ 0 & \text{otherwise} \end{cases}$$

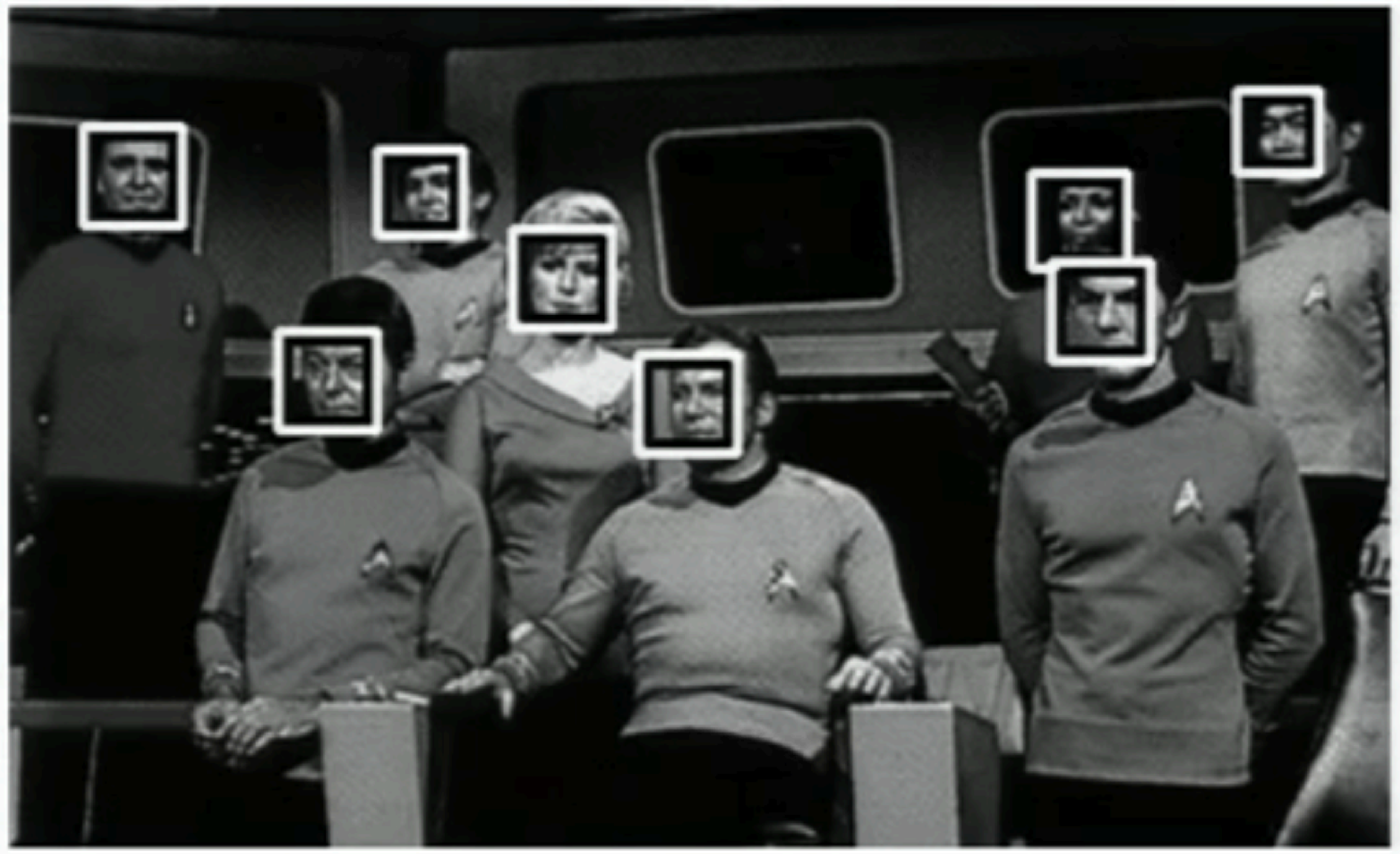
Relevant feature   Irrelevant feature

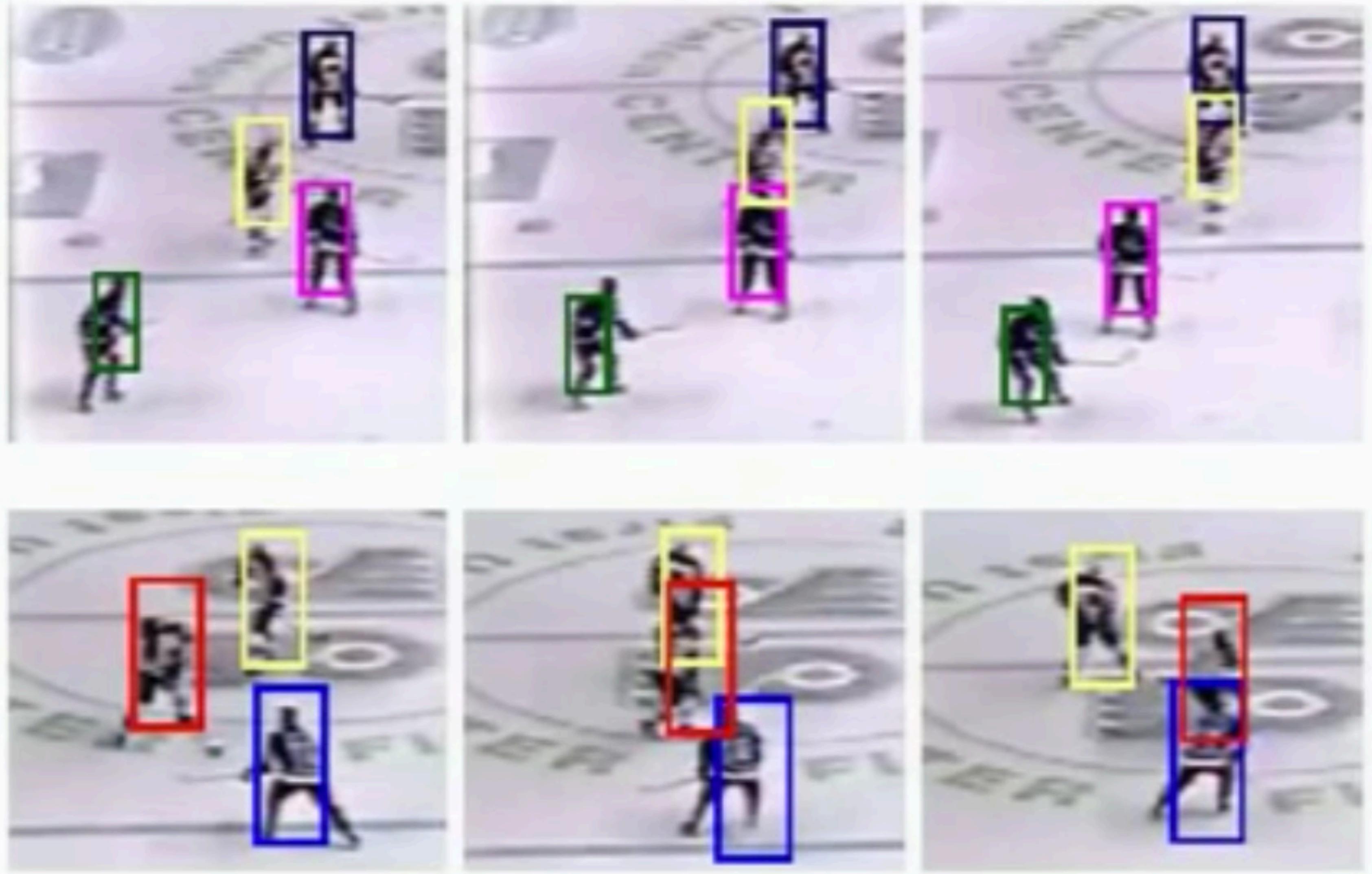
Computer Science

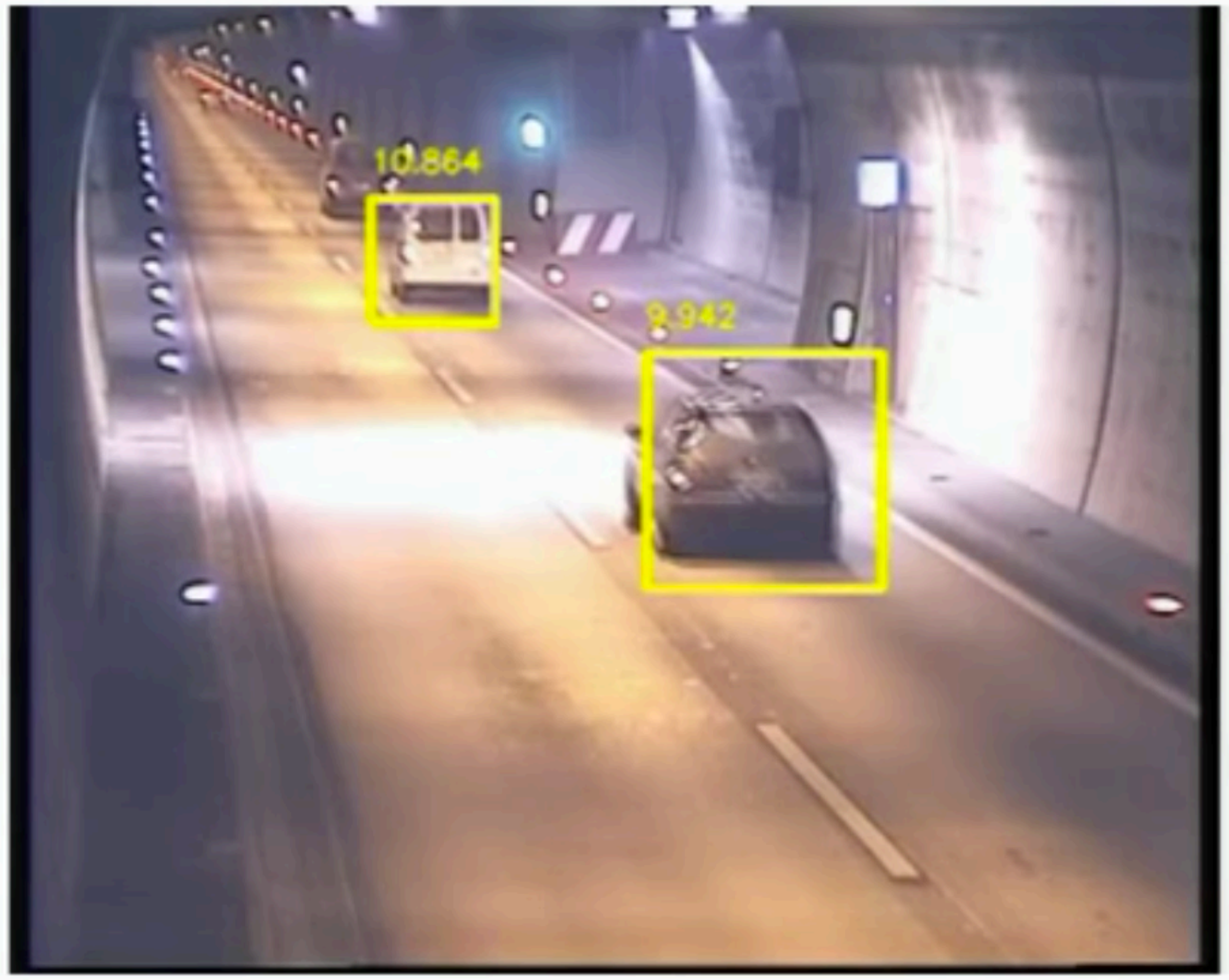
CPSC 540
Nando De Freitas

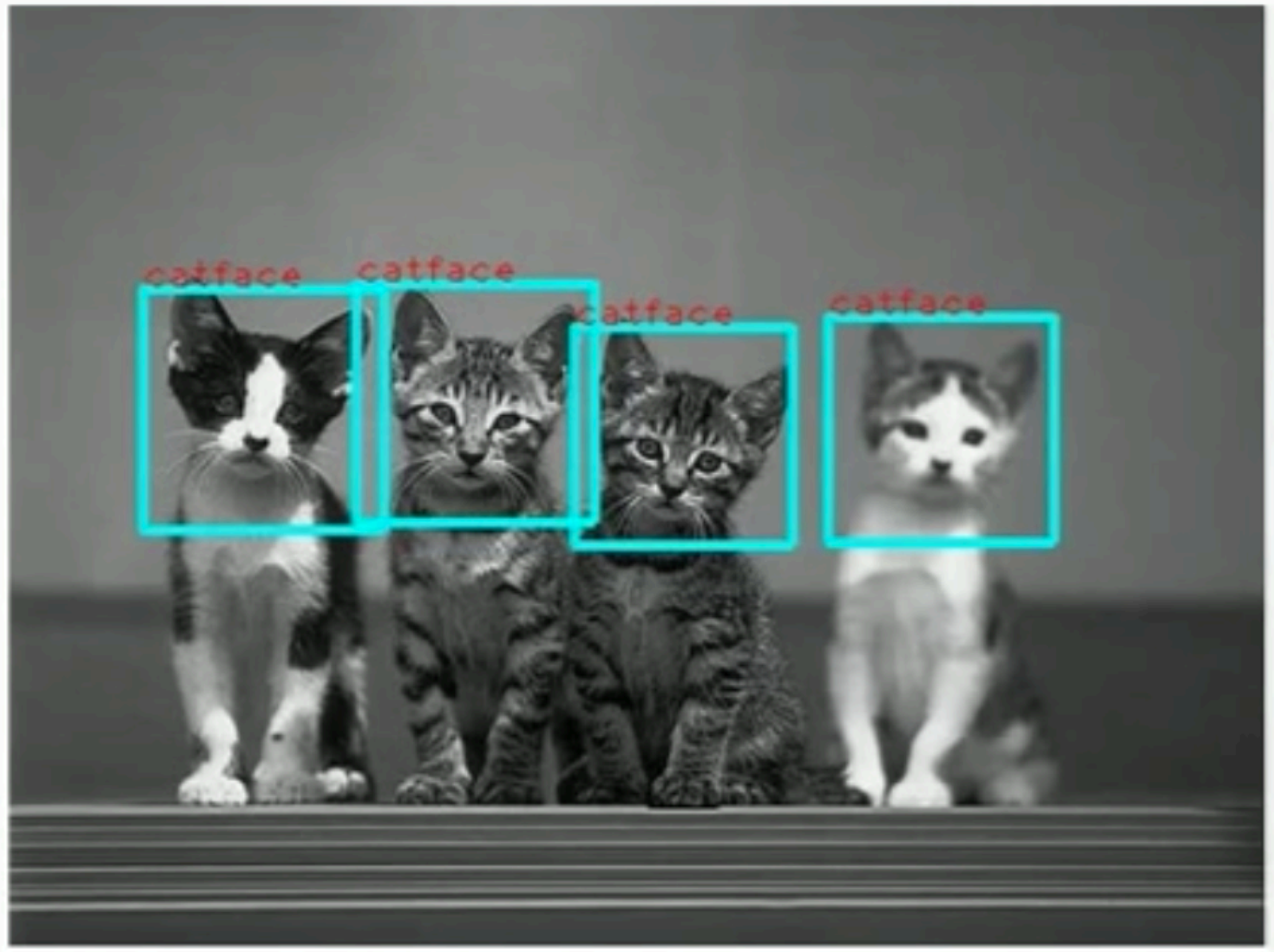
Feb 26 2013
12:42

Object detection







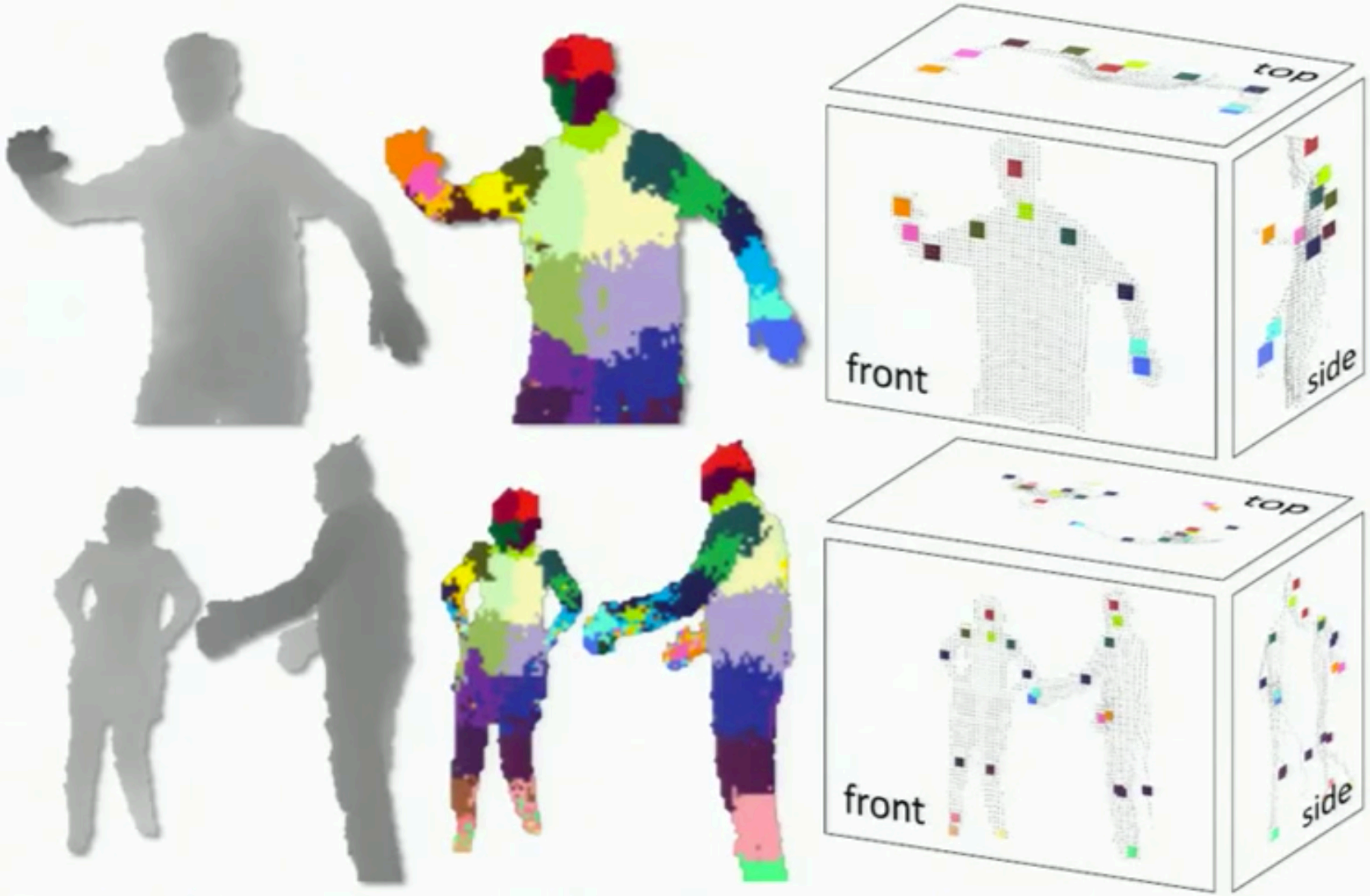


Computer Science

CPSC 540
Nando De Freitas

Feb 26 2013
12:47

Random Forests and the Kinect



depth image ➡ body parts ➡ 3D joint proposals

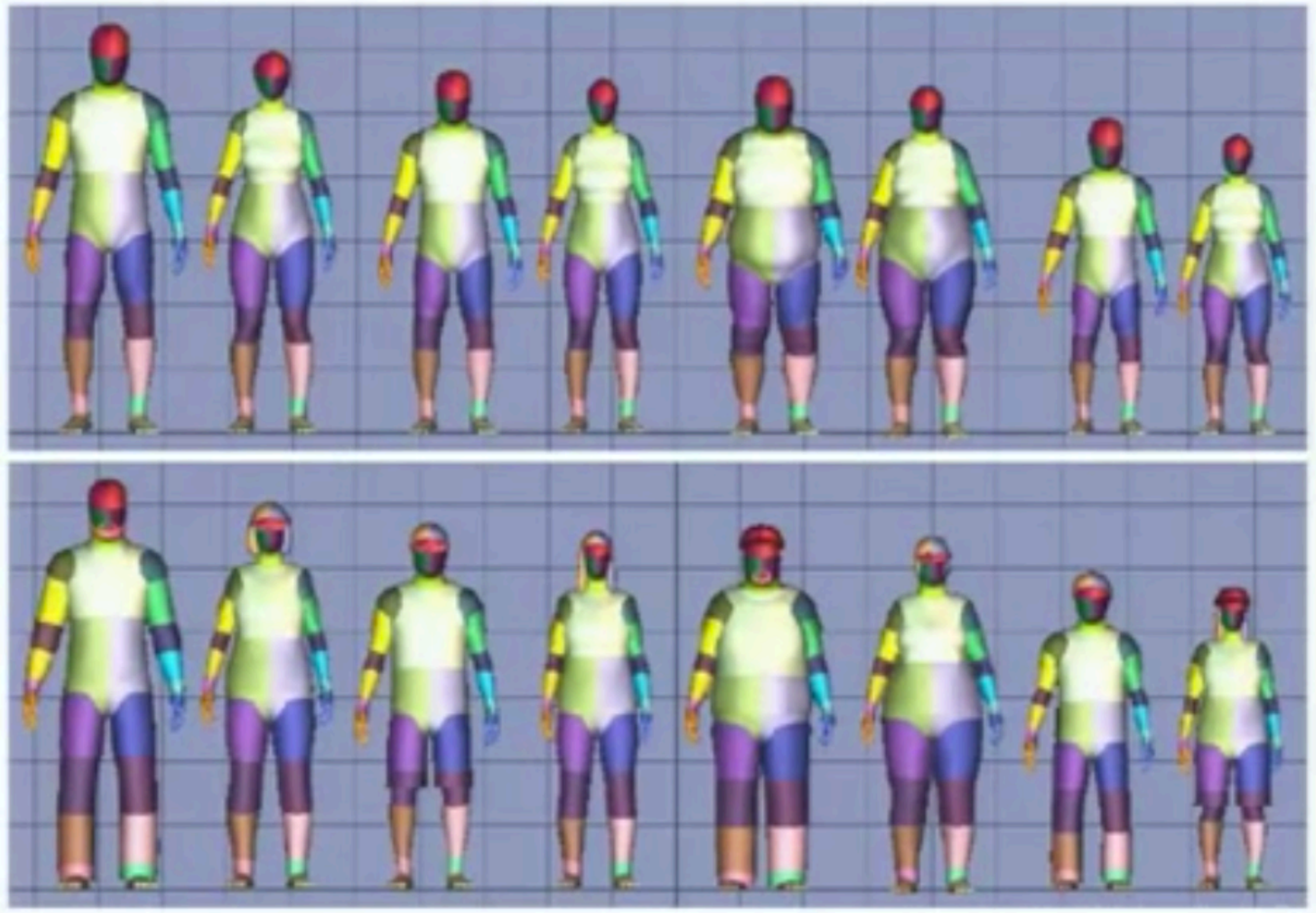
[Jamie Shotton et al 2011]

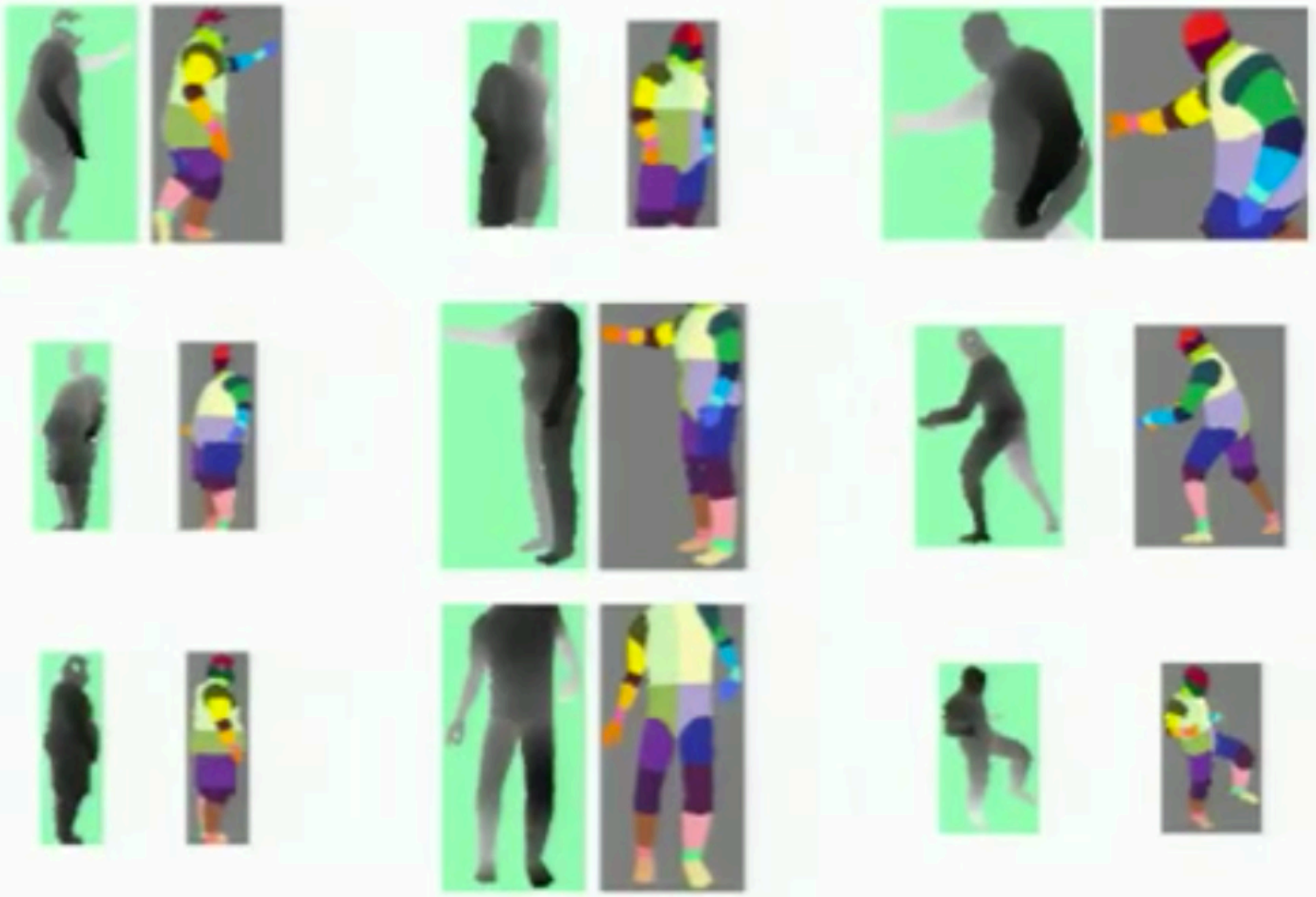
CPSC 540
Nando De Freitas

Feb 26 2013
12:48

Random Forests and the Kinect

Lesson 1: Use computer graphics to generate plenty of data.





[Jamie Shotton et al 2011]

Computer Science

CPSC 540
Nando De Freitas

Feb 26 2013
12:53

Performance on train and test data

[Jamie Shotton et al 2011]

Computer Science

CPSC 540
Nando De Freitas

Feb 26 2013
13:03

Referências:

Decision trees - https://www.youtube.com/watch?v=-dCtJjlEEgM
Random Forests - https://www.youtube.com/watch?v=3kYujfDgmNk
Random Forests Applications - https://www.youtube.com/watch?v=zFGPjRPwyFw