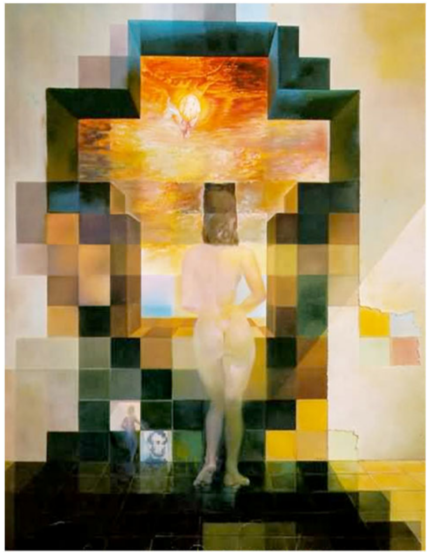
Face detection

Some slides courtesy Kaihang Wu and Antonio Torralba

Faces



Sinha *et al.*: Face Recognition by Humans: Nineteen Results Researchers Should Know About

Fig. 1. ***Unlike current machine-based systems, human observers are able to handle significant degradations in face images. For instance, subjects are able to recognize more than half of all familiar faces shown to them at the resolution depicted here. Individuals shown in order are: Michael Jordan, Woody Allen, Goldie Hawn, Bill Clinton, Tom Hanks, Saddam Hussein, Elvis Presley, Jay Leno, Dustin Hoffman, Prince Charles, Cher, and Richard Nixon.***

The Importance of Eyebrows

Fig. 5. *Sample stimuli from Sadr et al.'s [70] experiment assessing the contribution of eyebrows to face recognition: original images of President Richard M. Nixon and actor Winona Ryder, along with modified versions lacking either eyebrows or eyes.*

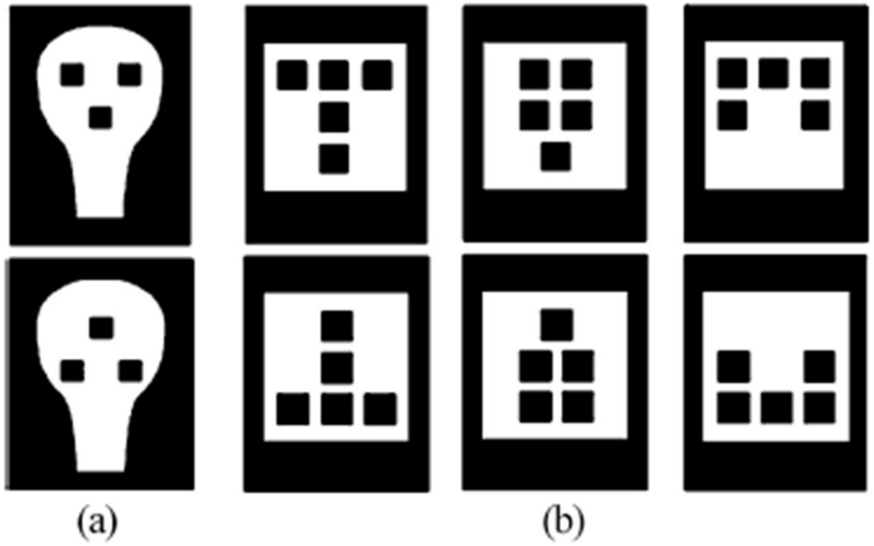
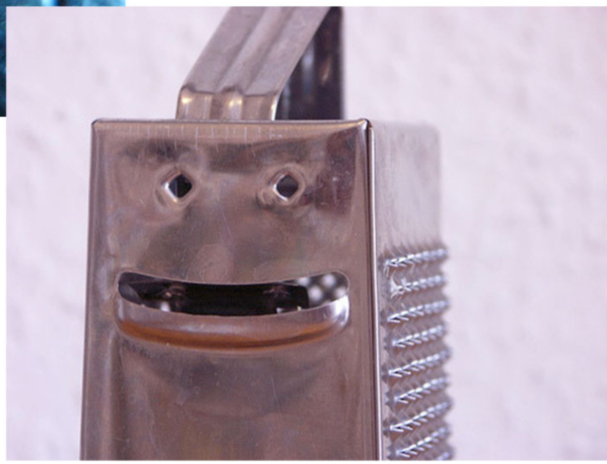


Fig. 15. *(a) Newborns preferentially orient their gaze to face-like pattern on top, rather than one shown on bottom, suggesting some innately specified representation for faces (from [36]). (b) As a counterpoint to idea of innate preferences for faces, Simion et al. [73] have shown that newborns consistently prefer top-heavy patterns (left column) over bottom-heavy ones (right column). It is unclear whether this is the same preference exhibited in earlier work, and if it is, whether it is face-specific or some other general-purpose or artifactual preference.*

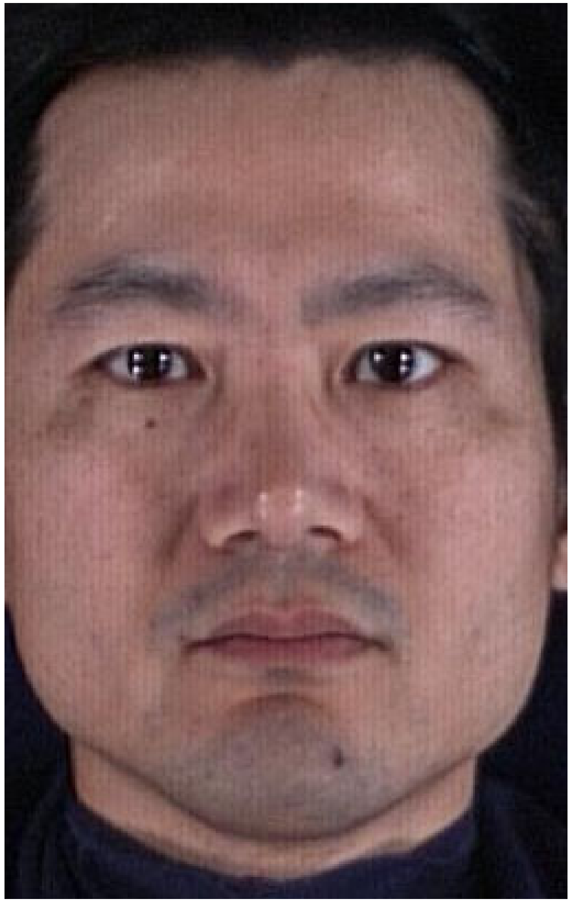
Faces everywhere



http://www.marcofolio.net/imagedump/faces_everywhere_15_images_8_illusions.html

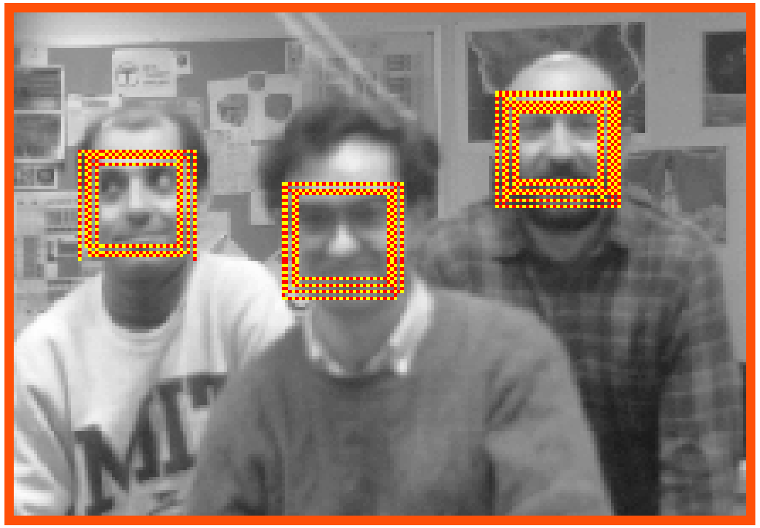
Why is that nobody has reported finding an arm, or an ear on one of the Mars pictures?

Challenges: illumination



slide credit: S. Ullman

FACE DETECTION



Face Detection

- Goal: Identify and locate human faces in an image (usually gray scale) regardless of their position, scale, in plane rotation, orientation, pose and illumination
- The first step for any automatic face recognition system
- A very difficult problem!
- First aim to detect upright frontal faces with certain ability to detect faces with different pose, scale, and illumination
- One step towards Automatic Target Recognition or generic object recognition

Where are the faces, if any?

Why Is Face Detection Difficult?

- **Pose**: Variation due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.
- **Presence or absence of structural components**: Facial features such as beards, mustaches, and glasses may or may not be present, and there is a great deal of variability amongst these components including shape, color, and size.
- **Facial expression**: The appearance of faces are directly affected by a person's facial expression.
- **Occlusion**: Faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.
- **Image orientation**: Face images directly vary for different rotations about the camera's optical axis.
- **Imaging conditions**: When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

Challenges

- Challenging Problem
- Image conditions:
  size, lighting, resolution,distortion
- Out-of-plane / In-plane rotation
- Partial occlusions:
  sunglasses, long hair, hats, masks

PIE - Pose, Illumination, and Expression






Rapid Object Detection Using a Boosted Cascade of Simple Features

Paul Viola Michael J. Jones
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA

Most of this work was done at Compaq CRL before the authors moved to MERL

Manuscript available on Blackboard:

http://citeseer.ist.psu.edu/cache/papers/cs/23183/http:zSzzSzwww.ai.mit.eduzSzpeoplezSzviolazSzresearchzSzpublicationszSzICCV01-Viola-Jones.pdf/viola01robust.pdf

What is novel about this approach?

- Feature set (… is huge about 16,000,000 features)
- Efficient feature selection using AdaBoost
- New image representation: Integral Image
- Cascaded Classifier for rapid detection
  - Hierarchy of Attentional Filters

The combination of these ideas yields the fastest known face detector for gray scale images.

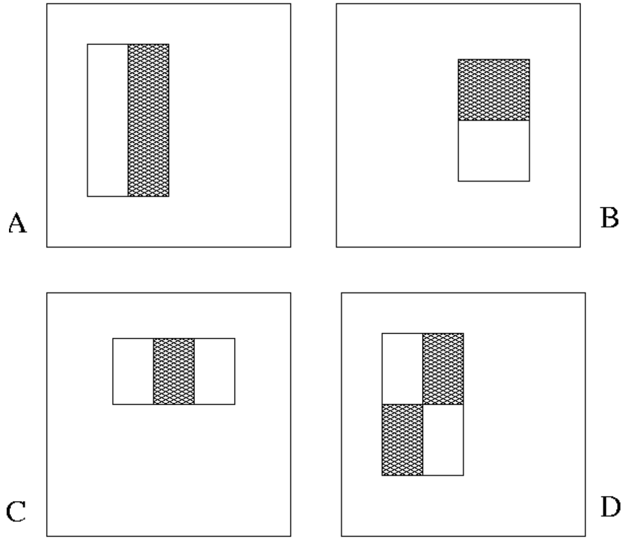
Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Image Features

"Rectangle filters"

Similar to Haar wavelets

Differences between sums of pixels in adjacent rectangles



$$h_t(x) = \begin{cases} +1 & \text{if } f_t(x) > \theta_t \\ -1 & \text{otherwise} \end{cases}$$

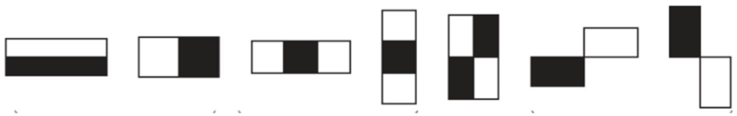
$$\boxed{160{,}000 \times 100 = 16{,}000{,}000}$$

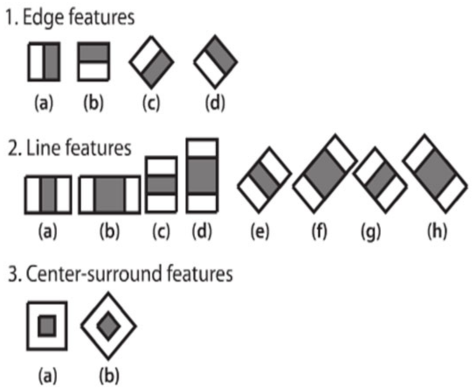
Unique Features

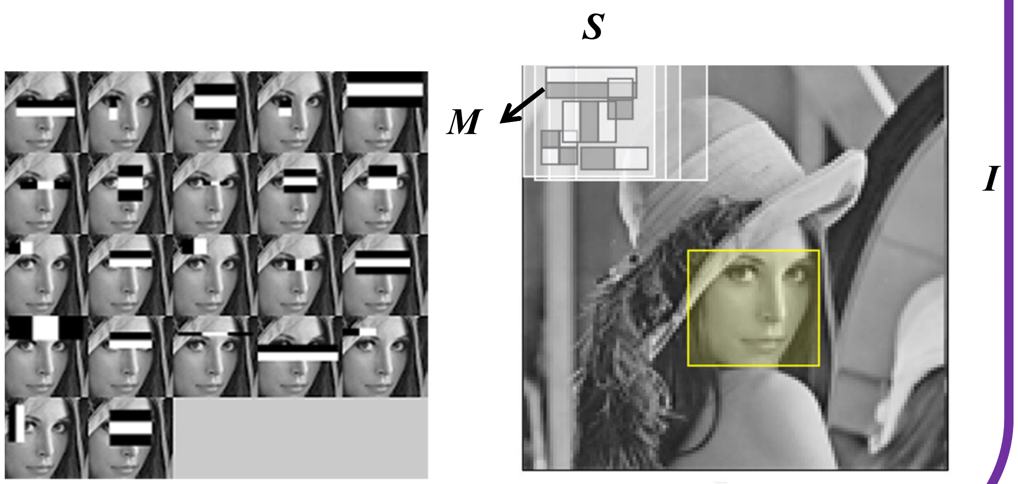
Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Haar Features

- Rectangle Features
  - Subtract sum of pixels in white area from the sum of pixels in black area
- Three original types of features and several new features are generally used
- Extended feature set







Integral Image

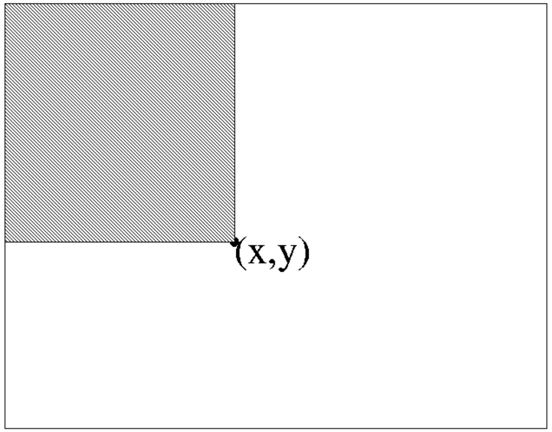
- Define the Integral Image

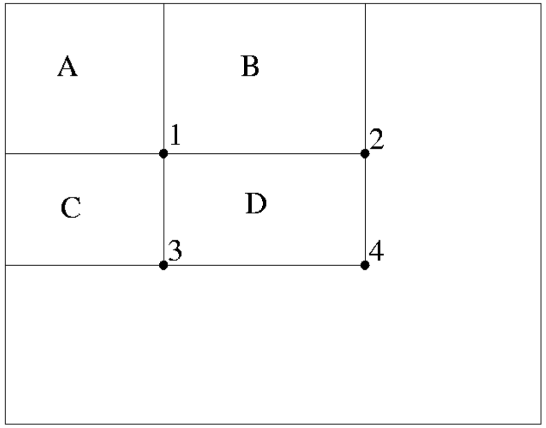
$$I'(x,y) = \sum_{\substack{x' \le x \\ y' \le y}} I(x',y')$$

- Any rectangular sum can be computed in constant time:

$$\begin{aligned} D &= 1 + 4 - (2 + 3) \\ &= A + (A + B + C + D) - (A + C + A + B) \\ &= D \end{aligned}$$

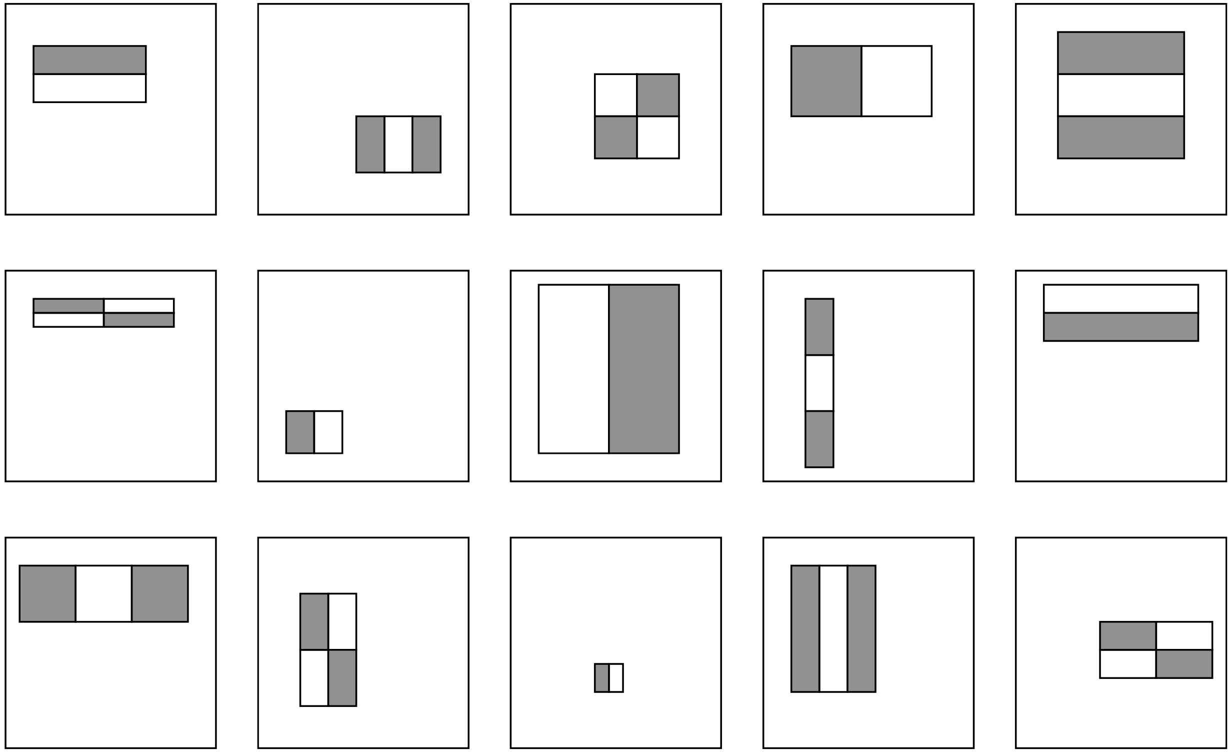
- Rectangle features can be computed as differences between rectangles





Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Huge “Library” of Filters



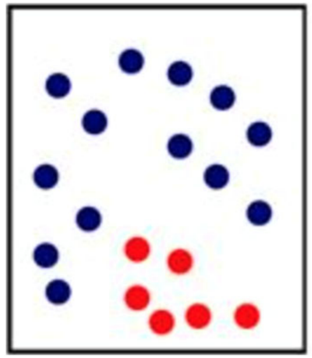
Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

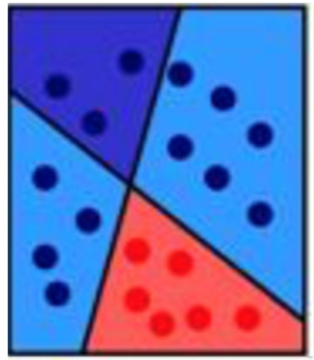
AdaBoost for Efficient Feature Selection

- Our Features = Weak Classifiers
- For each round of boosting:
  - Evaluate each rectangle filter on each example
  - Sort examples by filter values
  - Select best threshold for each filter (min error)
    - Sorted list can be quickly scanned for the optimal threshold
  - Select best filter/threshold combination
  - Weight on this feature is a simple function of error rate
  - Reweight examples
  - (There are many tricks to make this more efficient.)

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Adaboost – Machine Learning classifier





- Given: example N images labeled face/non-face
  - Initially, all weights set equally w = 1/N.
- Repeat N times
  - Step 1: choose the most efficient weak classifier with the lowest detection error
  - Step 2: Update the weights to emphasize the examples which were incorrectly classified by the first weak classifier and consider the next best classifier. This makes the next weak classifier focus on “harder” examples
- Final (strong) classifier is a weighted combination of the N “weak” classifiers
  - Weighted according to their importance

The final strong classifier

$$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots$$

Weak classifier

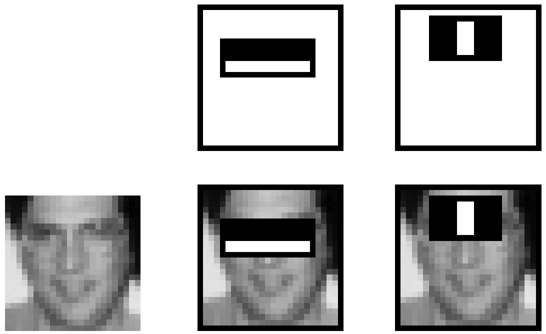
Strong classifier **Image** **Weight**

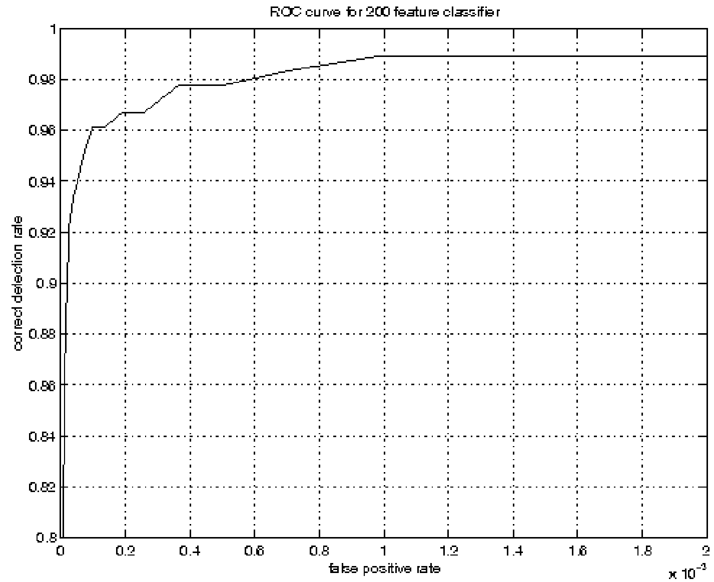
Example Classifier for Face Detection

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.
Need to add more features,
but then that slows it down.



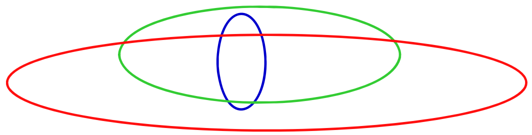


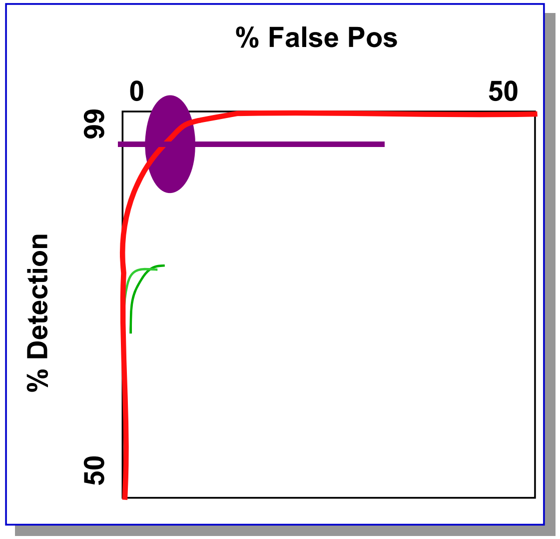
ROC curve for 200 feature classifier

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

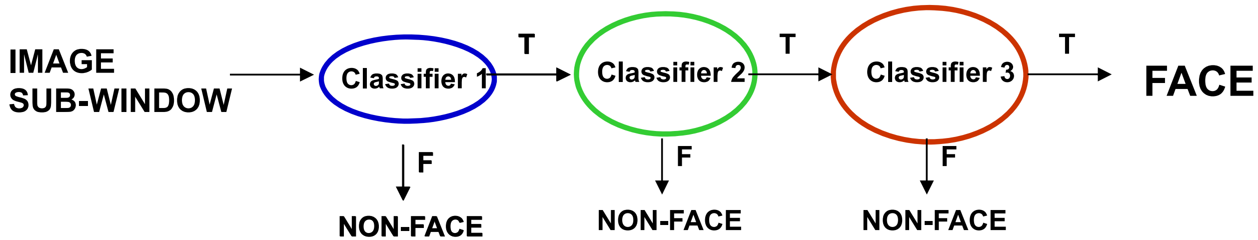
Develop fast, accurate classifier using a cascade

- Given a nested set of classifier hypothesis classes



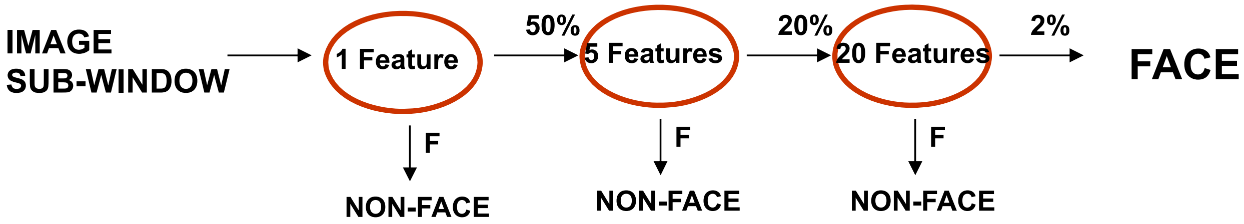


- Computational Risk Minimization



Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

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)

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

A Real-time Face Detection System

Training faces: 4916 face images (24 x 24 pixels) plus vertical flips for a total of 9832 faces

Training non-faces: 350 million sub-windows from 9500 non-face images

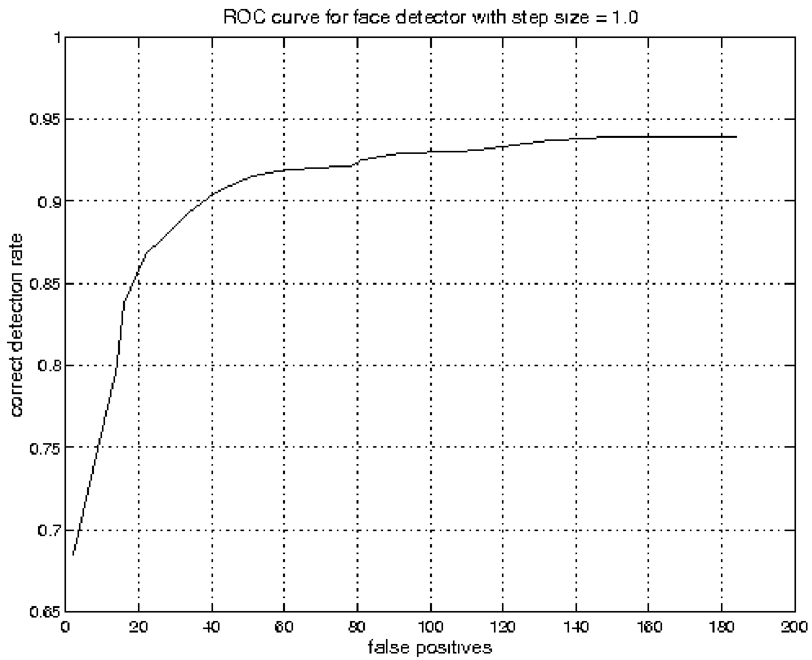
Final detector: 38 layer cascaded classifier The number of features per layer was 1, 10, 25, 25, 50, 50, 50, 75, 100, …, 200, …

Final classifier contains 6061 features.

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Accuracy of Face Detector

Performance on MIT+CMU test set containing 130 images with 507 faces and about 75 million sub-windows.



Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Comparison to Other Systems

False Detections / Detector	10	31	50	65	78	95	110	167
Viola-Jones	76.1	88.4	91.4	92.0	92.1	92.9	93.1	93.9
Viola-Jones (voting)	81.1	89.7	92.1	93.1	93.1	93.2	93.7	93.7
Rowley-Baluja-Kanade	83.2	86.0				89.2		90.1
Schneiderman-Kanade				94.4				

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Speed of Face Detector

Speed is proportional to the average number of features computed per sub-window.

On the MIT+CMU test set, an average of 9 features out of a total of 6061 are computed per sub-window.

On a 700 Mhz Pentium III, a 384x288 pixel image takes about 0.067 seconds to process (15 fps).

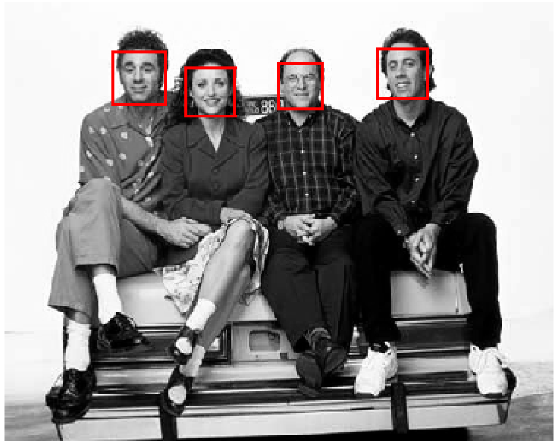
Roughly 15 times faster than Rowley-Baluja-Kanade and 600 times faster than Schneiderman-Kanade.

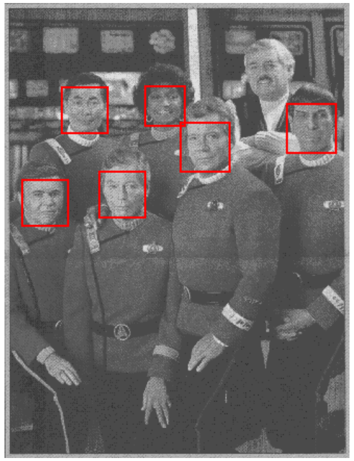
Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

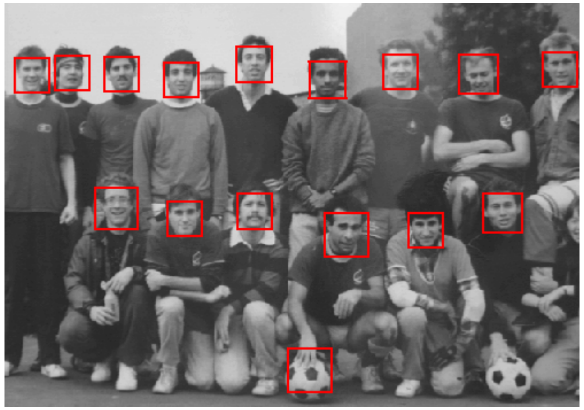
Output of Face Detector on Test Images

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

More Examples





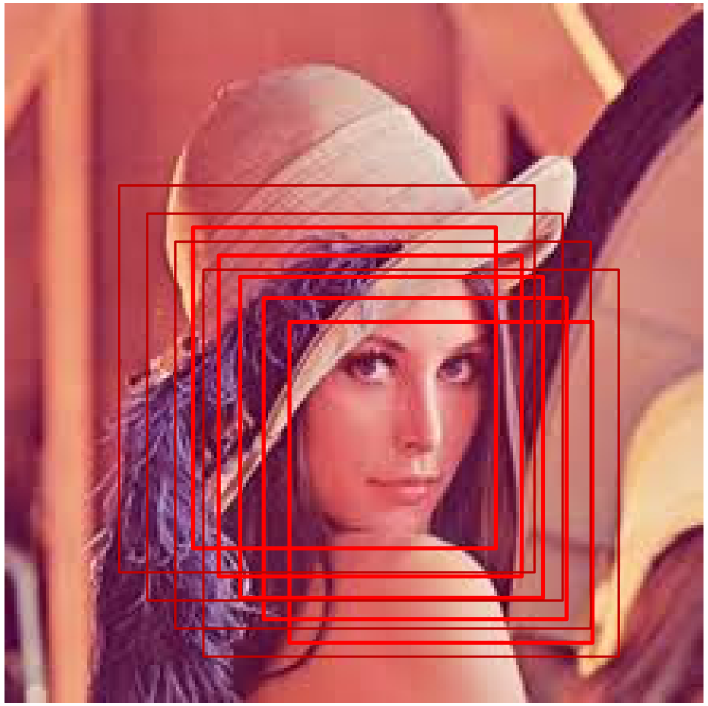


Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

Slide and rescale window

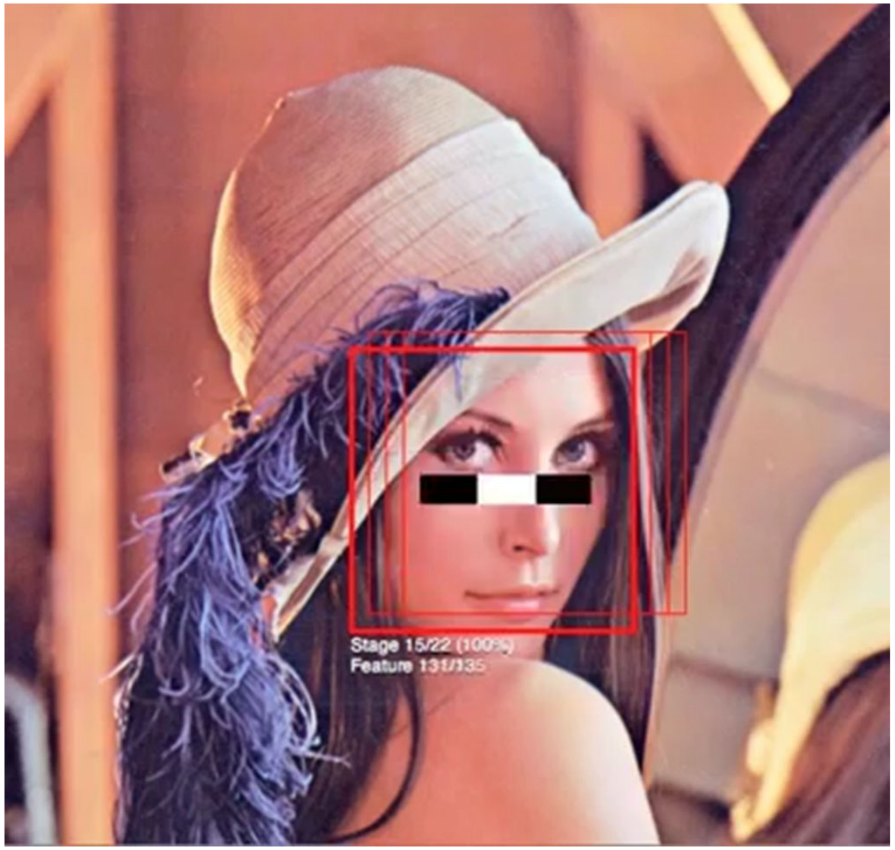
- Slide smallest window across image with stepsize (stride) of several pixels for speed
- Increase window size by, say, 20% and repeat
- Continue until you reach largest window size

Non-Maximal Suppression



If you set a threshold for 'faceness' then you will get many detections over your threshold as you pass over a face.
You need to suppress the weaker detections and keep the strongest response.

Viola-Jones Cascade Demo



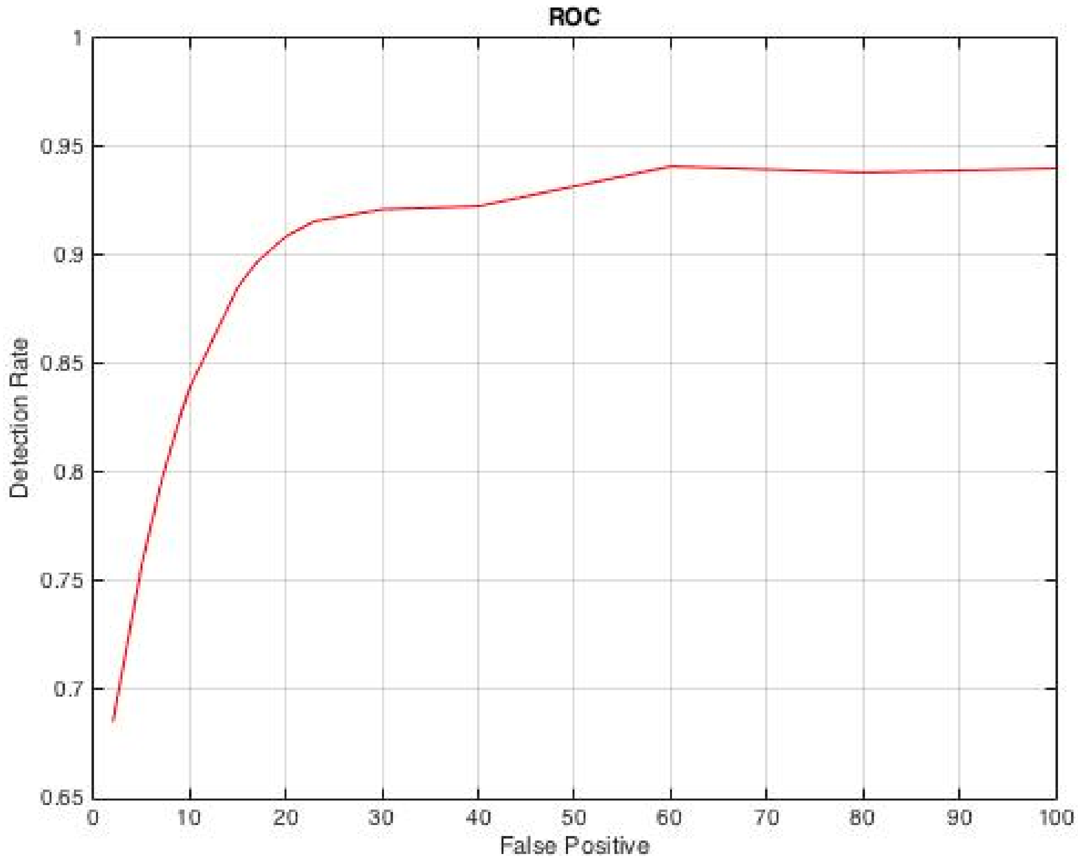
Comments

- Good for near frontal faces with up to 15 degrees pose angle
- Faces are often not well-centred due to stride of cascade algorithm
- Often need feature based aligner for face recognition
- For frontal faces, we can align the eyes
- The eyes can also be used to correct for in-plane rotation (tilt)
- For non-frontal faces, what do we align?
- Difficult to adapt to high pose angle detection because that would require more a more complex first stage which slows the algorithm and increases false acceptance

Our Performance of a single-node classifier

Examples for face samples:

Examples for non-face samples:



Our Test Performance on 4 datasets

Table 1. Datasets

Dataset	Number of images	Average image dimension	Number of annotated faces
FDDB	2845	377 x 399	5171
CMU	180	421 x 422	721
Yale	165	320 x 243	165
ORL	400	112 x 92	400

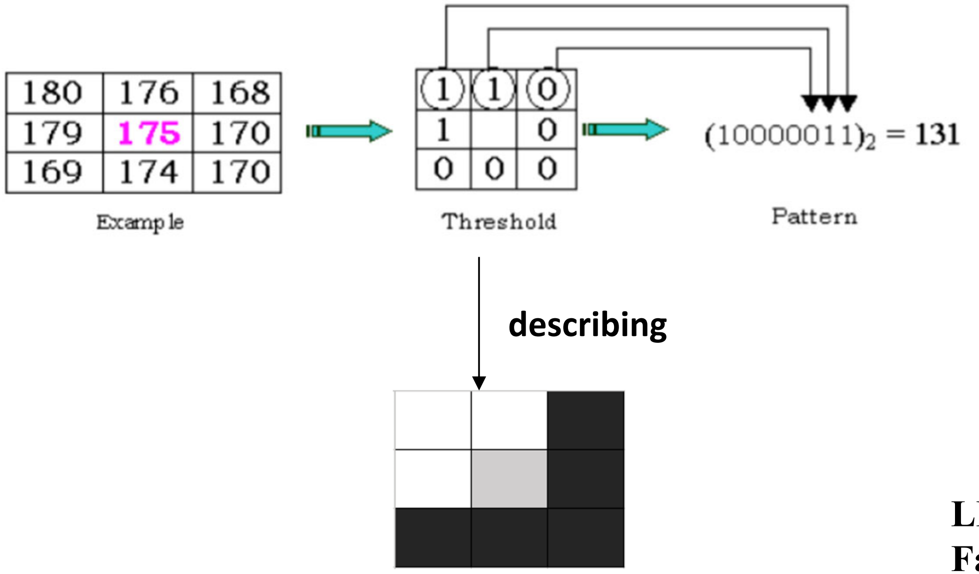
Table 2. The comparison of performance

Test set performance								
Traing set	FDDB		CMU		Yale		ORL	
stage	Time Precision	Detection Rate	Time Precision	Detection Rate	Time Precision	Detection Rate	Time Precision	Detection Rate
1	34998s	0.512	13344s	0.751	10464s	0.584	23829s	0.662
3	56032s	0.527	28976s	0.774	25121s	0.561	35185s	0.690
10	~23h	0.654	~54000s	0.876	~53560s	0.640	~62062s	0.835
20	~50h	0.740	~27h	0.958	~27h	0.769	~31h	0.940

Let’s try LBP features

- Comparative result (Time)

- LBP features:



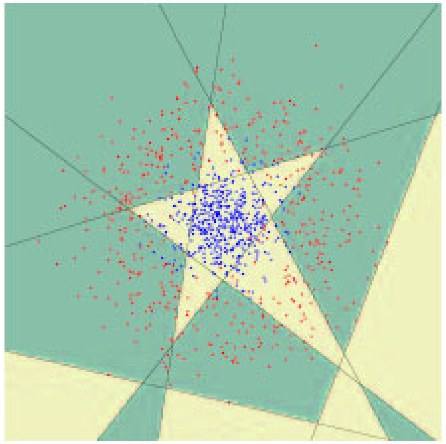
Evaluation performance		
Feature	V.J. algorithm	
	Time(s)	Detection Rate
Haar	2204.4	0.937
LBP	1113.33	0.870

LBP is Local Binary Patterns
Faster than Haar but slightly less accurate

References

1. Min R, Hadid A, Dugelay J-L. Efficient Detection of Occlusion prior to Robust Face Recognition. *The Scientific World Journal*. 2014;2014:519158. doi:10.1155/2014/519158.
2. Zhang, Zhanpeng, et al. "Facial landmark detection by deep multi-task learning." *Computer Vision–ECCV 2014*. Springer International Publishing, 2014. 94-108.
3. J. Chang-yeon, "Face Detection using LBP features," *Final Project Report,* vol. 77, 2008.
4. J. Šochman and J. Matas, "AdaBoost and face detection," *CZECH TECHNICAL UNIVERSITY. República Checa,* 2003.
5. P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, pp. 137-154, 2004.

Adaboost



Strong Classifiers formed from Weak Classifiers

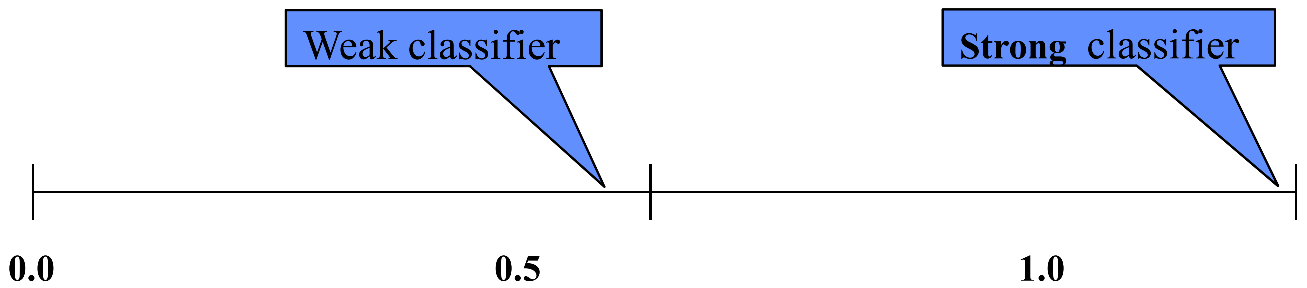
Can a crowd be smarter than the participants in the crowd?

Pedigree of Adaboost

- **AdaBoost**, short for "Adaptive Boosting", is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire who won the Gödel Prize in 2003 for their work. It can be used in conjunction with many other types of learning algorithms to improve their performance.
- Integral element of Viola-Jones Face Detection which won most significant paper over 10 years at CVPR in 2011
- It is easy and it actually works!
- See Patrick Winston talk at MIT https://www.youtube.com/watch?v=UHBmv7qCey4

Binary Classifier

- Answers question such as “Is this a face or non-face?”
- Assume we have a set of classifiers of the form:
  - h [-1,1]
    - i.e. the classifier h looks at input and gives -1 or 1 as output
- Lets look at the error rates



Voting Classifier

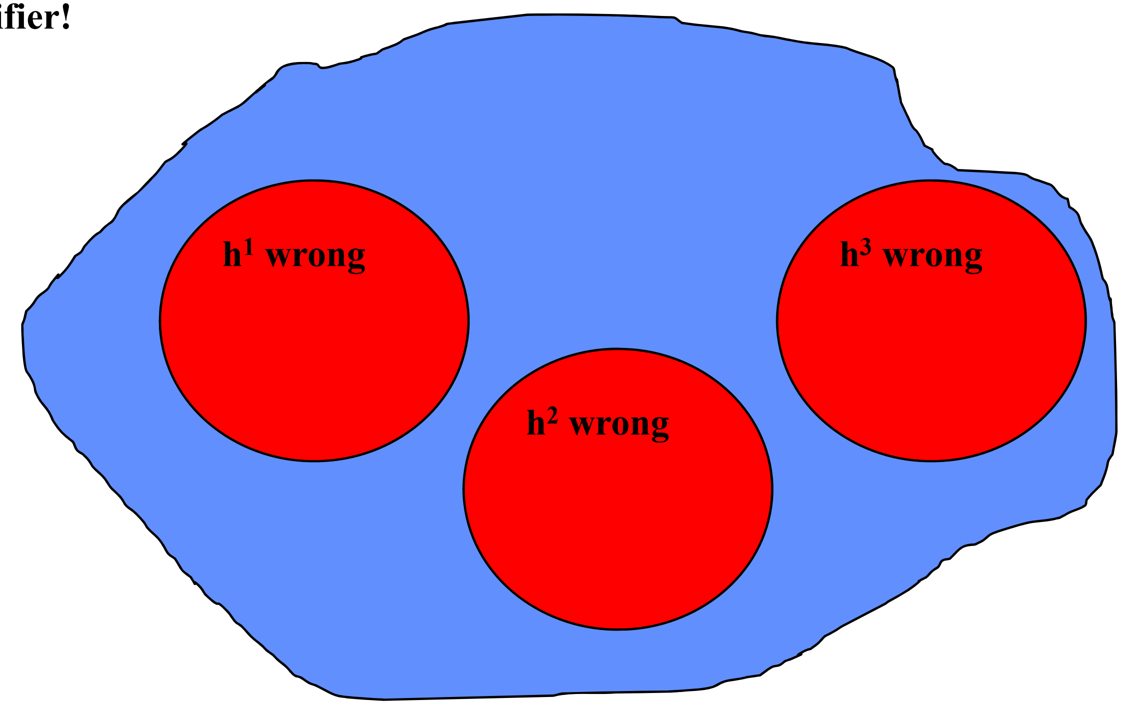
- Let's try to improve performance by creating a few classifiers and taking a vote

$$H(x) = \mathrm{sgn}(h^1(x) + h^2(x) + h^3(x))$$

Correct if 2 out of 3 are correct

Sample Set

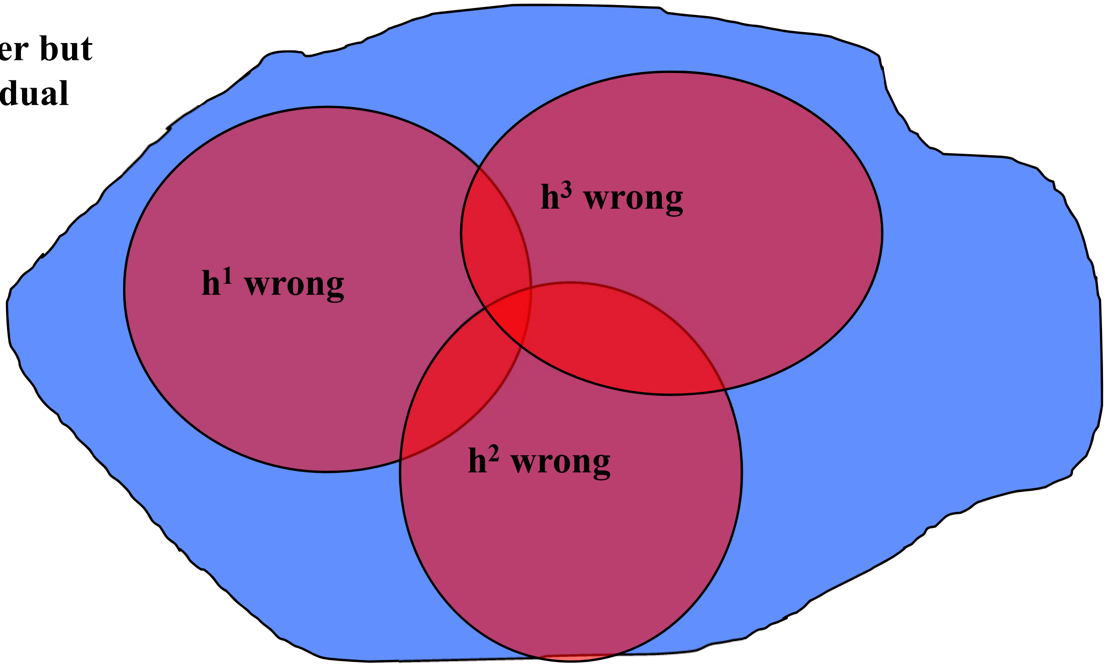
Perfect Classifier!



Sample Set

Imperfect classifier but better than individual Weak classifiers

Is this always True?



Always better or equal to the single worst classifier because intersection is a subset of the set

Choose Best Classifiers

- Now how do we choose the best classifiers for the vote?
- We test classifiers on the data to see which performs best
- This is fine for the first classifier, but we need the second classifier to fix the errors in the first etc.

IDEA 1

Data $\rightarrow h^1$

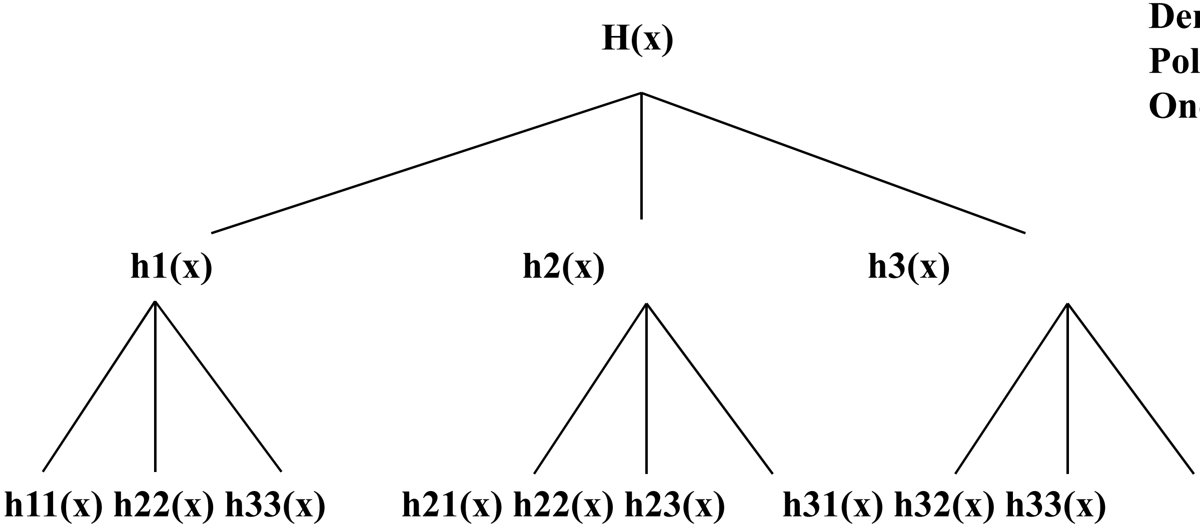
Data with exaggerated h^1 errors $\rightarrow h^2$

Data with exaggerated $h^1 \neq h^2 \rightarrow h^3$

Extend the Vote

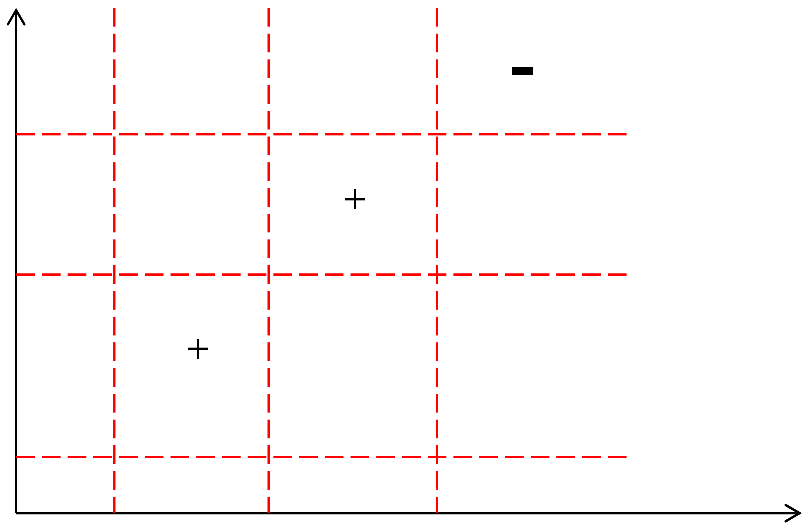
Wisdom of a crowd

Representative
Democracy or
Political Convention
One vote one value



IDEA 2

Decision Tree Stumps



How many tests?

12

Tests = Number of lines x 2 x number of dimensions

Why use Decision Tree Stumps?

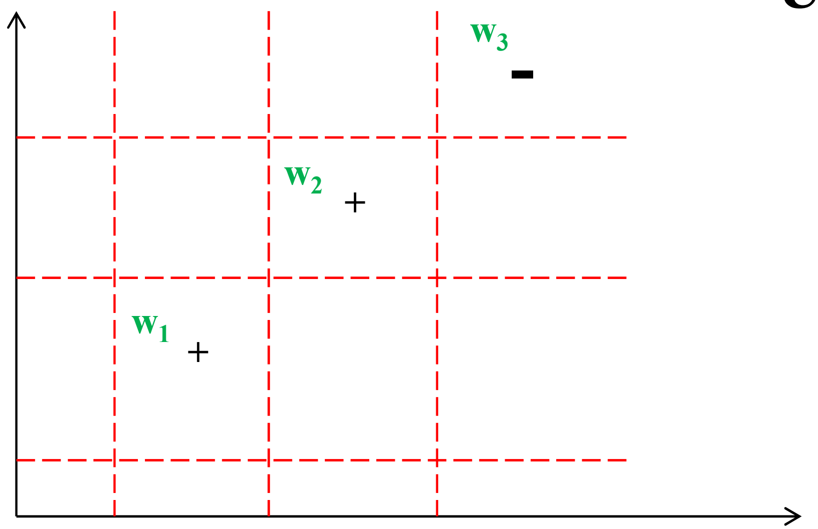
- Could use any class of classifiers
  - Neural nets
  - Hyperplanes (linear functions)
- Decision trees are easy to explain and enumerate
- Boosting works on all classifiers

Error Rate

$$\text{Error Rate, } \in \; = \sum_{wrong} \frac{1}{N}$$

So for N=50 samples, if we get 2 examples wrong, ϵ is 2/50 = 4%

How do we Exaggerate Errors?



Use error weights

$$w_i^1 = \frac{1}{N}$$

Enforces a distribution

$$\in \; = \sum_{wrong} w_i \;\text{ and } \sum w_i = 1$$

IDEA 3

Another Way to Extend the Vote

$$H(x) = \mathrm{sgn}(h^1(x) + h^2(x) + h^3(x) + \cdots)$$

But some voters are better than others!

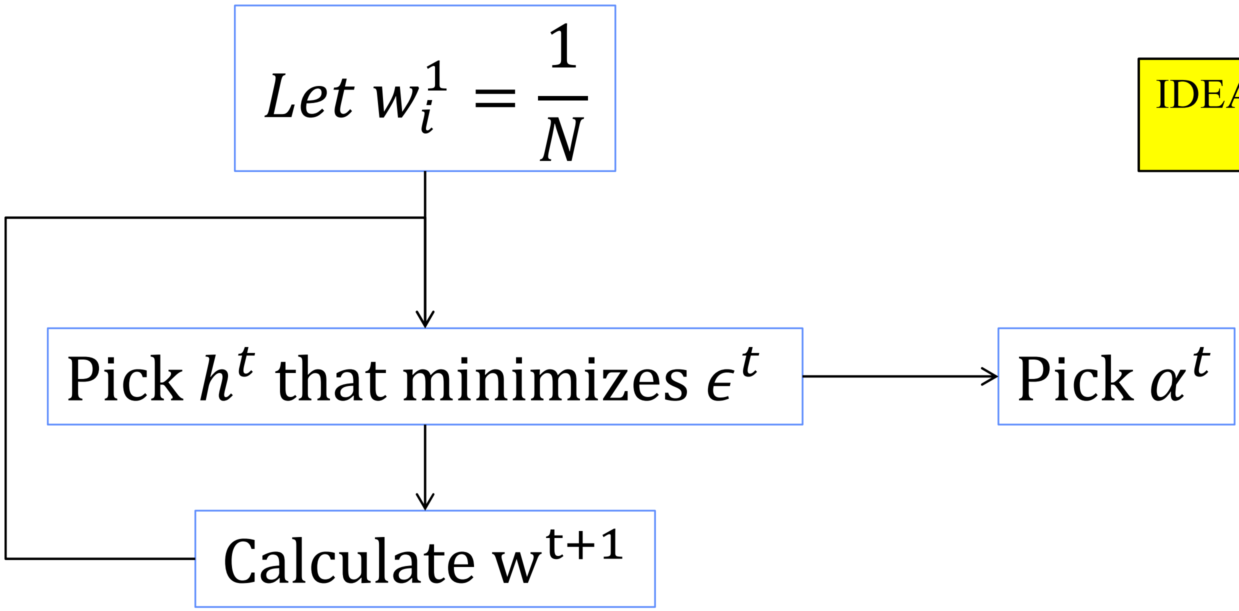
IDEA 4

$$H(x) = \mathrm{sgn}(\alpha^1 h^1(x) + \alpha^2 h^2(x) + \alpha^3 h^3(x) + \cdots)$$

Wisdom of a weighted crowd of experts

Method Outline

1. **Pick the best classifier from the set**
2. **Reweight errors and repeat**



IDEA **5**

Reweighting

$$w_i^{t+1} = \frac{w_i^t}{Z} e^{\alpha^t h^t(x) y(x)}$$

Z is normalisation factor

IDEA **6**

y(x) is function that is [-1,+1]
depending if h^t answer is wrong or right

How was this arrived at?
It took a top mathematician over a year of contemplation
Almost led to divorce!

Error Bound

Minimum error bound for whole thing if we choose

$$\alpha^t = \frac{1}{2} \frac{\ln(1-\epsilon^t)}{\epsilon^t}$$

Substitute into previous equation and the new weights are

$$w_i^{t+1} = \frac{w_i^t}{Z} * \begin{cases} \sqrt{\dfrac{\boldsymbol{\epsilon}^t}{1-\epsilon^t}} & \text{correct} \\ \sqrt{\dfrac{1-\epsilon^t}{\epsilon^t}} & \text{wrong} \end{cases}$$

Simplification

$$\sqrt{\frac{\epsilon^t}{1-\epsilon^t}}\sum_{correct}\omega_i^t + \sqrt{\frac{1-\epsilon^t}{\epsilon^t}}\sum_{wrong}\omega_i^t = Z = 2\sqrt{\epsilon^t(1-\epsilon^t)}$$

$$\omega_i^{t+1} = \frac{\omega_i^t}{2}\frac{1}{1-\epsilon}\ correct = \frac{1}{2}\frac{1}{1-\epsilon}\sum_{correct}\omega_i^t = \frac{1}{2}$$

$$= \frac{\omega_i^t}{2}\frac{1}{\epsilon}\ wrong = \frac{1}{2}$$

Easy: Sum of all weights for correct/wrong is always ½!!
No need for log or exp, simply rescale the weights

Example