The Viola/Jones Face Detector
(2001)

- A widely used method for real-time object detection.
- Training is slow, but detection is very fast.

(Most slides from Paul Viola)
Classifier is Learned from Labeled Data

• Training Data
  – 5000 faces
    • All frontal
  – 300 million non faces
    • 9400 non-face images
  – Faces are normalized
    • Scale, translation

• Many variations
  – Across individuals
  – Illumination
  – Pose (rotation both in plane and out)
Key Properties of Face Detection

• Each image contains 10 - 50 thousand locs/scales
• Faces are rare 0 - 50 per image
  – 1000 times as many non-faces as faces
• Extremely small # of false positives: 10^{-6}
**AdaBoost**

- Given a set of weak classifiers originally: \( h_j(x) \in \{+1, -1\} \)
  - 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
AdaBoost

Freund & Shapire

Final classifier is linear combination of weak classifiers
AdaBoost:
Super Efficient Feature Selector

• Features = Weak Classifiers
• Each round selects the optimal feature given:
  – Previous selected features
  – Exponential Loss
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
The Integral Image

- 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:
  \[ \text{sum} = A - B - C + D \]
• Only 3 additions are required for any size of rectangle!
  – This is now used in many areas of computer vision
Feature Selection

• For each round of boosting:
  – Evaluate each rectangle filter on each example
  – Sort examples by filter values
  – Select best threshold for each filter (min $Z$)
  – Select best filter/threshold (= Feature)
  – Reweight examples

• $M$ filters, $T$ thresholds, $N$ examples, $L$ learning time
  – $O(MTL(MTN))$ Naïve Wrapper Method
  – $O(MN)$ Adaboost feature selector
Example Classifier for Face Detection

A classifier with 200 rectangle features was learned using AdaBoost.

95% correct detection on test set with 1 in 14084 false positives.

Not quite competitive...

ROC curve for 200 feature classifier
Building Fast Classifiers

• Given a nested set of classifier hypothesis classes

• Computational Risk Minimization

![Diagram of nested classifier hypothesis classes and computational risk minimization with image sub-window classifiers and decision tree.]
Cascaded Classifier

- A 1 feature classifier achieves 100% detection rate and about 50% false positive rate.
- A 5 feature classifier achieves 100% detection rate and 40% false positive rate (20% cumulative)
  – using data from previous stage.
- A 20 feature classifier achieve 100% detection rate with 10% false positive rate (2% cumulative)
Output of Face Detector on Test Images
Solving other “Face” Tasks

Facial Feature Localization

Profile Detection

Demographic Analysis
Feature Localization Features

- Learned features reflect the task
Profile Detection
Profile Features
Review: Colour

• Spectrum of illuminant and surface
• Human colour perception (trichromacy)
• Metameric lights, Grassman’s laws
• RGB and CIE colour spaces
• Uniform colour spaces
• Detection of specularities
• Colour constancy
Review: Invariant features

- Scale invariance, using image pyramid
- Orientation selection
- Local region descriptor (vector formation)
- Matching with nearest and 2nd nearest neighbours
- Object recognition
- Panorama stitching
Review: Classifiers

- Bayes risk, loss functions
- Histogram-based classifiers
- Kernel density estimation
- Nearest-neighbor classifiers
- Neural networks

Viola/Jones face detector

- Integral image
- Cascaded classifier