Window-Based Detection (Face Detection)

Slides adapted from Kristen Grauman, Lana Lazebnik and Paul Viola

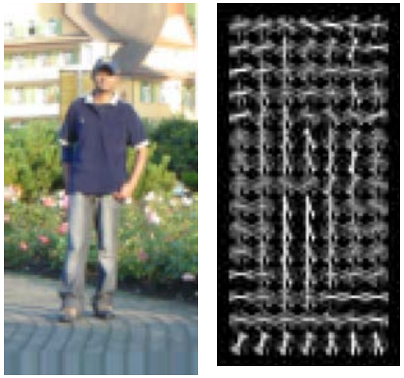
Roadmap

- Previously…
  - Instance recognition
    - Local features: detection and description
    - Local feature matching, scalable indexing
    - Spatial verification
  - Generic Object Recognition
  - Supervised Classification

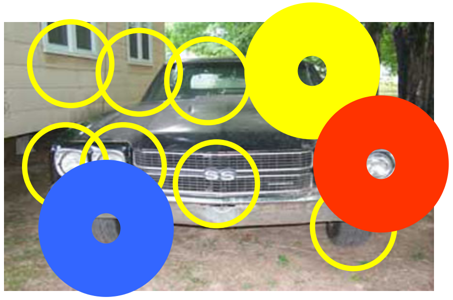
- Today…
  - Window-based generic object detection
  - Boosting classifiers
  - *Case Study: Face Detection*

Generic Category Recognition: Basic Framework

- Build/train object model
  - Choose a representation
  - Learn or fit parameters of model / classifier
- Generate candidates in new image
- Score the candidates



Window-based



Part-based

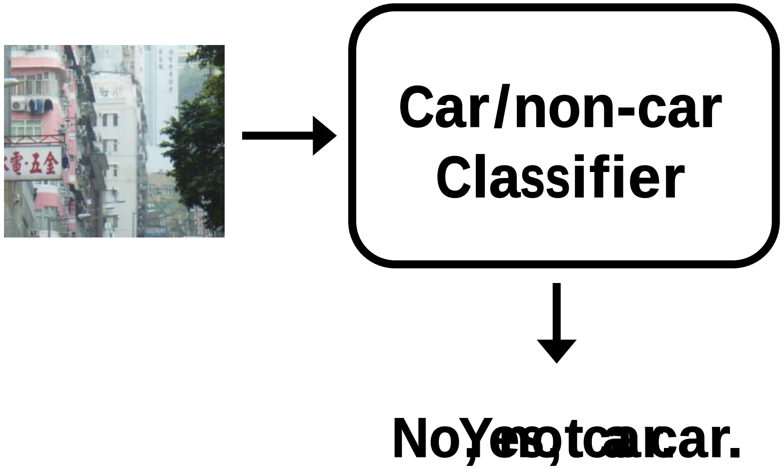
Window-Based Models

- Window
  - AKA Bounding Box, (Image) Region, (Image) Patch, Subwindow, etc.
- Object Model
  - Histogram of intensities
    - Sensitive to illumination
  - Vector of pixel intensities
    - Sensitive to small shifts
  - Global descriptors: HoG, GIST, etc.
  - Bag of Words (in the window)

Window-based "window" detector

Window-based models
Building an object model

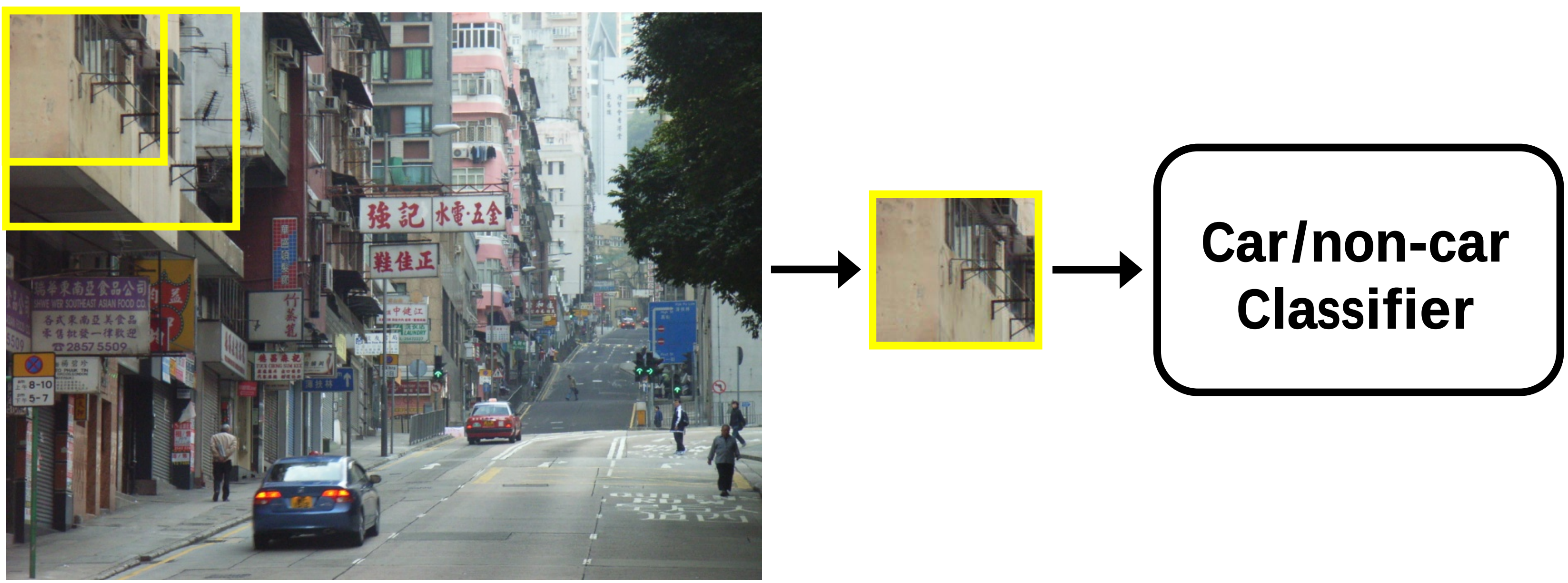
Given the representation, train a binary classifier



Generic Category Recognition: Basic Framework

- Build/train object model
  - Choose a representation
  - Learn or fit parameters of model / classifier
- **Generate candidates in new image**
- **Score the candidates**

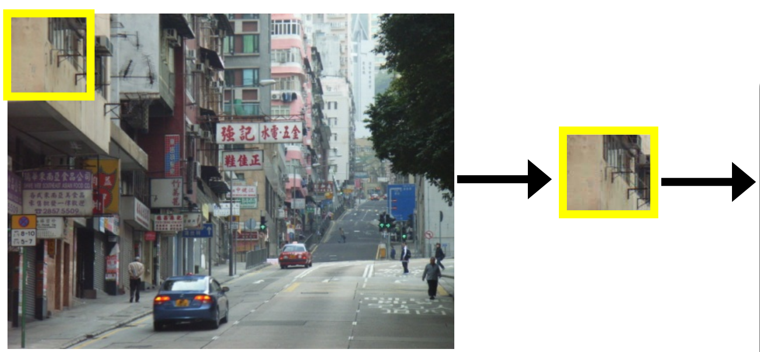
Window-based models
Generating and scoring candidates

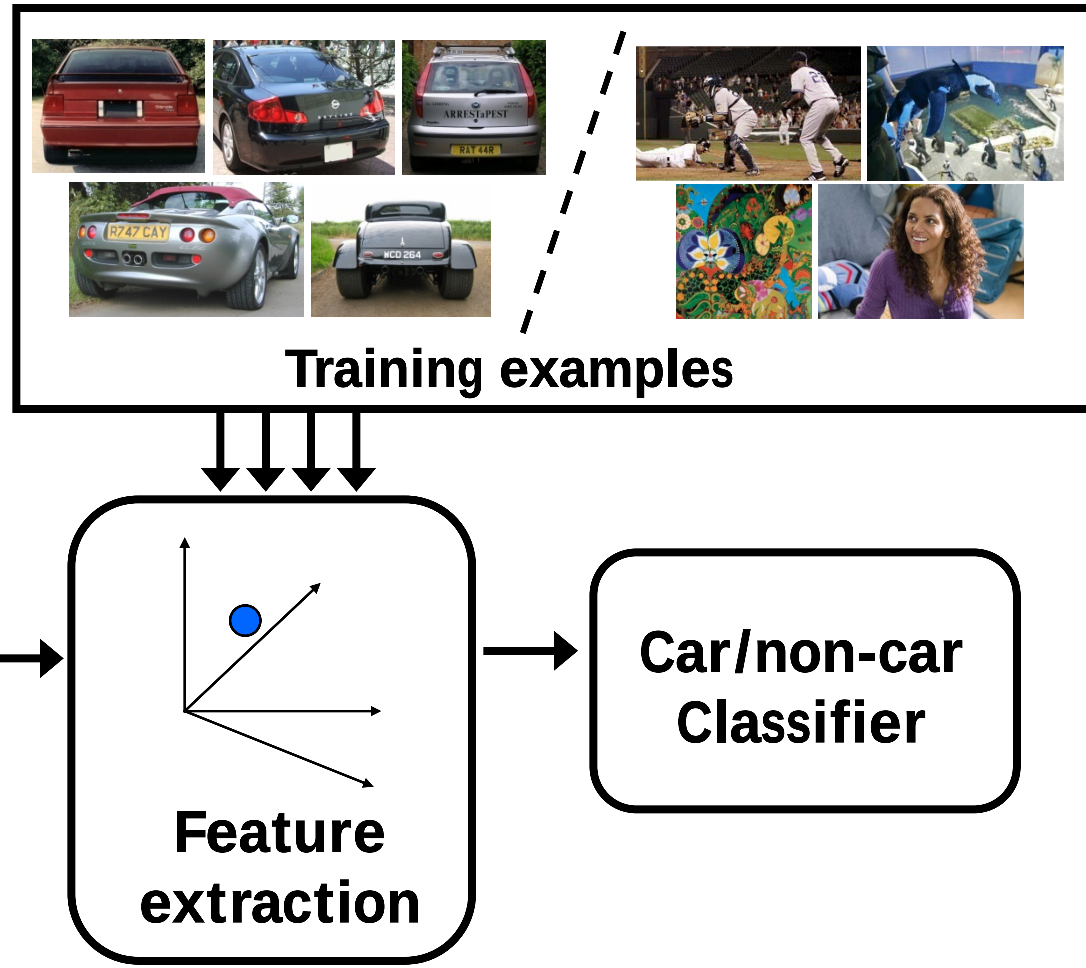


Window-based object detection: recap

Training:
1. Obtain training data
2. Define features
3. Define classifier

Given new image:
1. Slide window
2. Score by classifier





Case Study: Face Detection

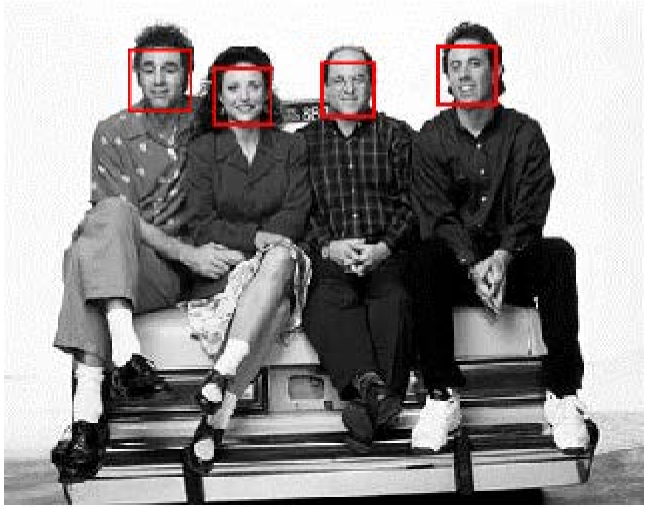
ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001

Rapid Object Detection using a Boosted Cascade of Simple Features

Paul Viola
viola@merl.com

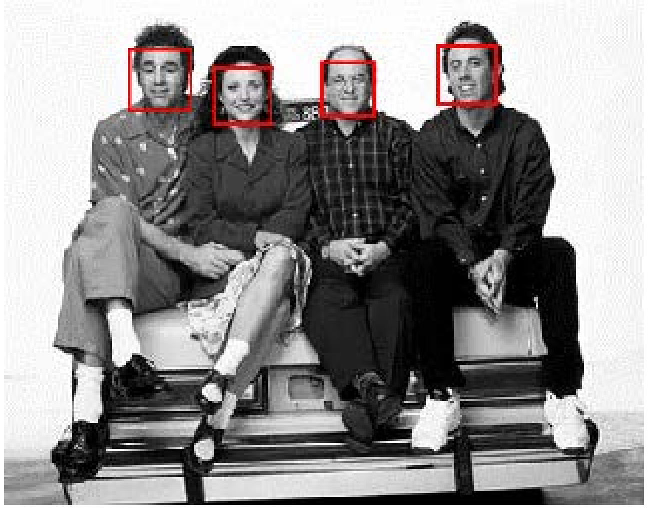
Michael Jones
mjones@crl.dec.com

- "Viola-Jones Detector"
  - **160th most cited paper in computer science**
    - #161 is "Dynamic Programming" (1957)

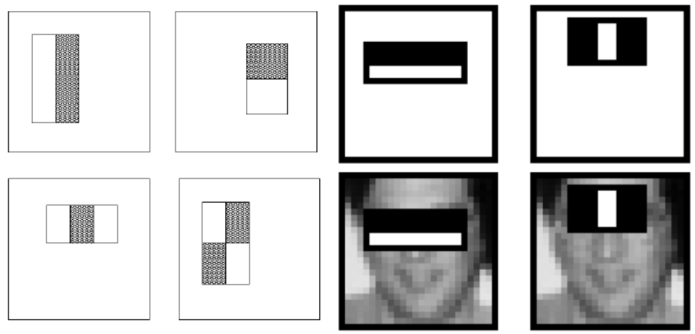


Viola-Jones Face Detector

- **3 Key Ideas:**
  - Efficient local texture features
    - integral images
  - Boosting
    - Select discriminative features to be weak classifiers
    - Combine them to create final classifier
  - Cascade Classification
    - Reject clear negatives quickly



Viola-Jones Detector: Features



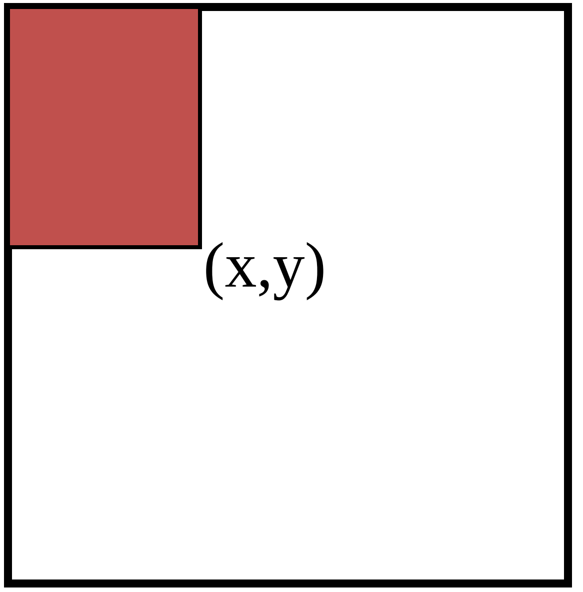
- Simple "Haar-like" features
  ∑(pixel intensities in black) – ∑(pixels intensities in white)

  – Features can vary in shape, position, # of boxes

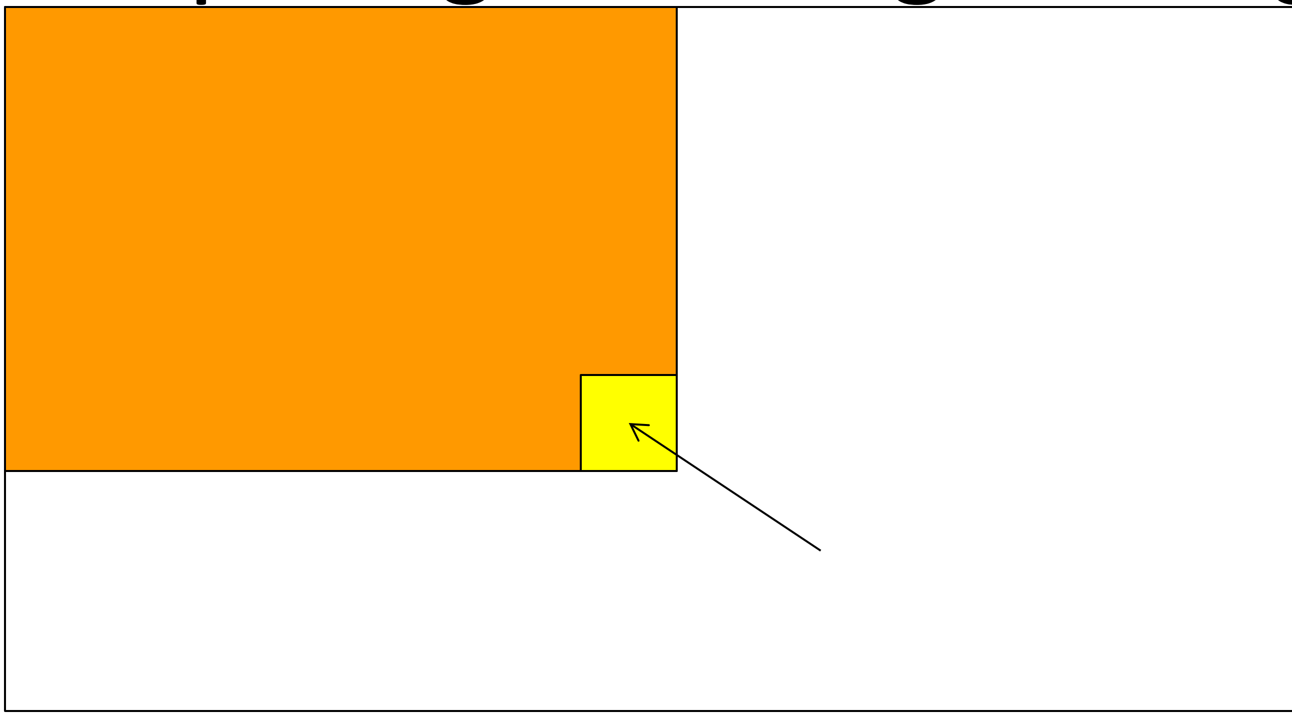
  – Result is a scalar value

Integral Images

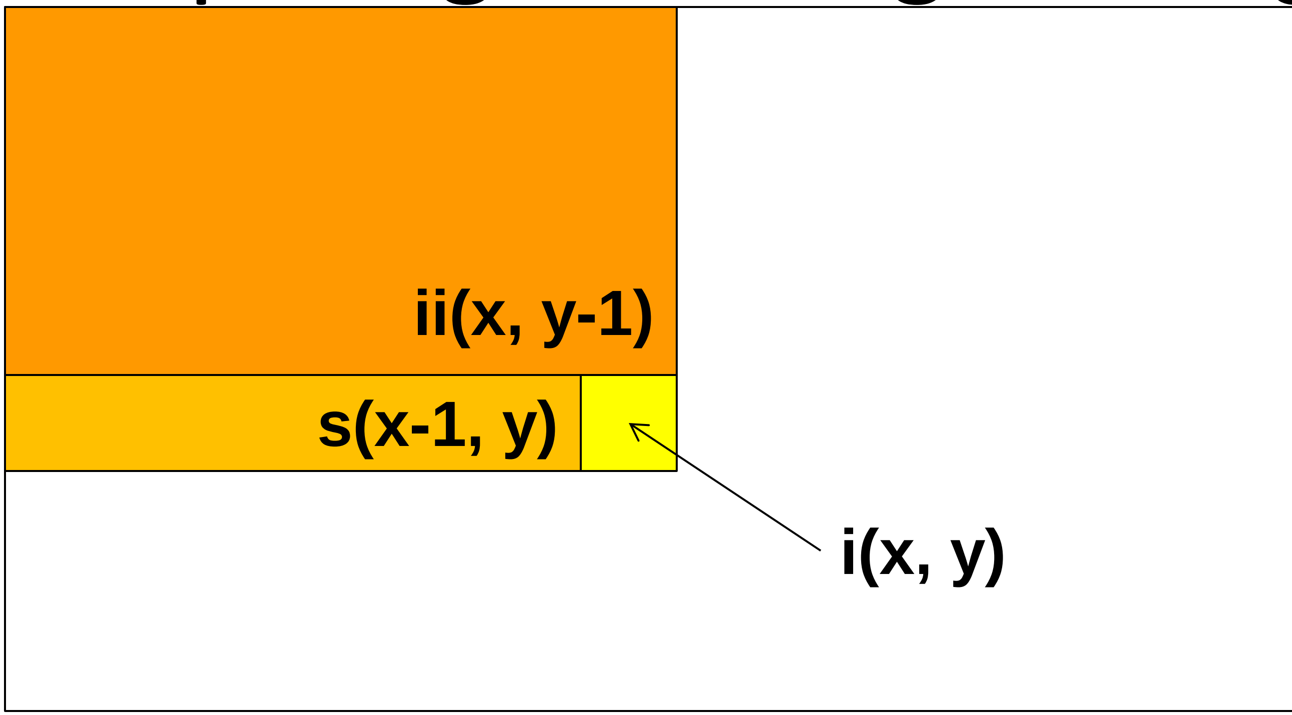
- The *integral image* computes a value at each pixel (*x*,*y*) that is the sum of the pixel values above and to the left of (*x*,*y*), inclusive
- This can quickly be computed in one pass through the image



Computing the integral image



Computing the integral image



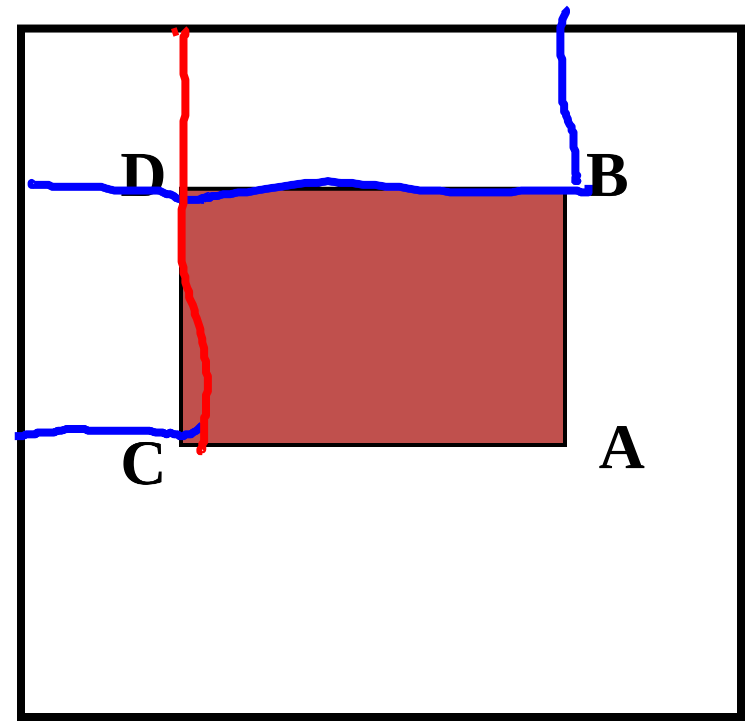
- Cumulative row sum: s(x, y) = s(x–1, y) + i(x, y)
- Integral image: ii(x, y) = ii(x, y–1) + s(x, y)

MATLAB: ii = cumsum(cumsum(double(i)), 2);

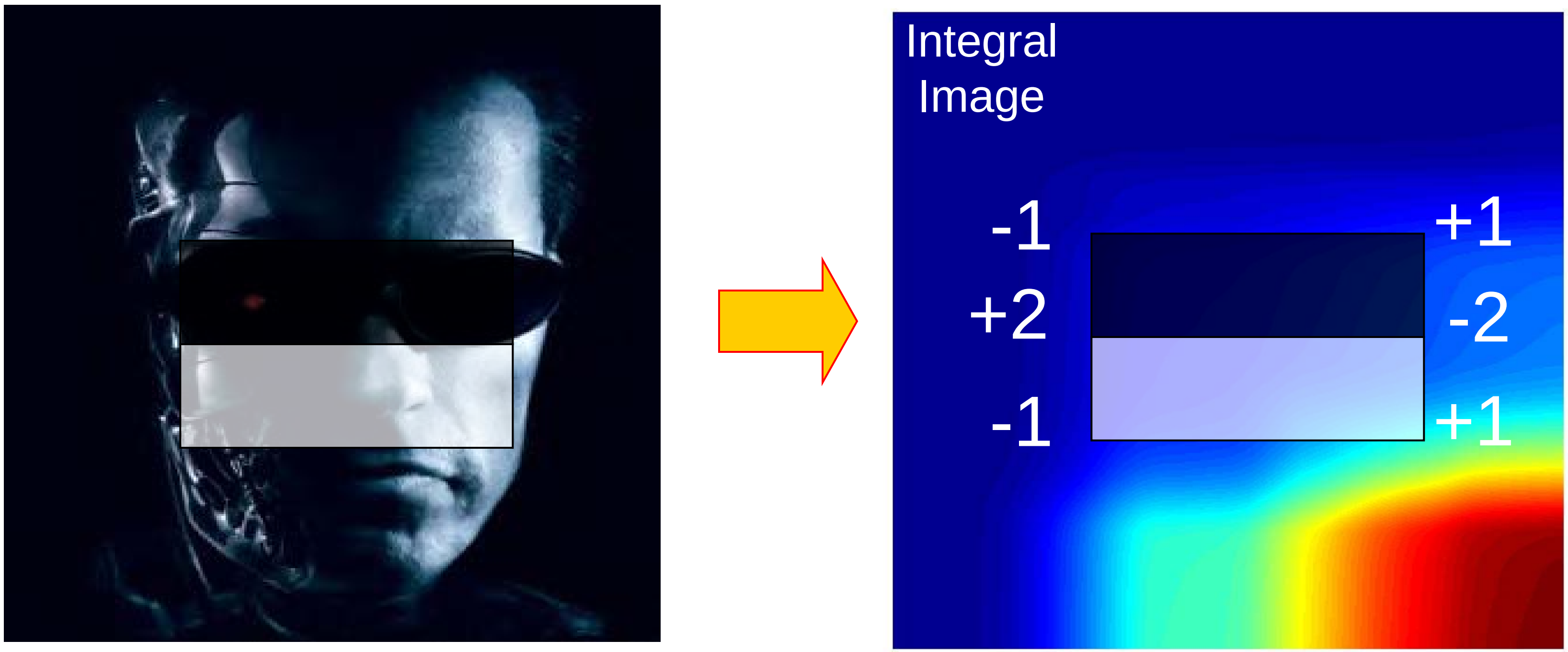
Efficiently Computing Rectangle Sum

- 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!

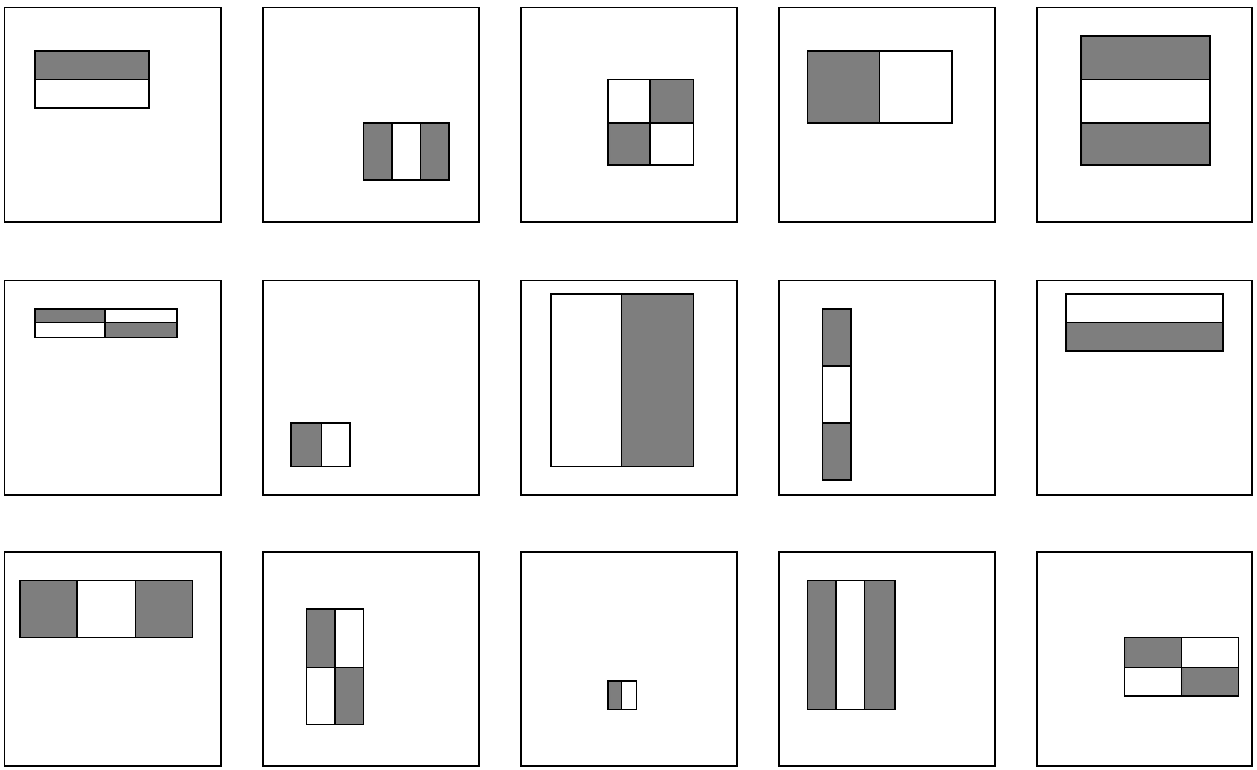


Evaluating Rectangle Features



Feature selection

- For a 24x24 detection region, the number of possible rectangle features is ~160,000!



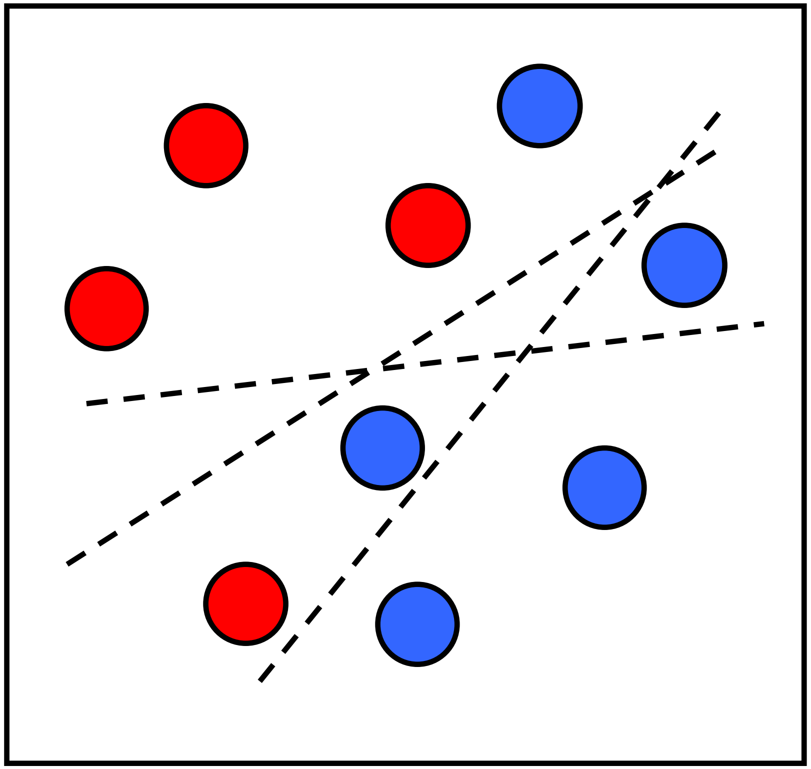
Feature selection

- For a 24x24 detection region, the number of possible rectangle features is ~160,000!
- 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?
  - Viola-Jones Key Idea #2
    - We can select the informative features **and** build a classifier

Supervised Learning: Classification

- Goal: Learn a decision boundary given labeled examples

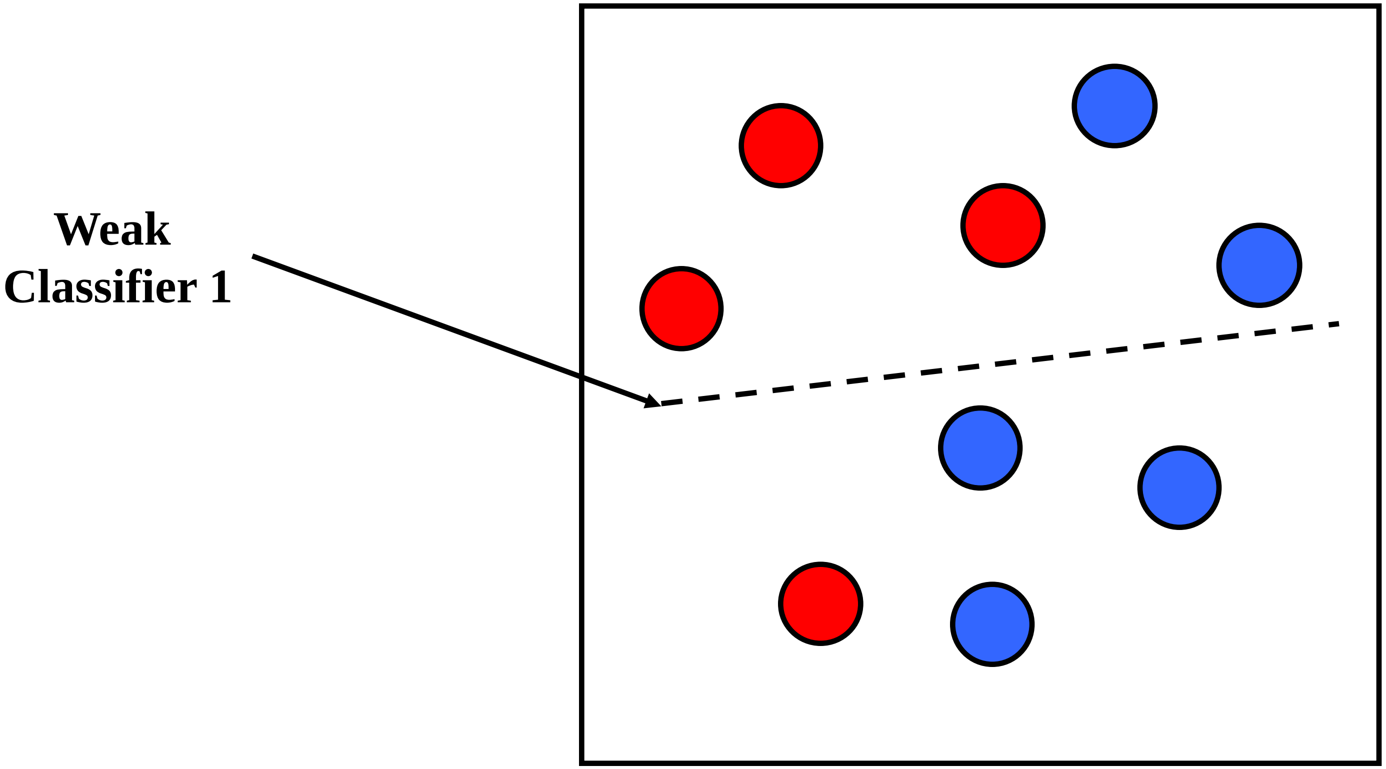
- Toy example
  - Can only use linear classifiers (lines)
  - How to separate red and blue examples?



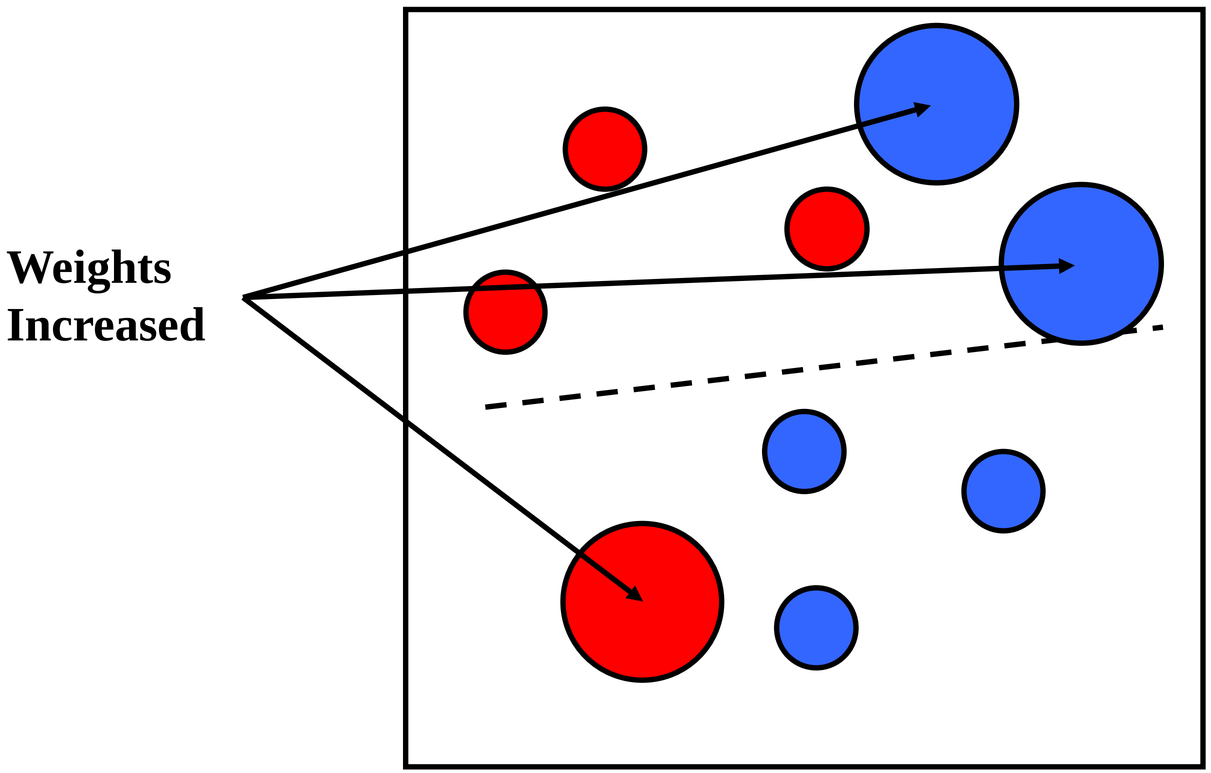
Boosting (Main Idea)

- Train classifiers in a sequence
  - a new classifier should focus on those cases which were incorrectly classified in the last round
  - combine the classifiers by letting them vote on the final prediction
  - each classifier should be very “weak”
    - Where “weak” means slightly better than random guess

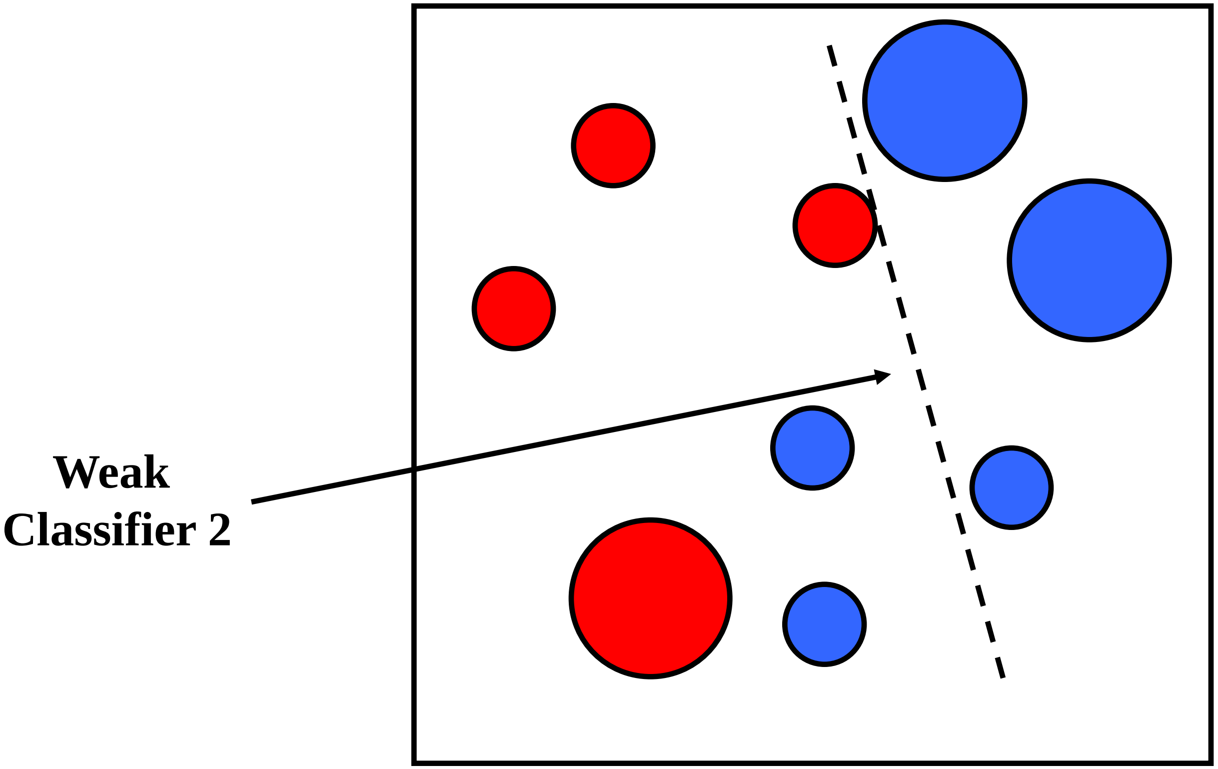
Boosting intuition



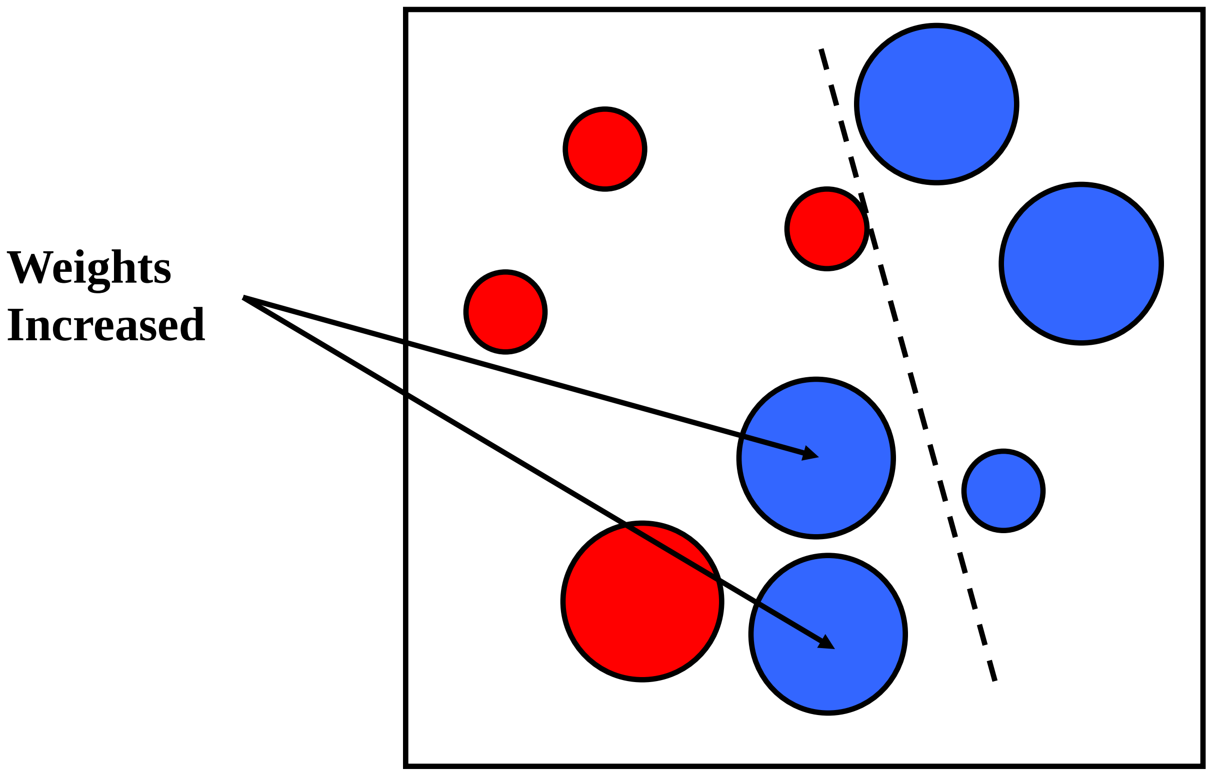
Boosting illustration



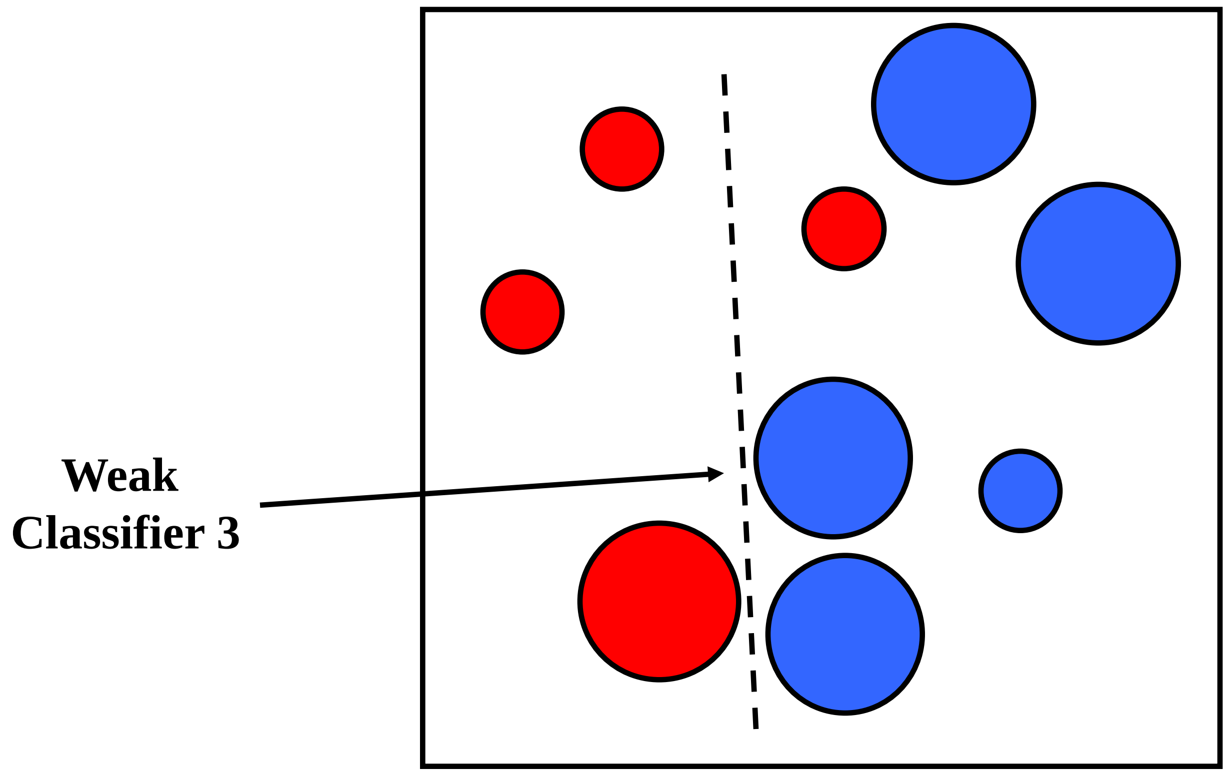
Boosting illustration



Boosting illustration

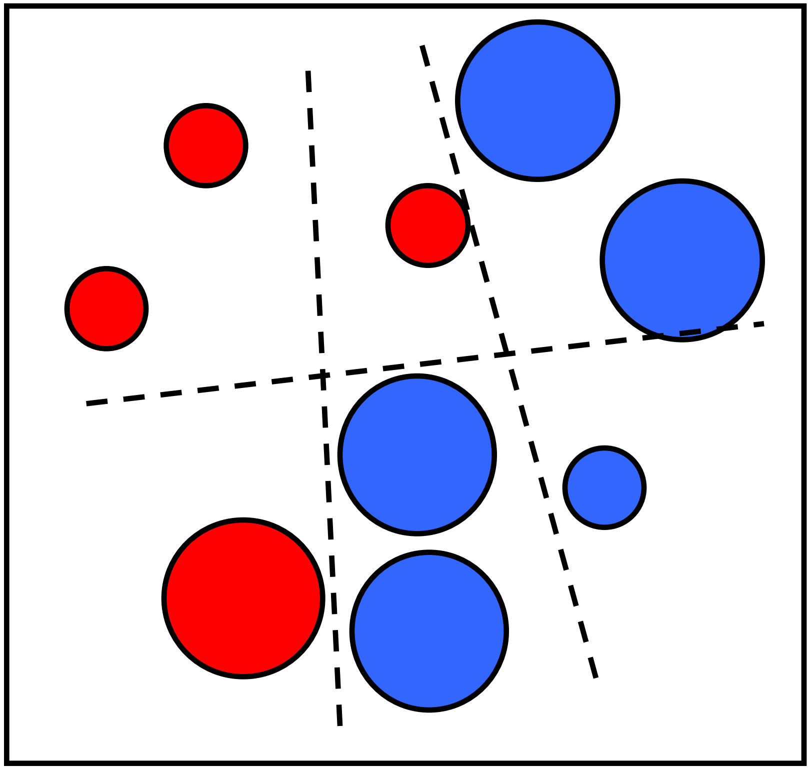


Boosting illustration



Boosting illustration

Final classifier is a combination of weak classifiers



Boosting: Training

- Initially, weight each training example equally
- In each boosting round:
  - Find the weak learner that achieves the lowest *weighted* training error
  - Raise weights of training examples misclassified by current weak learner
- Compute final classifier as linear combination of all weak learners (weight of each learner is directly proportional to its accuracy)

- *Exact formulas for re-weighting and combining weak learners depend on the particular boosting scheme (e.g., AdaBoost)*

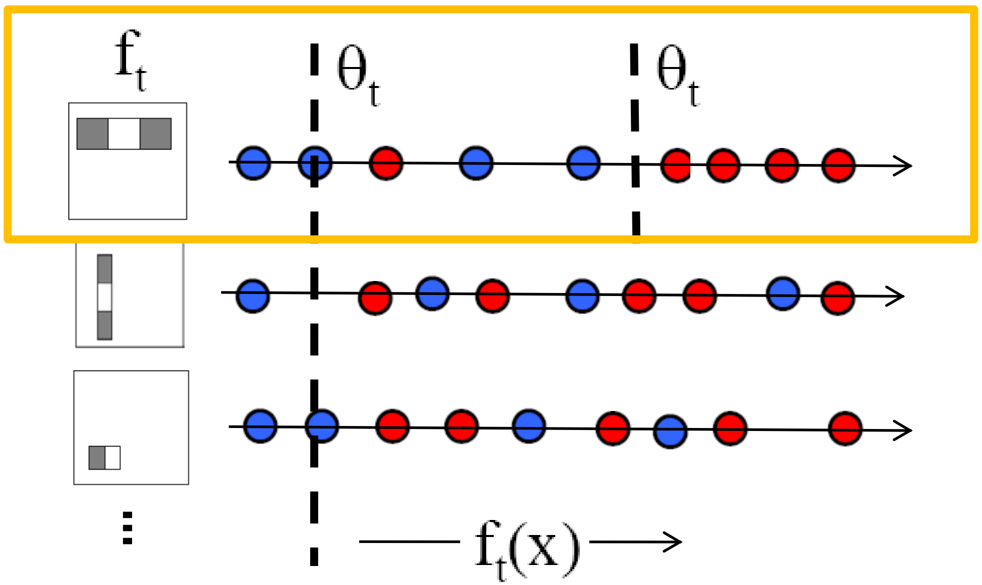
Boosting: Pros and Cons

- Advantages of boosting
  - Integrates classification with feature selection
  - Complexity of training is linear in the number of training examples
  - Flexibility in the choice of weak learners, boosting scheme
  - Testing is fast
  - Easy to implement

- Disadvantages
  - Needs many training examples
  - Often found not to work as well as an alternative discriminative classifier, support vector machine (SVM)
    - especially for many-class problems

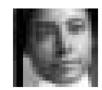
Viola-Jones: Boosting

- **Want to select the single rectangle feature and threshold that best separates positive (faces) and negative (non-faces) training examples, in terms of *weighted* error.**



Outputs of a possible rectangle feature on faces and non-faces.

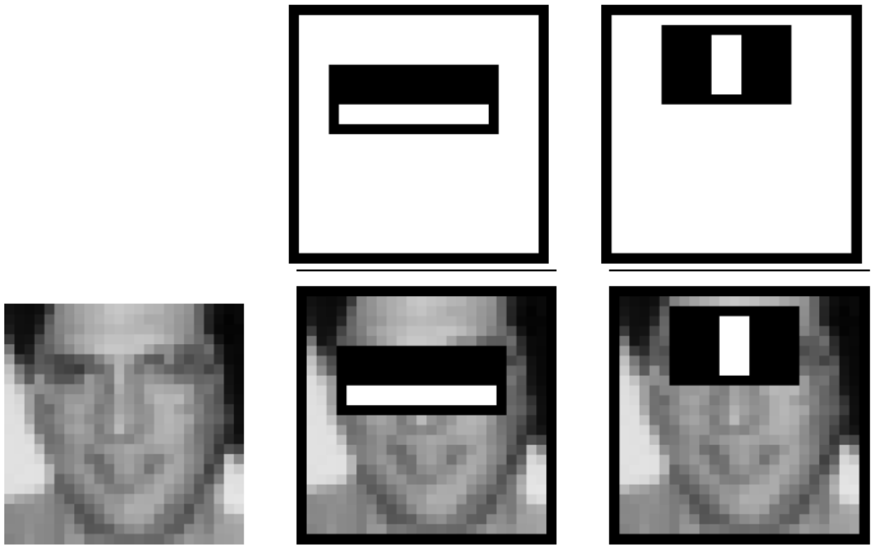
Resulting weak classifier:



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

For next round, reweight the examples according to errors, choose another filter/threshold combo.

Viola-Jones Face Detector: Results

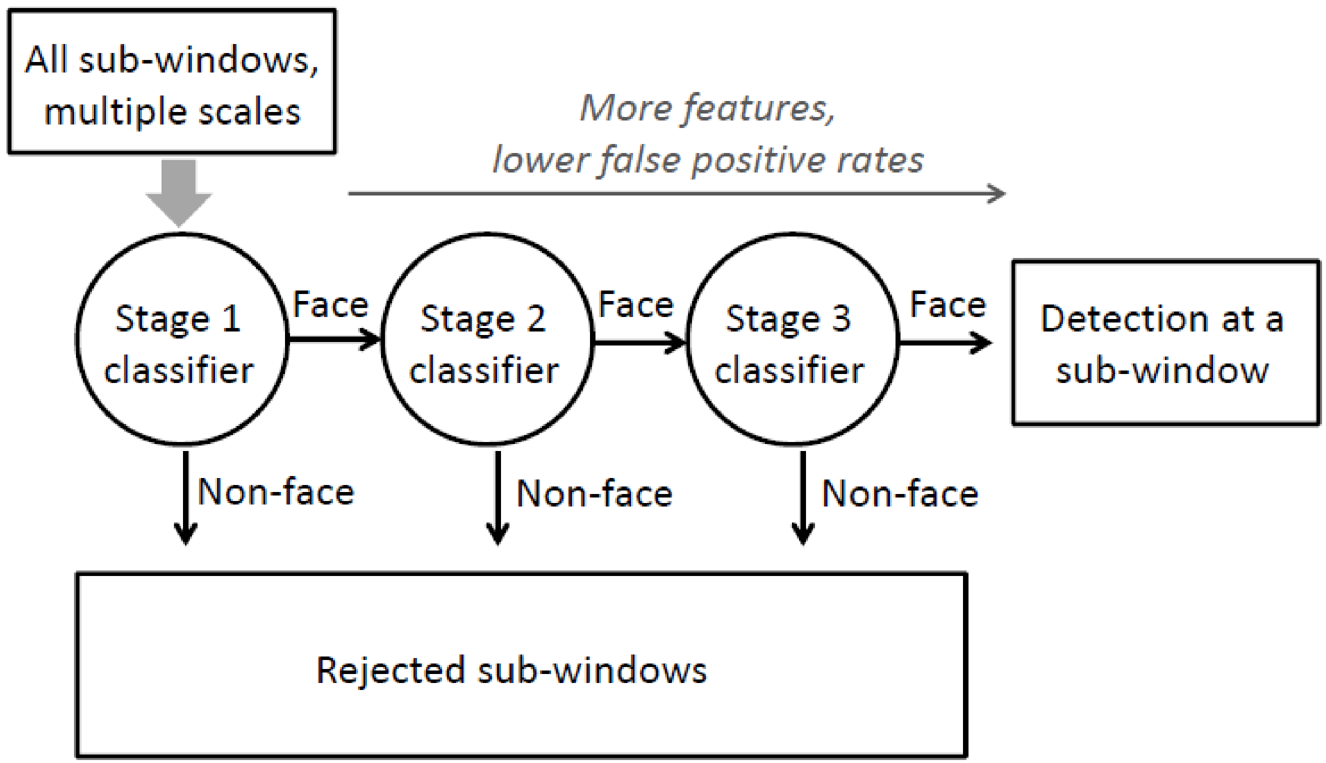


First two features selected

Even if the filters are fast to compute, each new image has a **lot** of possible windows to search.

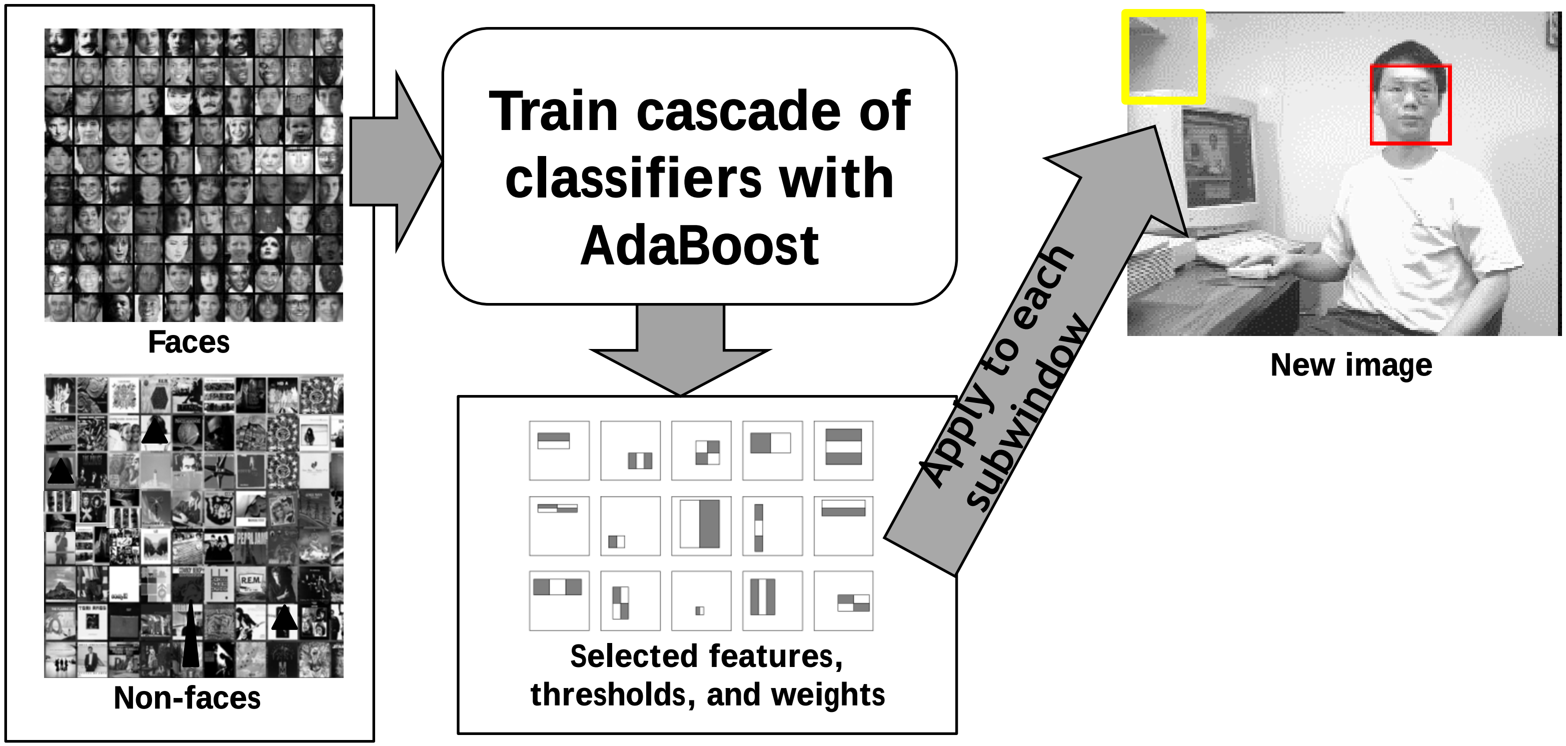
How can detection be made more efficient?

Cascading classifiers for detection



- Form a *cascade* with low false negative rates early on
- Apply less accurate but faster classifiers first to immediately discard windows that clearly appear to be negative

Viola-Jones detector: summary



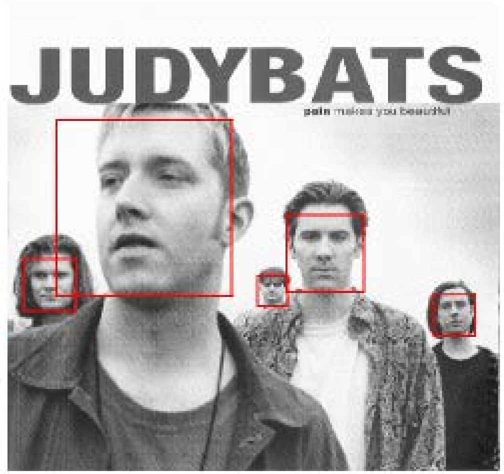
Train with 5K positives, 350M negatives

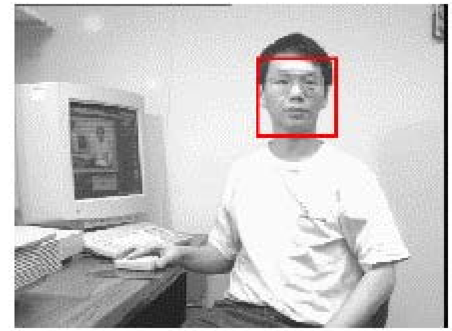
Real-time detector using 38 layer cascade

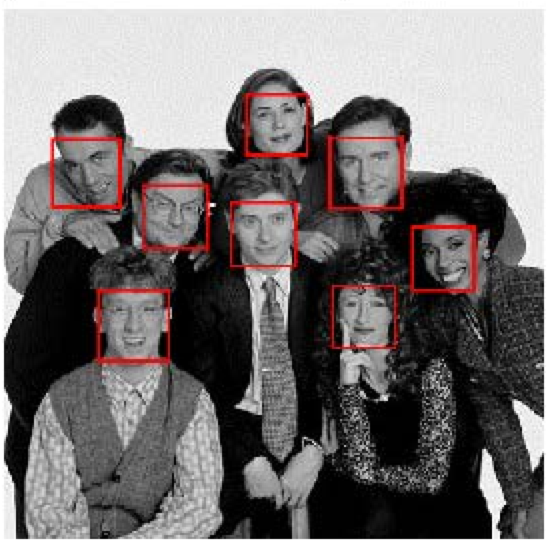
6061 features in all layers

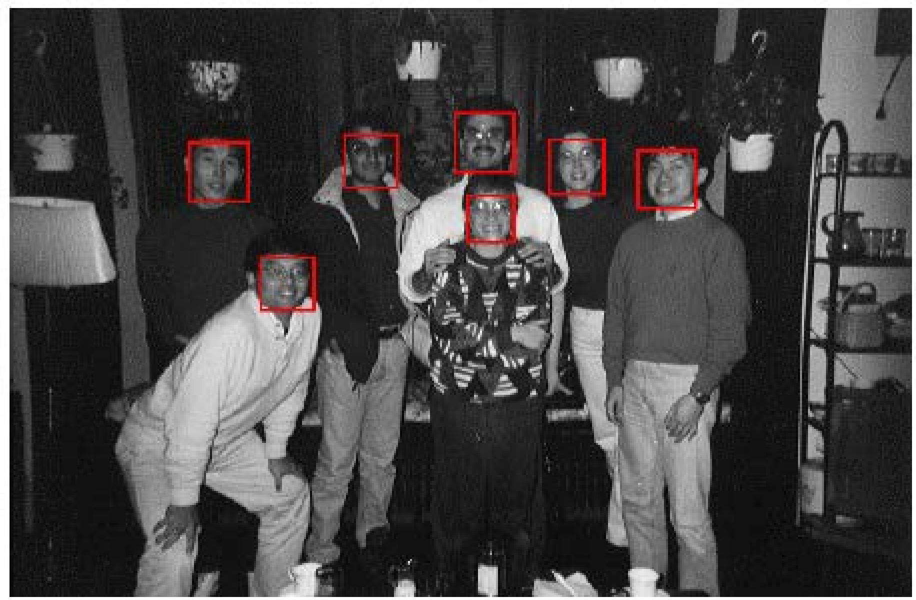
[Implementation available in OpenCV]

Viola-Jones Face Detector: Results

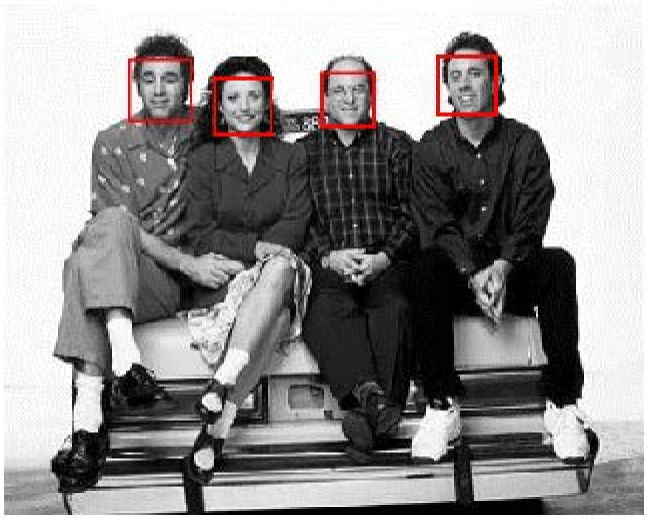


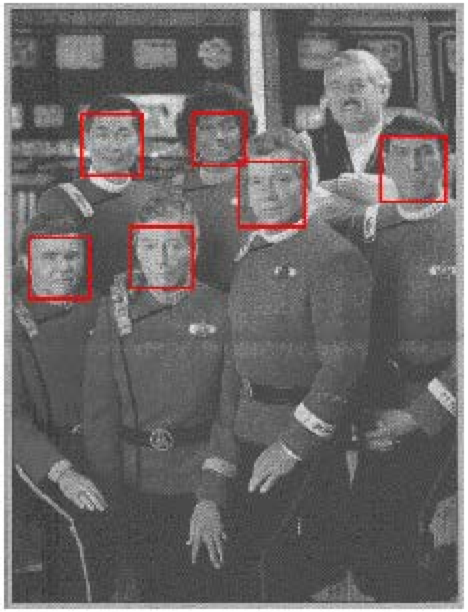


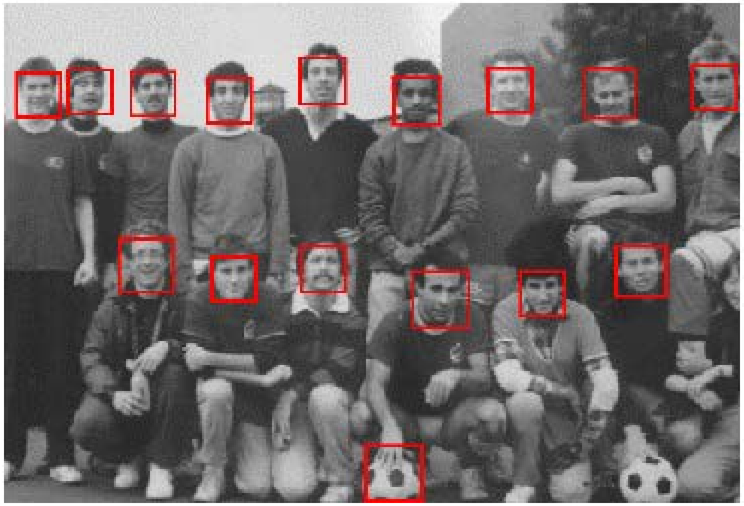


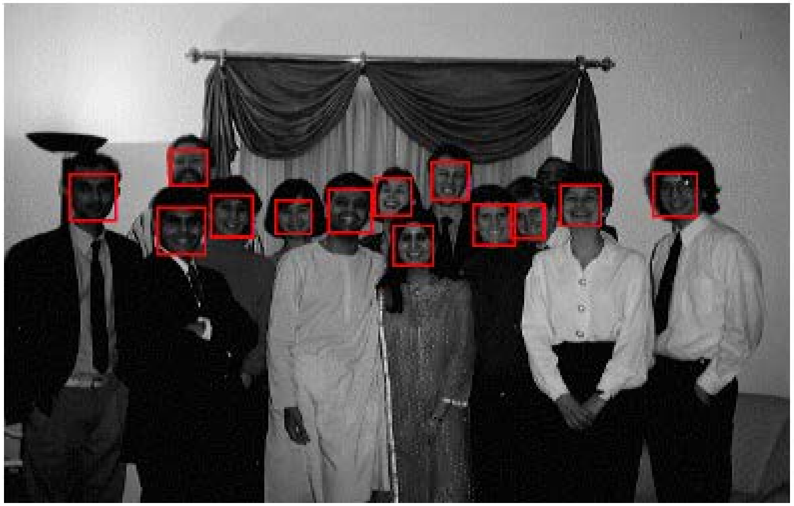


Viola-Jones Face Detector: Results





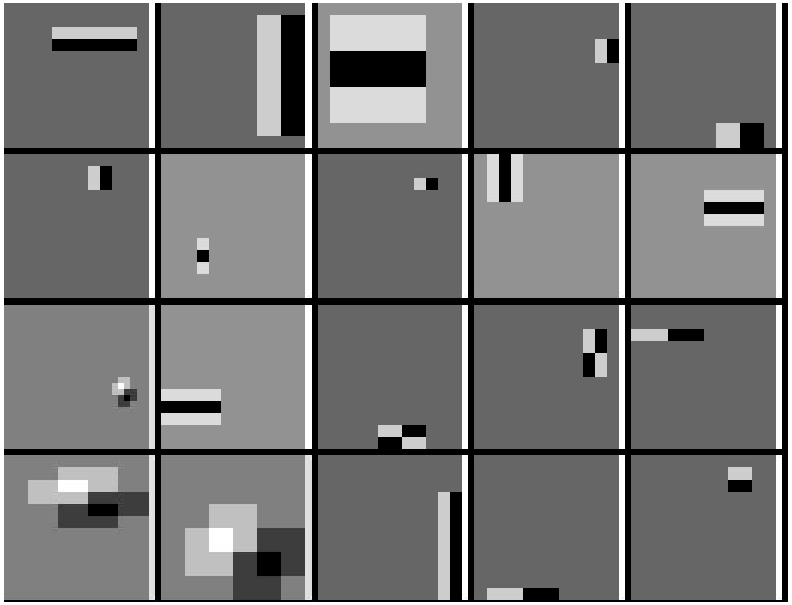




Viola-Jones Face Detector: Results

Detecting profile faces?

Can we use the same detector?



Viola-Jones Face Detector: Results

Google now erases faces, license plates on Map Street View

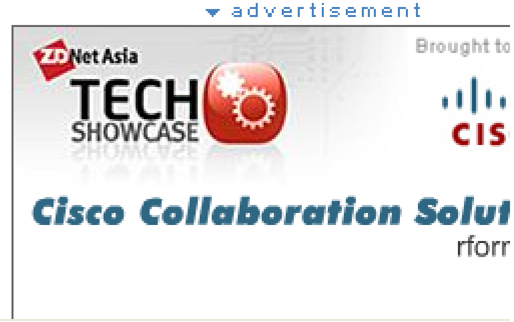
By Elinor Mills, CNET News.com
Friday, August 24, 2007 01:37 PM

Google has gotten a lot of flack from privacy advocates for photographing faces and license plate numbers and displaying them on the Street View in Google Maps. Originally, the company said only people who identified themselves could ask the company to remove their image.

But Google has quietly changed that policy, partly in response to criticism, and now anyone can alert the company and have an image of a license plate or a recognizable face removed, not just the owner of the face or car, says Marissa Mayer, vice president of search products and user experience at Google.

"It's a good policy for users and also clarifies the intent of the product," she said in an interview following her keynote at the Search Engine Strategies conference in San Jose, Calif., Wednesday.

The policy change was made about 10 days after the launch of the product in late May, but was not publicly announced, according to Mayer. The company is removing images only when someone notifies them and not proactively, she said. "It was definitely a big policy change inside."



Consumer application: iPhoto 2009

http://www.apple.com/ilife/iphoto/

Slide credit: Lana Lazebnik

Consumer application: iPhoto 2009

- 

Slide credit: Lana Lazebnik

Viola-Jones Detector: Summary

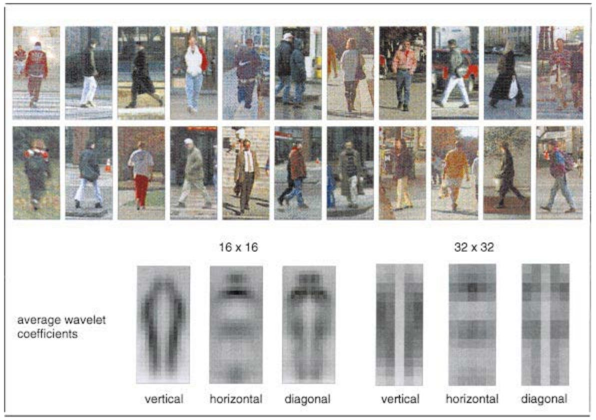
- A seminal approach to real-time object detection
- Training is slow, but detection is very fast
- Key ideas
  - *Integral images* for fast feature evaluation
  - *Boosting* for feature selection
  - *Attentional cascade* of classifiers for fast rejection of non-face windows

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

P. Viola and M. Jones. *Robust real-time face detection.* IJCV 57(2), 2004.

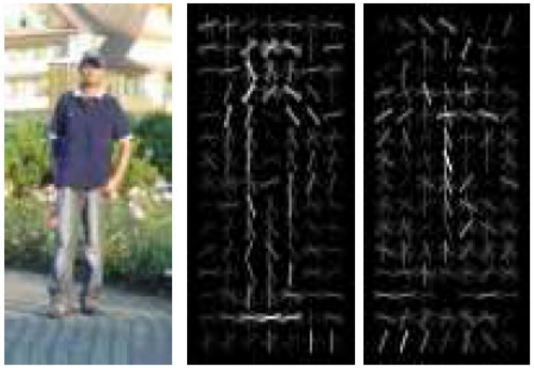
Window-Based Detection

- Other categories are amenable to window-based detection
- Detecting upright, walking humans
  - Using sliding window's appearance/texture; e.g.,



SVM with Haar wavelets [Papageorgiou & Poggio, IJCV 2000]

Space-time rectangle features [Viola, Jones & Snow, ICCV 2003]



SVM with HoGs [Dalal & Triggs, CVPR 2005]

Window-Based Detection: Strengths

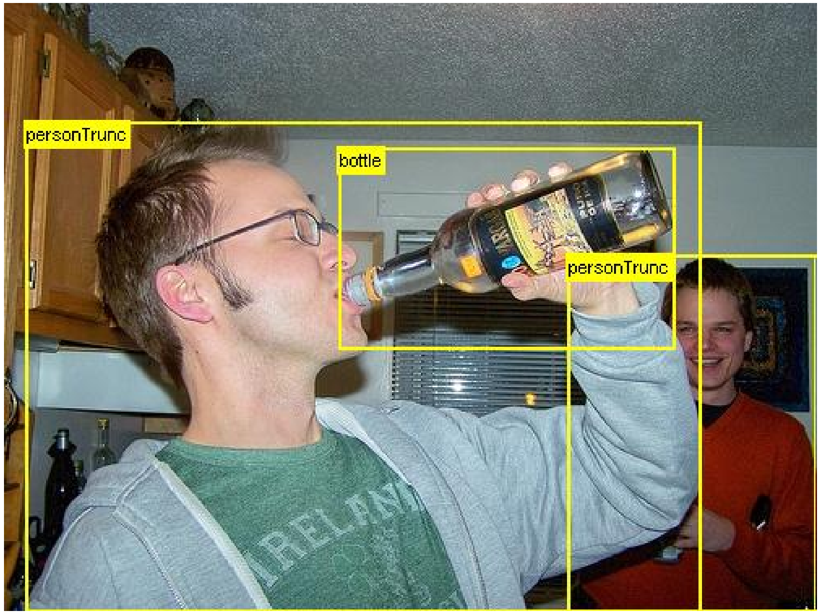
- Sliding window detection and global appearance descriptors:
  - Simple detection protocol to implement
  - Good feature choices critical
  - Past successes for certain classes

Window-Based Detection: Limitations

- High computational complexity
  - For example: 250,000 locations x 30 orientations x 4 scales = 30,000,000 evaluations!
  - If training binary detectors independently, means cost increases linearly with number of classes
- With so many windows, false positive rate better be low

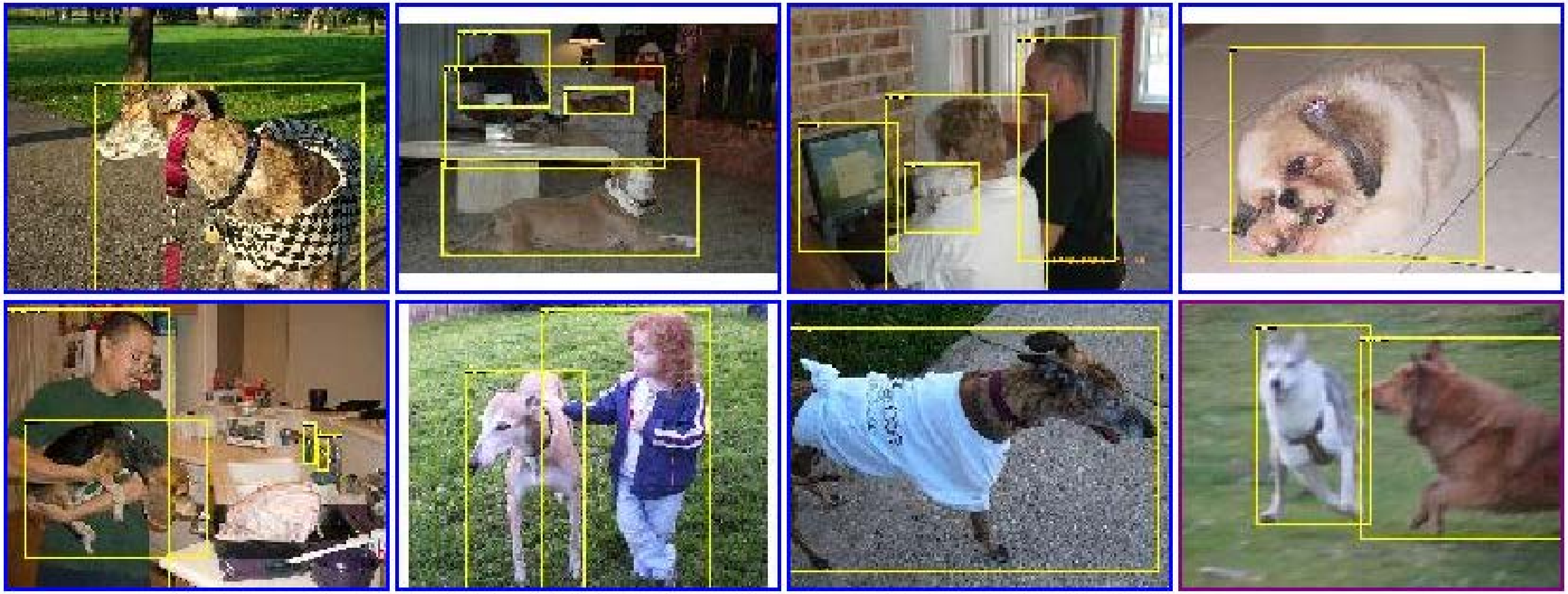
Limitations (continued)

- Not all objects are “box” shaped



Limitations (continued)

- Non-rigid, deformable objects not captured well with representations assuming a fixed 2d structure; or must assume fixed viewpoint



Kristen Grauman

Limitations (continued)

- If considering windows in isolation, context is lost

Sliding window

Limitations (continued)

- In practice, often entails large, cropped training set (expensive)
- Global (window) descriptor can lead to sensitivity to partial occlusions

Summary

- Basic pipeline for window-based detection
  - Model/representation/classifier choice
  - Sliding window and classifier scoring
- Viola-Jones face detector
  - Exemplar of basic paradigm
  - Key ideas: rectangular features, boosting for feature selection, cascade classification
- Now that we can detect faces, can we say who it is?
  - Next class…