The Viola/Jones Face Detector

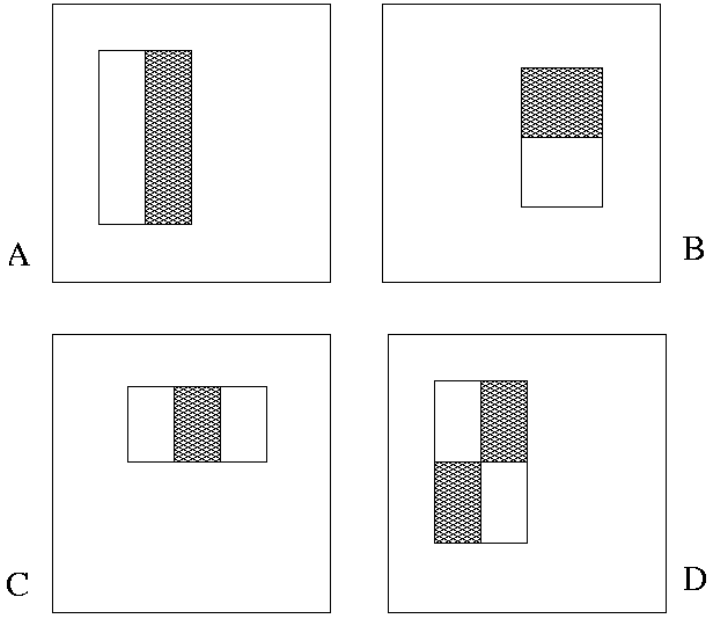
- A “paradigmatic” method for real-time object detection
- Training is slow, but detection is very fast
- Key ideas
  - *Integral images* for fast feature evaluation
  - *Boosting* for feature selection
  - *Attentional cascade* for fast rejection of non-face windows

P. Viola and M. Jones. *Rapid object detection using a boosted cascade of simple features*. CVPR 2001.

Slides by Robert Fergus

Image Features

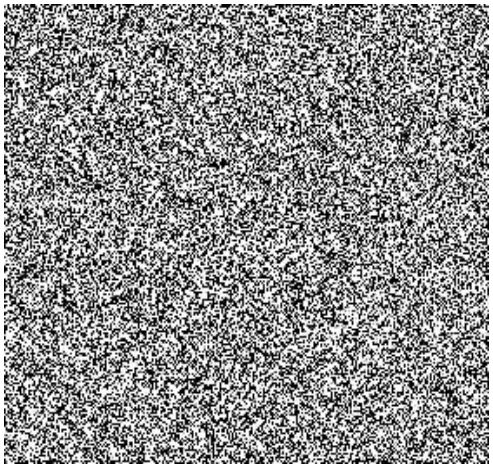
"Rectangle filters"



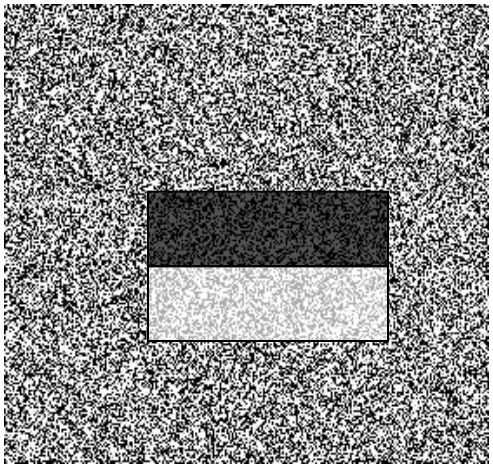
Value =

∑ (pixels in white area) –
∑ (pixels in black area)

Example



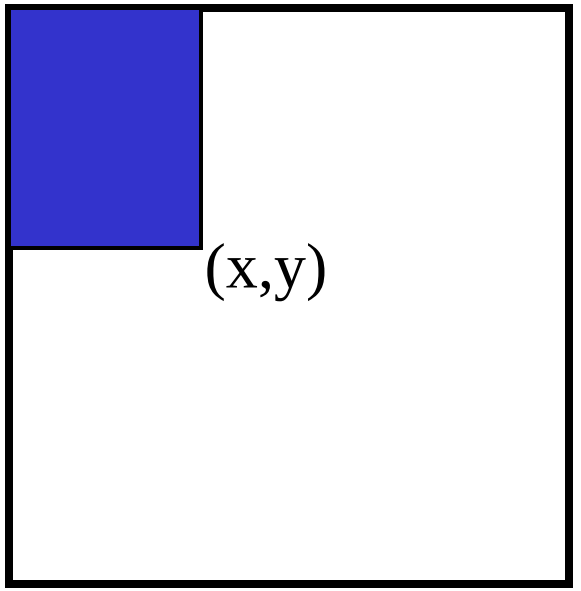
Source



Result

Fast computation with integral images

- The *integral image* computes a value at each pixel (*x*,*y*) that is the sum of the pixel values above and to the left of (*x*,*y*), inclusive
- This can quickly be computed in one pass through the image

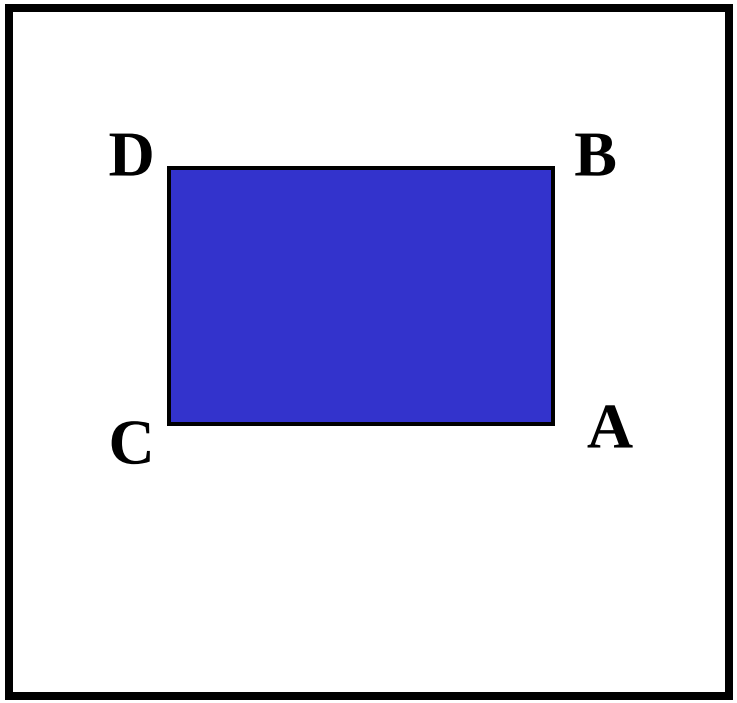


Computing sum within a rectangle

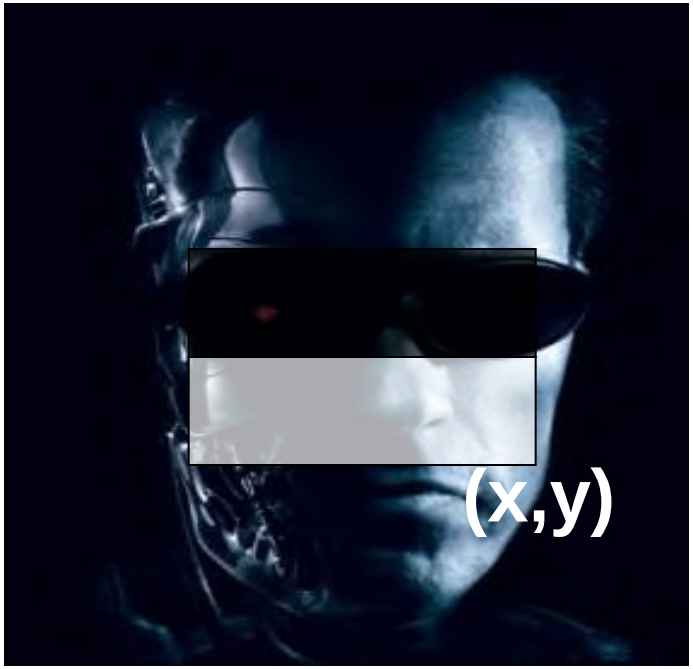
- Let A,B,C,D be the values of the integral image at the corners of a rectangle
- Then the sum of original image values within the rectangle can be computed as:

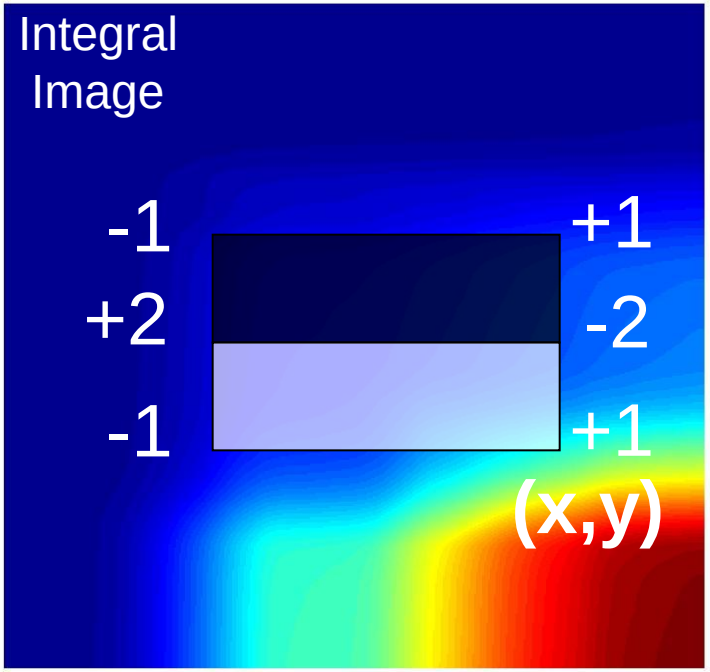
  sum = A – B – C + D

- Only 3 additions are required for any size of rectangle!
  - This is now used in many areas of computer vision



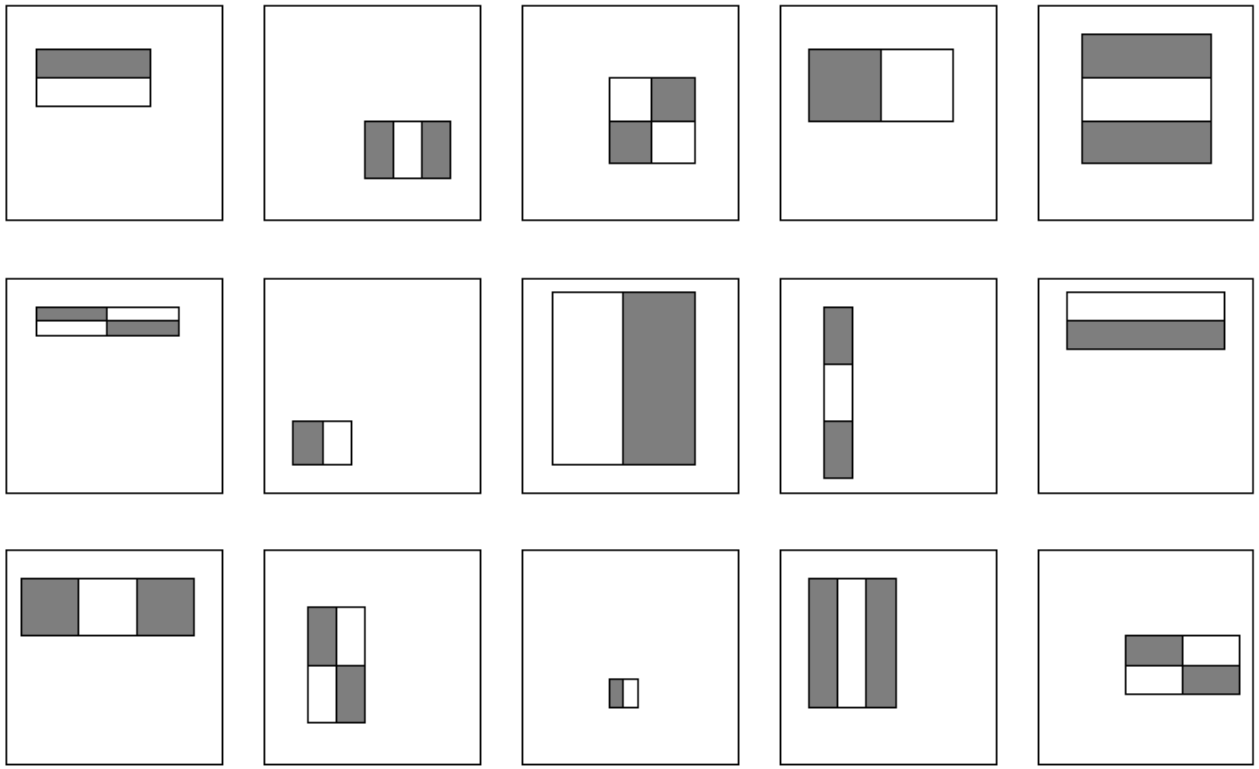
Example





Feature selection

- For a 24x24 detection region, the number of possible rectangle features is ~180,000!



Feature selection

- For a 24x24 detection region, the number of possible rectangle features is ~180,000!
- At test time, it is impractical to evaluate the entire feature set
- Can we create a good classifier using just a small subset of all possible features?
- How to select such a subset?

Boosting

- Boosting is a classification scheme that works by combining *weak learners* into a more accurate ensemble classifier
- *Weak learner*: classifier with accuracy that need be only better than chance
- We can define weak learners based on rectangle features:

Y. Freund and R. Schapire, A short introduction to boosting, *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780, September, 1999.

AdaBoost

- Given a set of weak classifiers originally : $h_j(\mathbf{x}) \in \{+1, -1\}$
  - None much better than random
- Iteratively combine classifiers
  - Form a linear combination

$$C(x) = \theta\left(\sum_t h_t(x) + b\right)$$

  - Training error converges to 0 quickly
  - Test error is related to training margin

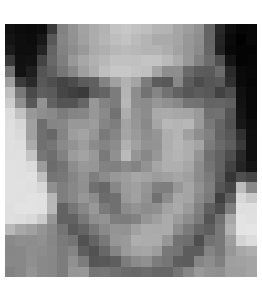
Y. Freund and R. Schapire, A short introduction to boosting, *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780, September, 1999.

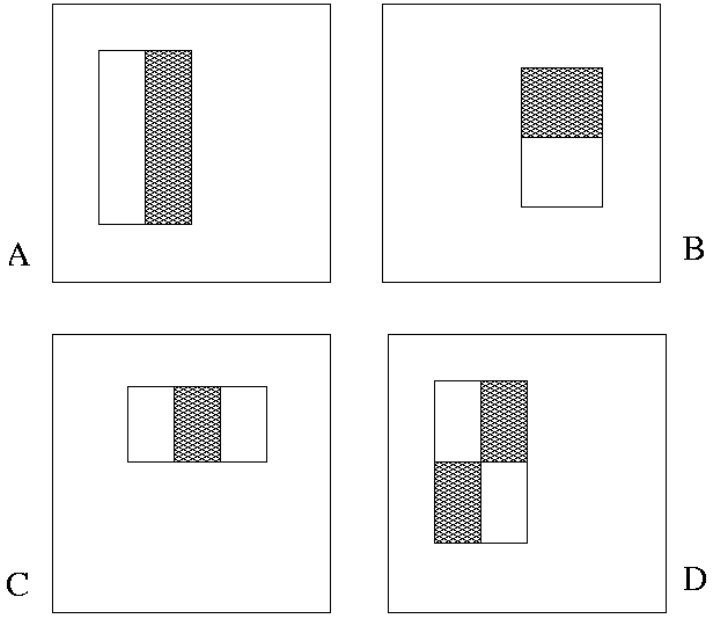
Boosted Face Detection: Image Features

"Rectangle filters"

Similar to Haar wavelets

Papageorgiou, et al.





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

$$C(x) = \theta\left(\sum_t h_t(x) + b\right)$$

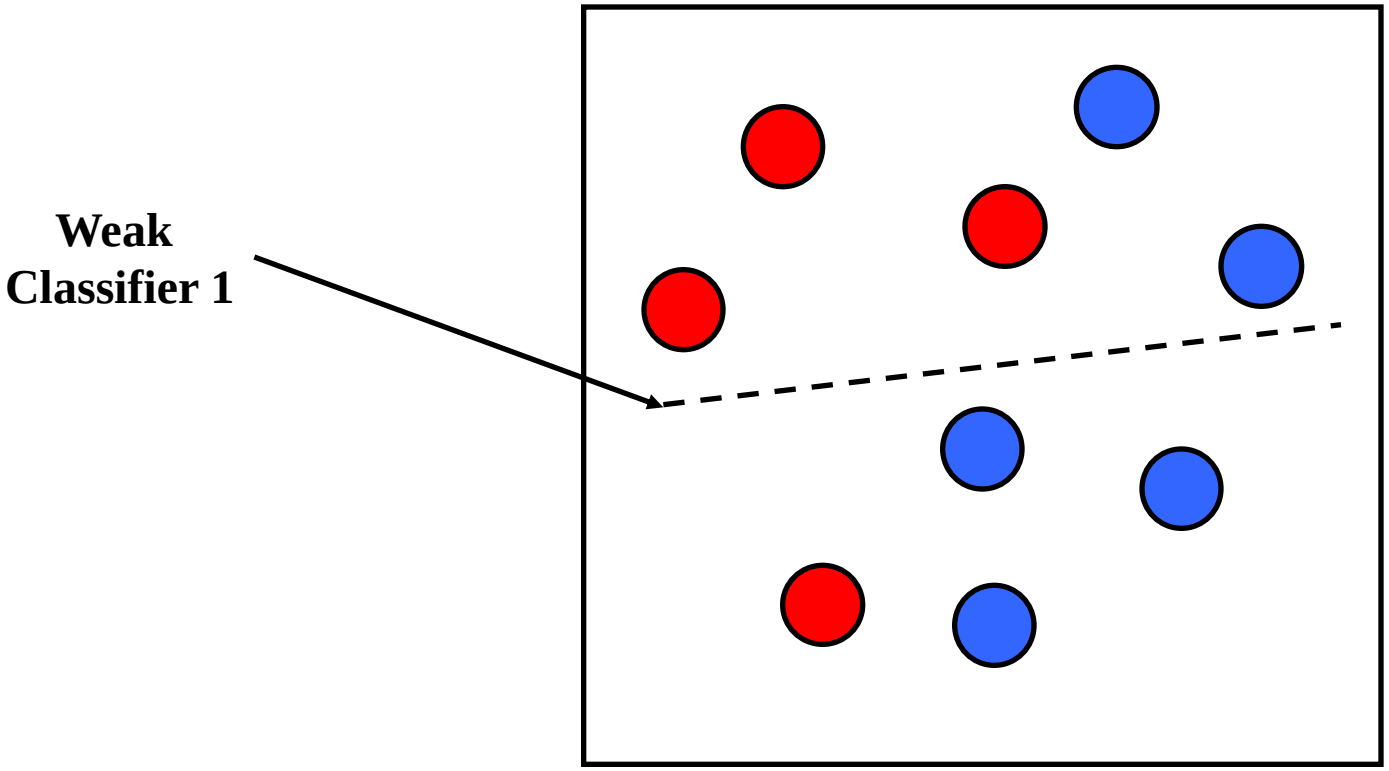
60,000 features to choose from

Boosting outline

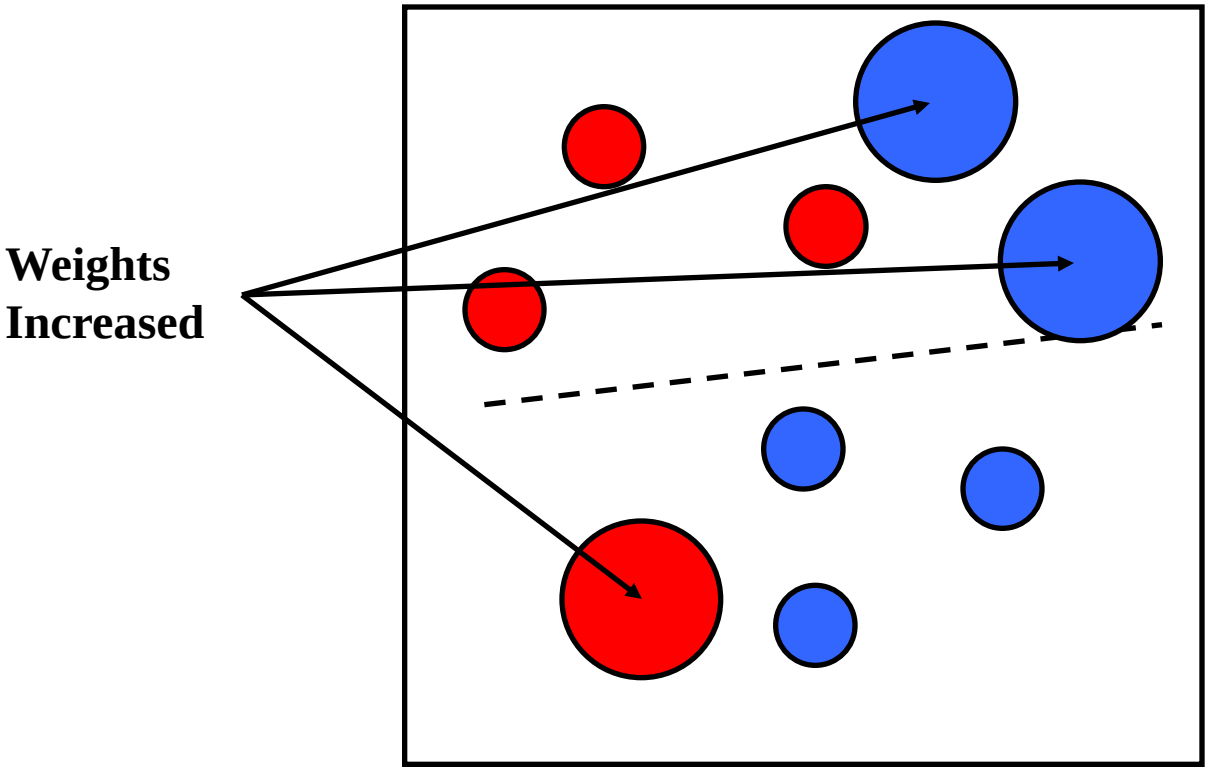
- Initially, give equal weight to each training example
- Iterative training procedure
  - Find best weak learner for current weighted training set
  - Raise the weights of training examples misclassified by current weak learner
- Compute final classifier as linear combination of all weak learners (weight of each learner is related to its accuracy)

Y. Freund and R. Schapire, A short introduction to boosting, *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780, September, 1999.

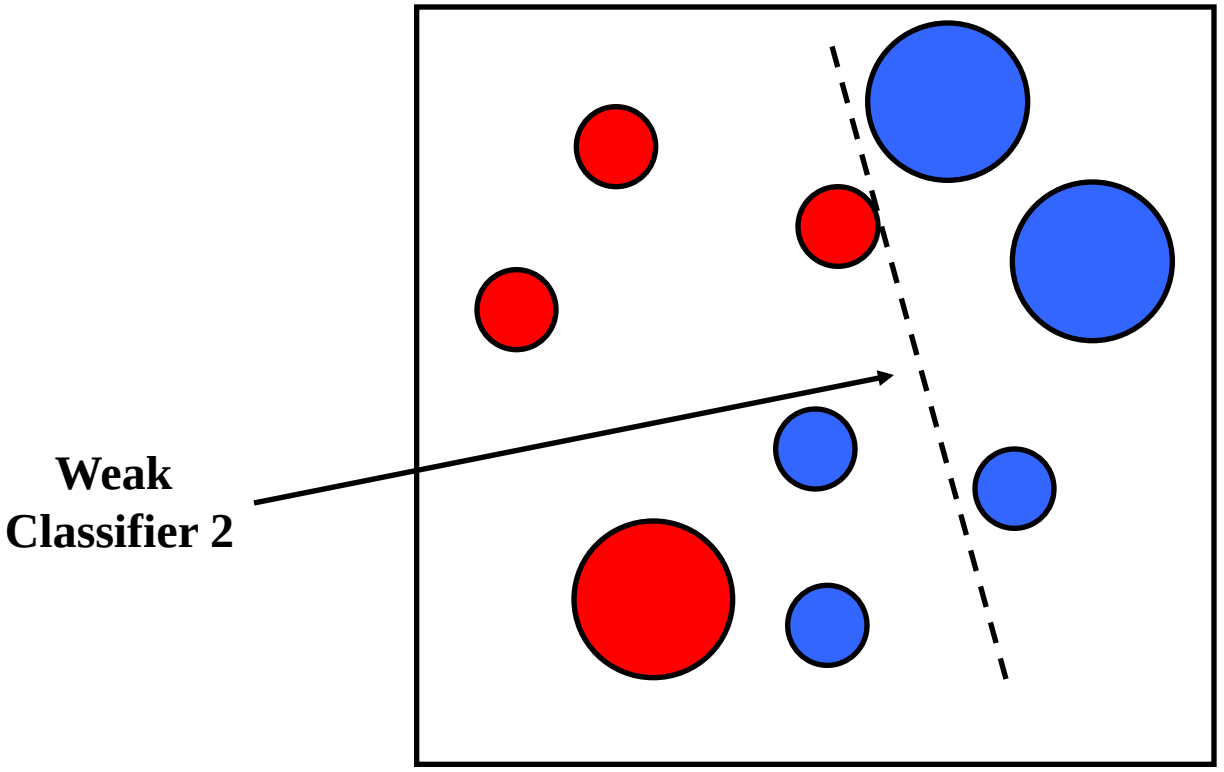
Boosting



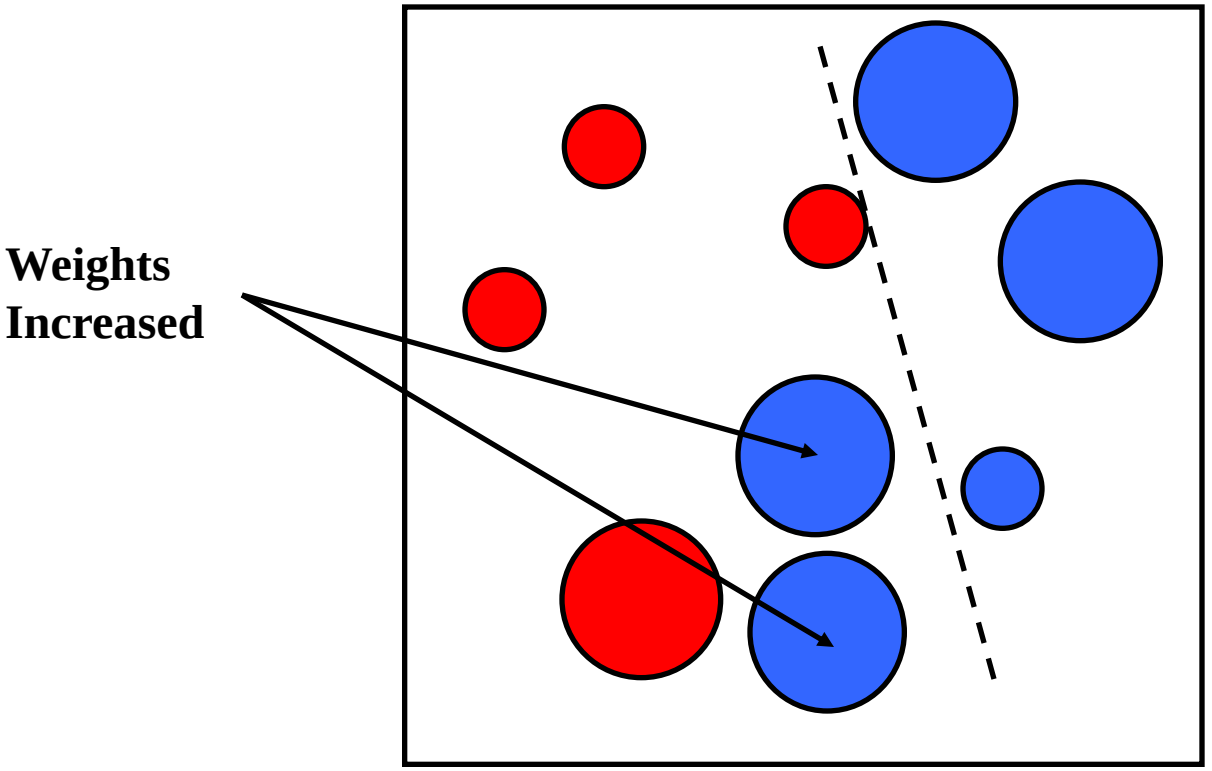
Boosting



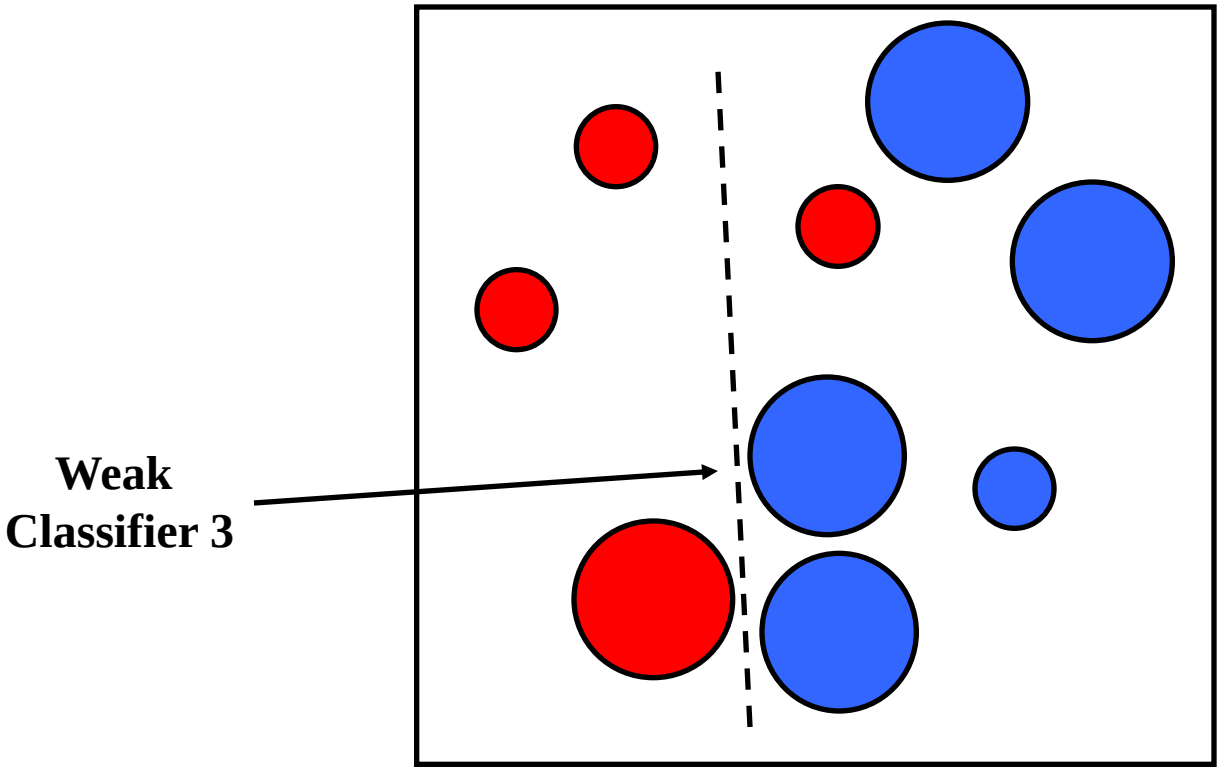
Boosting



Boosting

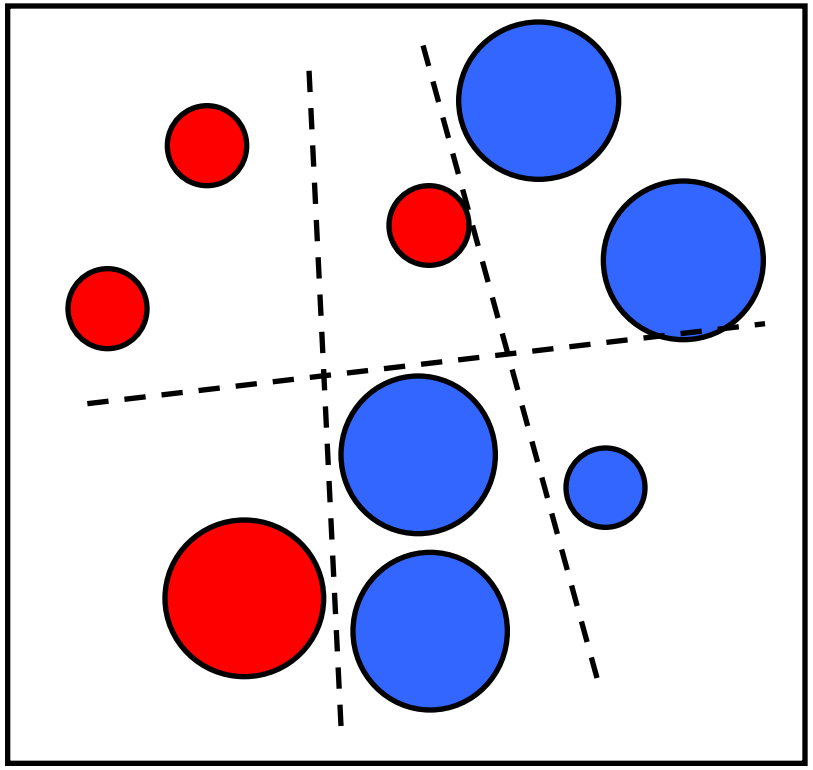


Boosting



Boosting

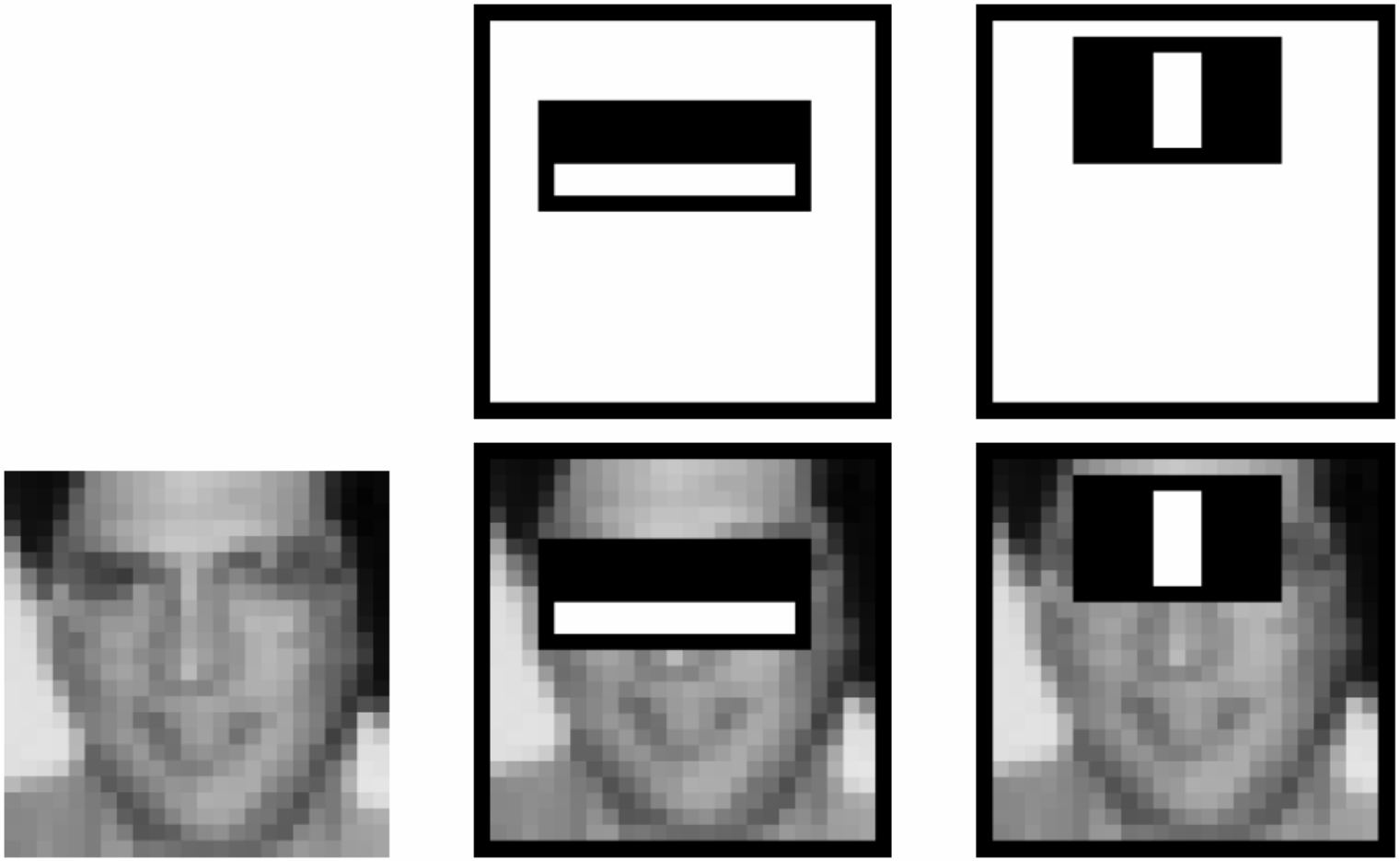
Final classifier is linear combination of weak classifiers



Boosting for face detection

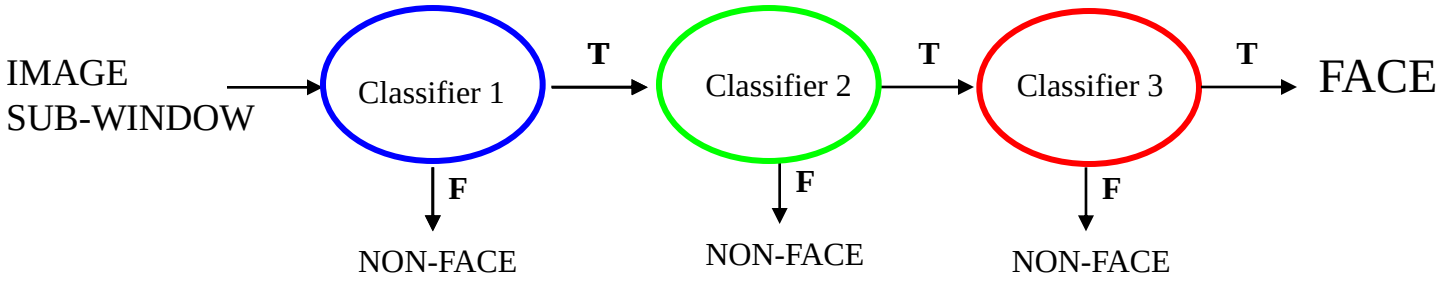
- For each round of boosting:
  - Evaluate each rectangle filter on each example
  - Select best threshold for each filter
  - Select best filter/threshold combination
  - Reweight examples
- Computational complexity of learning: $O(MNT)$
  - M filters, N examples, T thresholds

First two features selected by boosting



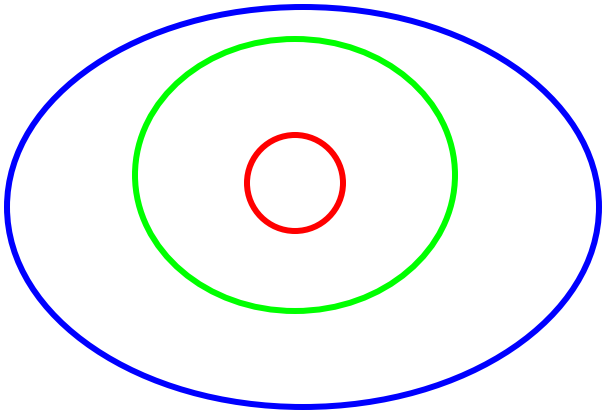
Cascading classifiers

- We start with simple classifiers which reject many of the negative sub-windows while detecting almost all positive sub-windows
- Positive results from the first classifier triggers the evaluation of a second (more complex) classifier, and so on
- A negative outcome at any point leads to the immediate rejection of the sub-window

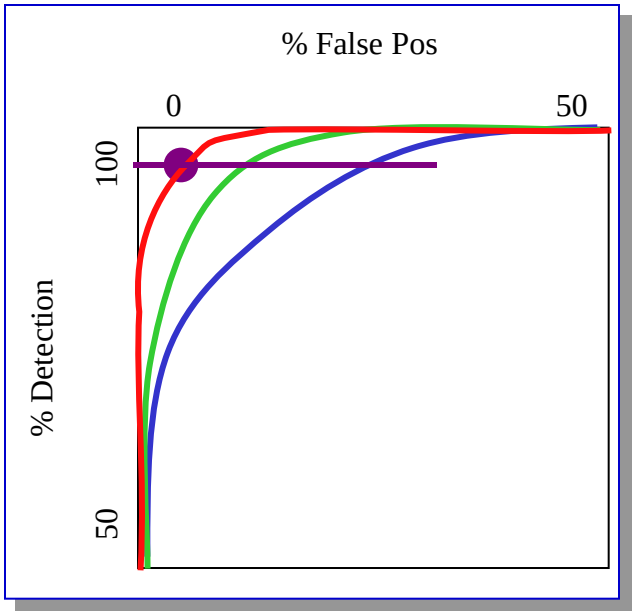


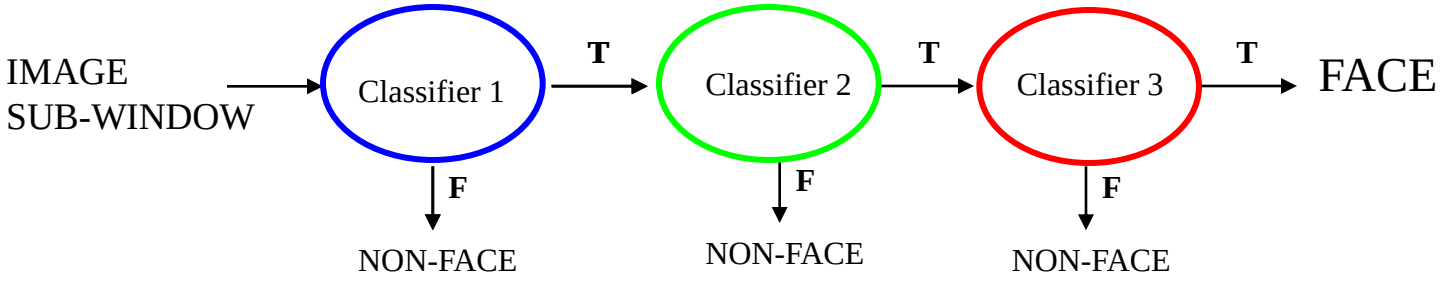
Cascading classifiers

- Chain classifiers that are progressively more complex and have lower false positive rates:



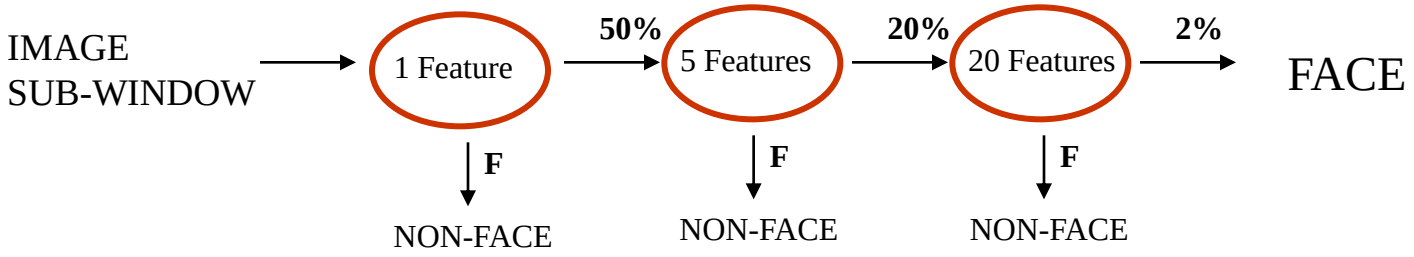
Receiver operating characteristic





Training the cascade

- Adjust weak learner threshold to minimize *false negatives* (as opposed to total classification error)
- Each classifier trained on false positives of previous stages
  - A single-feature classifier achieves 100% detection rate and about 50% false positive rate
  - A five-feature classifier achieves 100% detection rate and 40% false positive rate (20% cumulative)
  - A 20-feature classifier achieve 100% detection rate with 10% false positive rate (2% cumulative)



The implemented system

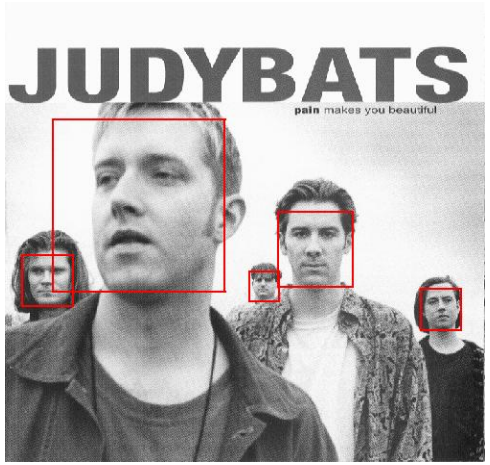
- Training Data
  - 5000 faces
    - All frontal, rescaled to 24x24 pixels
  - 300 million non-faces
    - 9500 non-face images
  - Faces are normalized
    - Scale, translation
- Many variations
  - Across individuals
  - Illumination
  - Pose

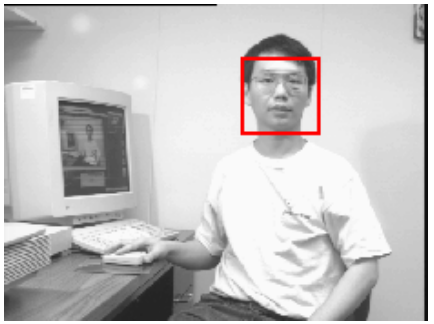
(Most slides from Paul Viola)

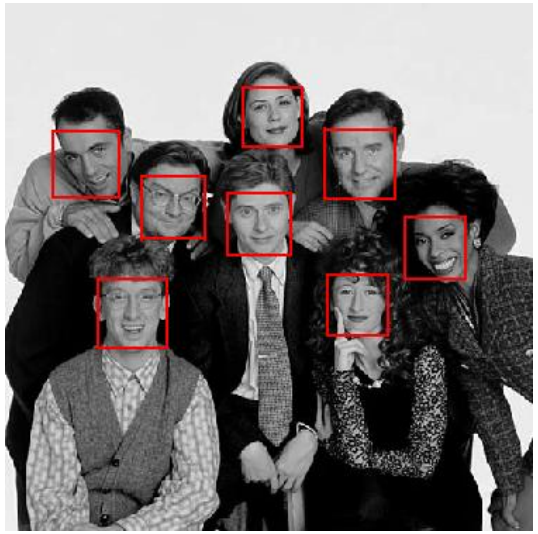
System performance

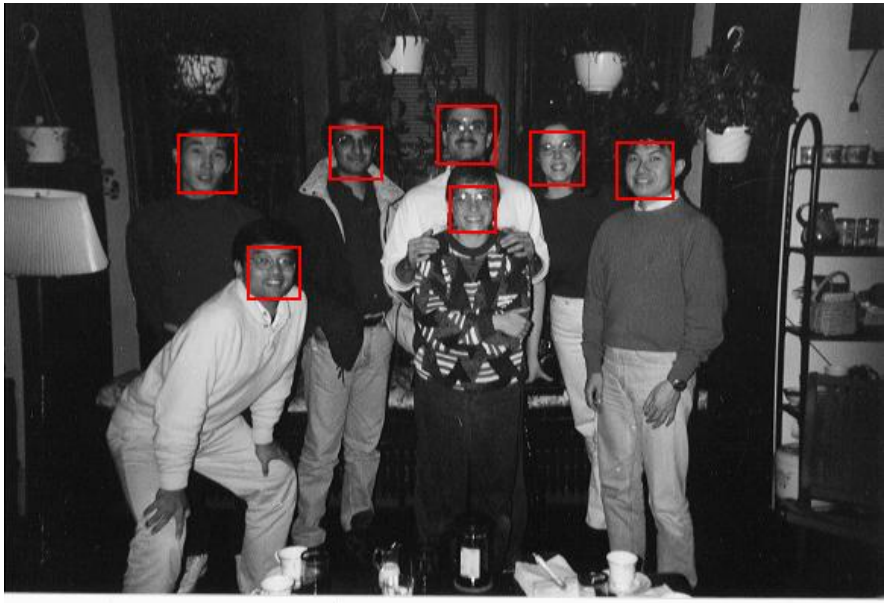
- Training time: “weeks” on 466 MHz Sun workstation
- 38 layers, total of 6061 features
- Average of 10 features evaluated per window on test set
- “On a 700 Mhz Pentium III processor, the face detector can process a 384 by 288 pixel image in about .067 seconds”
  - 15 Hz
  - 15 times faster than previous detector of comparable accuracy (Rowley et al., 1998)

Output of Face Detector on Test Images



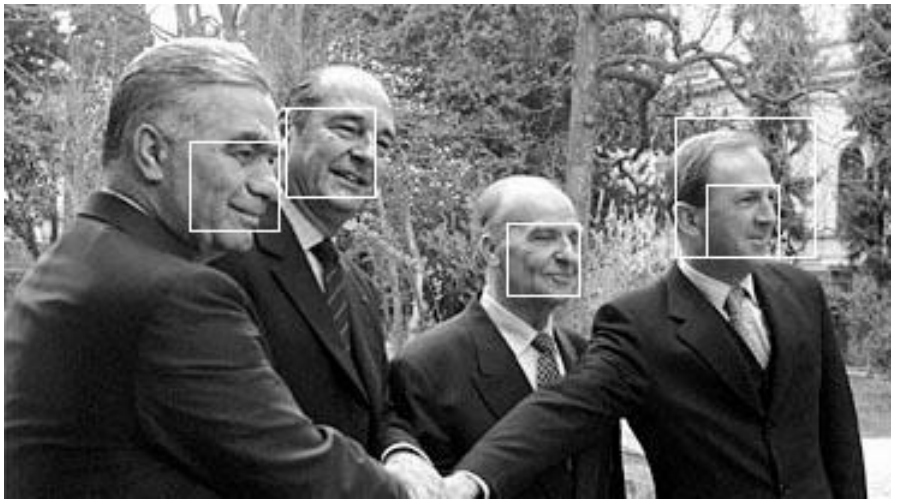






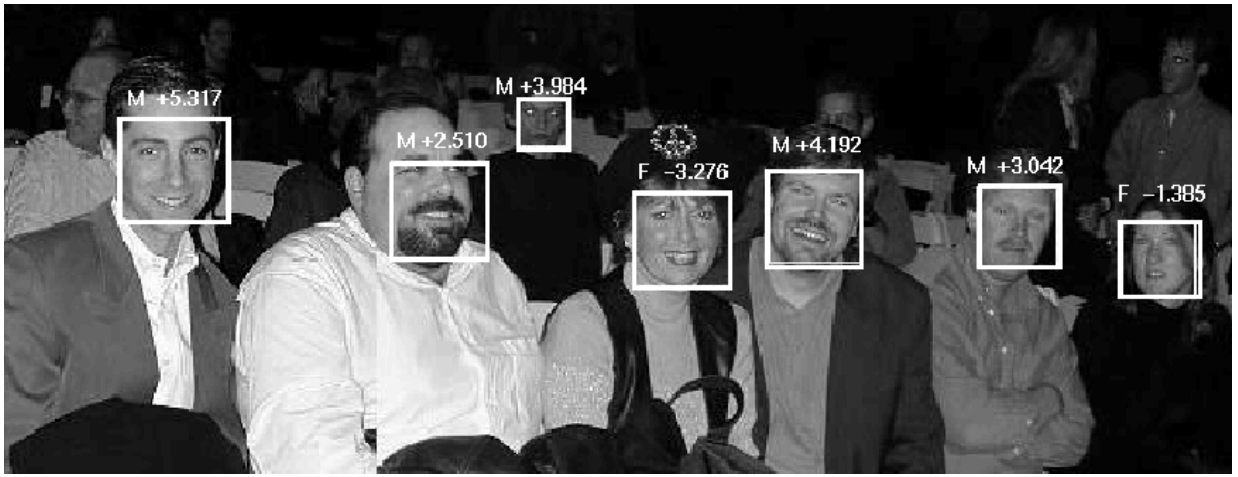
Other detection tasks

Facial Feature Localization

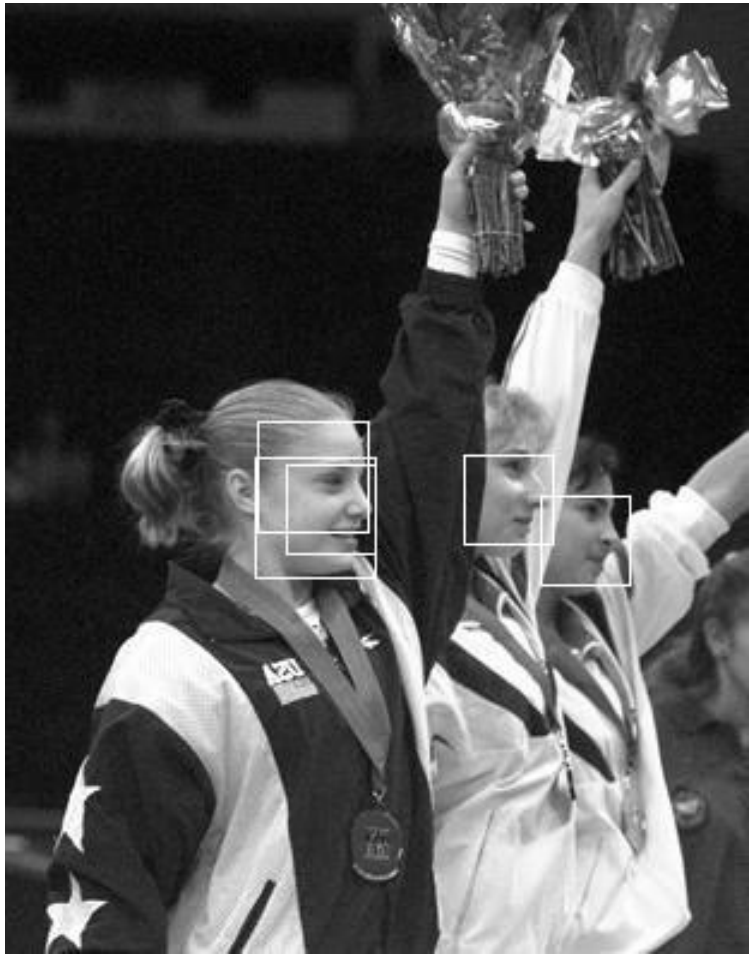


Profile Detection

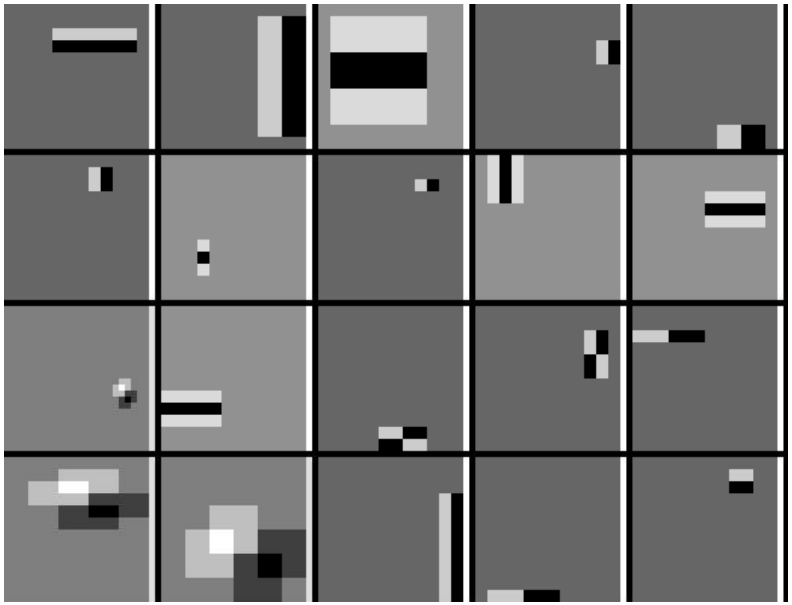
Male vs. female



Profile Detection



Profile Features



Summary: Viola/Jones detector

- Rectangle features
- Integral images for fast computation
- Boosting for feature selection
- Attentional cascade for fast rejection of negative windows

Overview

Face Recognition

- Brief review of Eigenfaces
- Active Appearance models

Face Detection

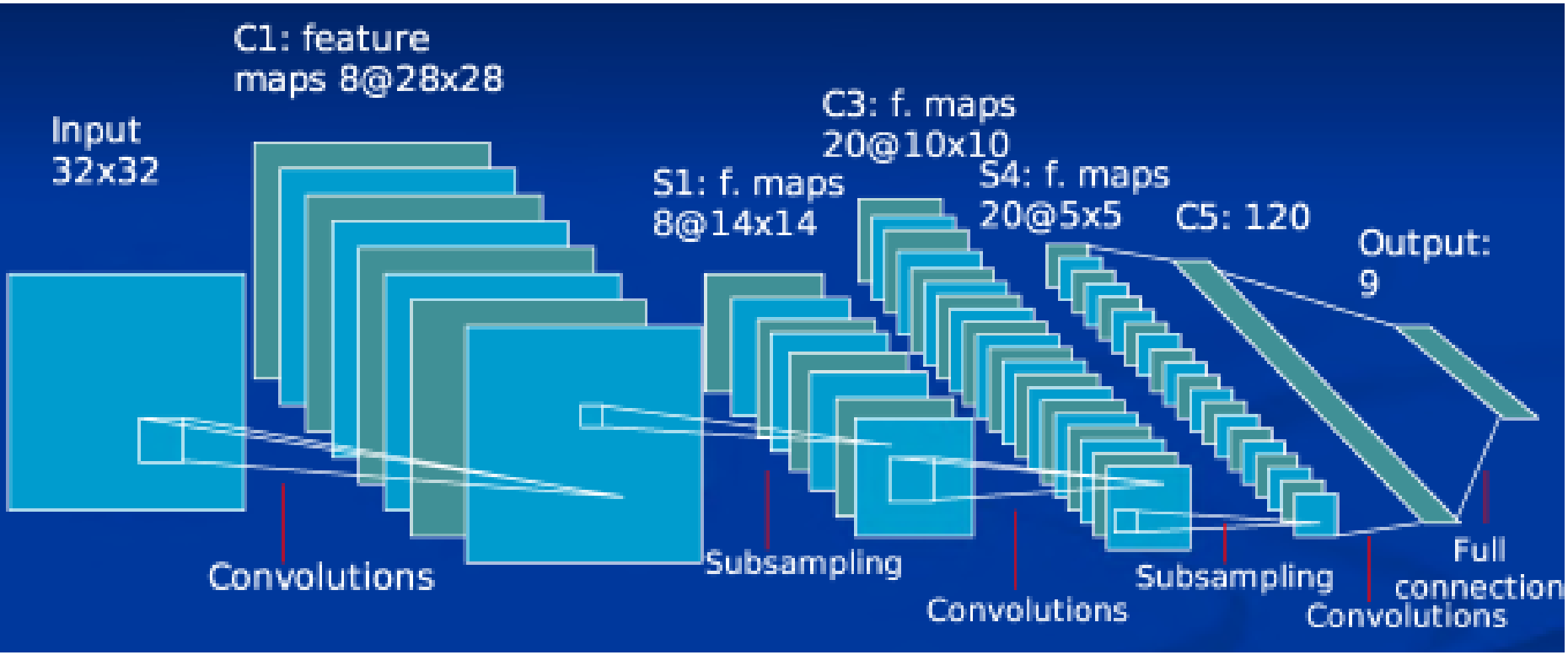
- Viola & Jones real-time face detector
- Convolutional Neural Networks

Specific Object Recognition

- SIFT based recognition

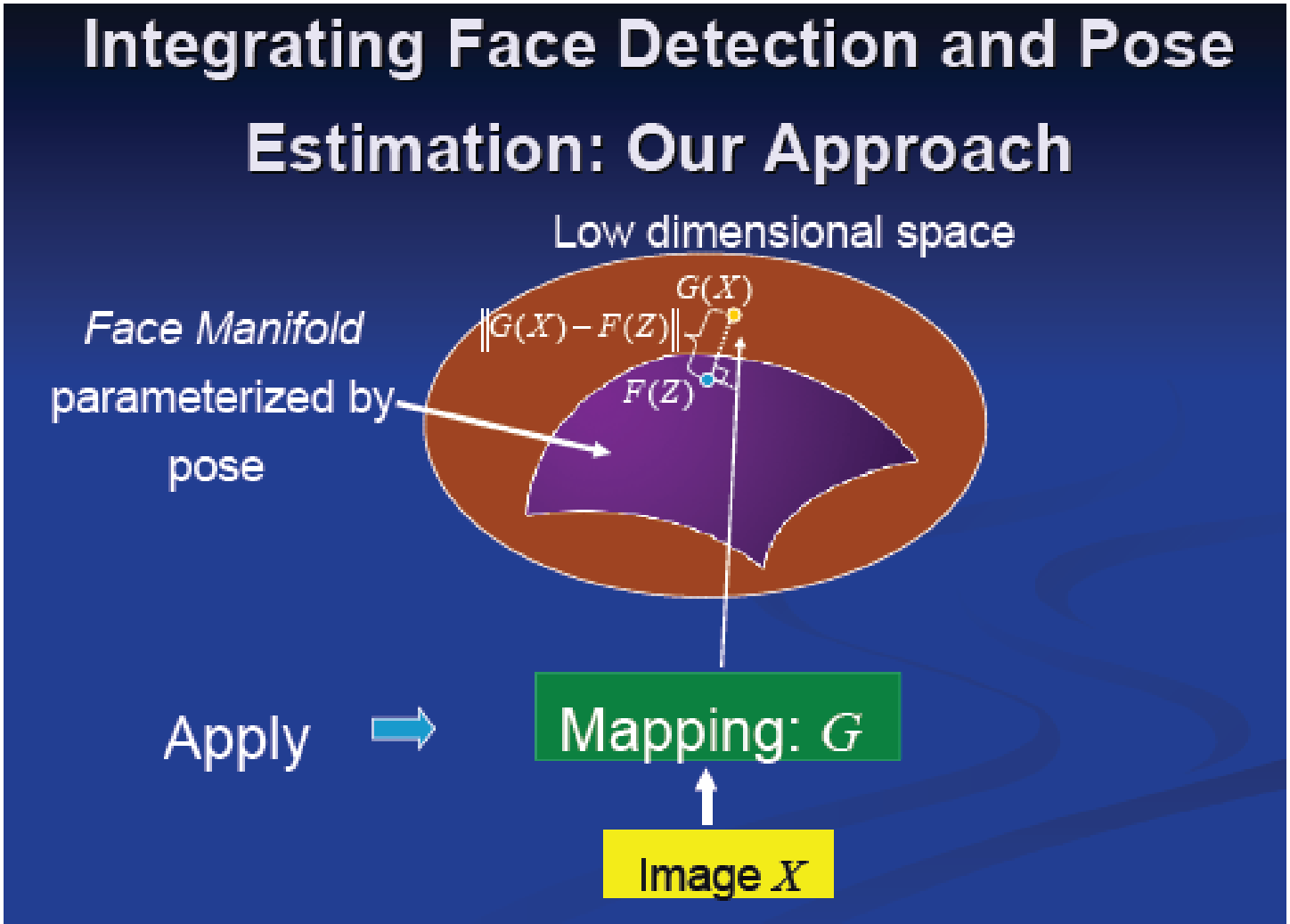
Osadchy, Miller, LeCun.
Face Detection and Pose Estimation, 2004

- Application of Convolutional Neural Networks to Face Detection



Osadchy, Miller, LeCun.
Face Detection and Pose Estimation, 2004

- Non-linear dimensionality reduction



Osadchy, Miller, LeCun.

Face Detection and Pose Estimation, 2004