UNIVERSITÀ DEGLI STUDI DI PADOVA

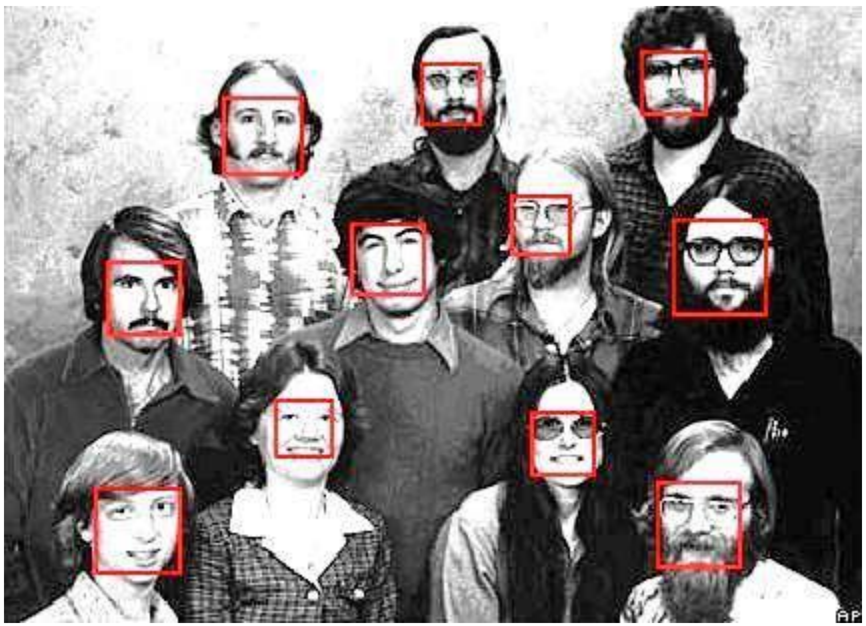
Face detection: the Viola-Jones approach

Stefano Ghidoni

Agenda

- The face detection problem
- The Viola-Jones approach
- Boosting and cascading

The task of face detection



Sliding window approach

- Window moving over the image
- Find matching locations

Challenges

- Huge number of pixels
- Multiple locations and scales
  - All combinations should be evaluated!
- Faces are an unlikely event: ~0-10 per image
- Key elements for an optimal face detector:
  - Fast processing for non-face candidates
  - Very low false positive rate: $<10^{-6}$ is required to avoid a false positive/image

Viola & Jones face detector: overview

- The Viola & Jones detector is a popular approach to face detection
  - Widely used approach to fast object detection
  - Particularly effective for face detection

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

Viola & Jones face detector: overview

- Key elements:
  - Haar features
    - Fast feature evaluation based on the **integral image**
      - In which context did we see the integral image? What are the advantages/disadvantages?

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

Viola & Jones face detector: overview

- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

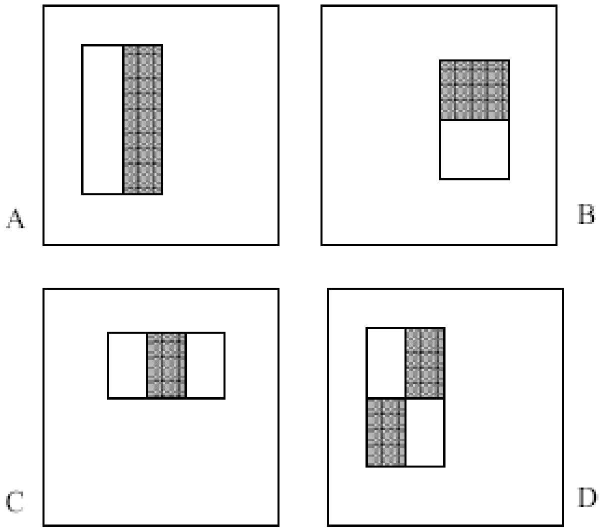
P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features." CVPR 2001
P. Viola and M. Jones, "Robust real-time face detection." IJCV 57(2), 2004

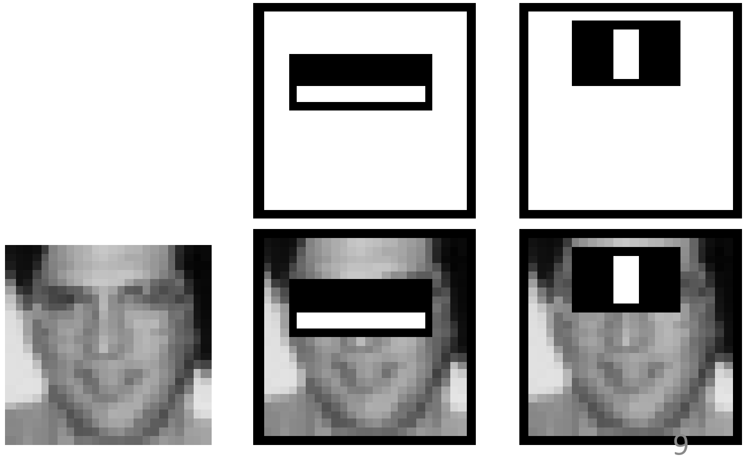
Haar features

- Rectangular filters
- Local feature: subtract the sum of pixels in the white areas from the sum of pixels in the black area

$$f(\boldsymbol{x}) = \sum_i p_b(i) - \sum_i p_w(i)$$

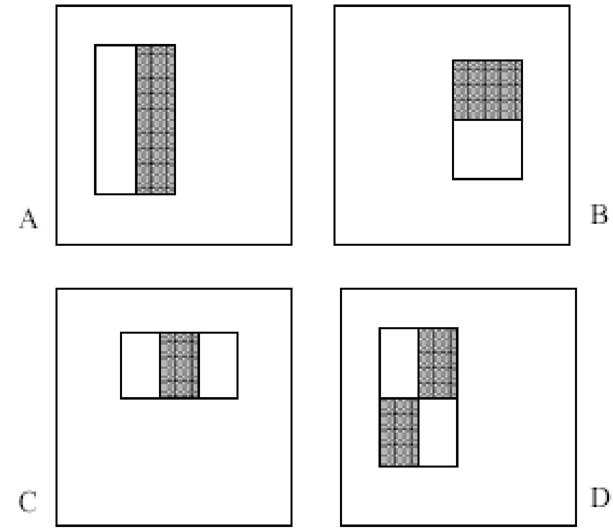
- 2-, 3- and 4-rectangle features
- Huge number of features!

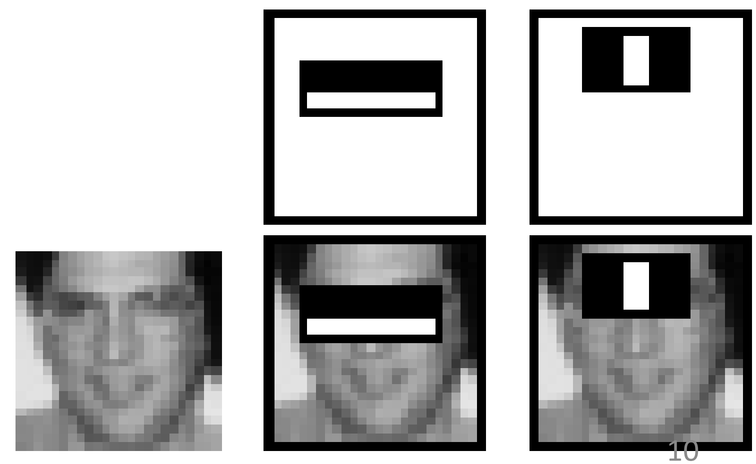




Haar features

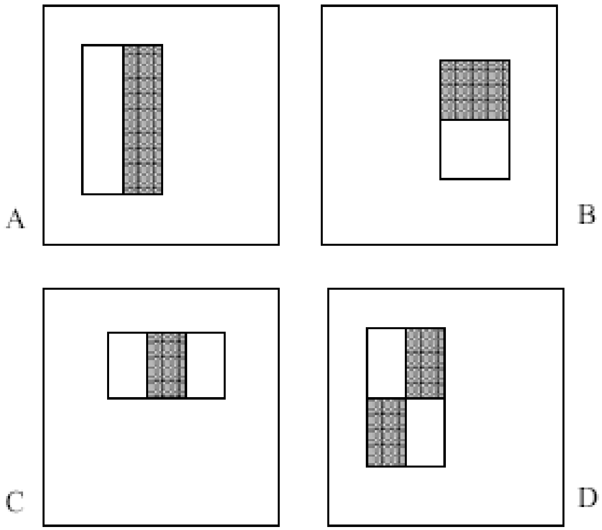
- Coarse features
- Sensitive to:
  - Edges
  - Bars
  - Other simple structures
- But: computationally efficient
  - A large number of features can be computed
  - Compensates for the coarseness

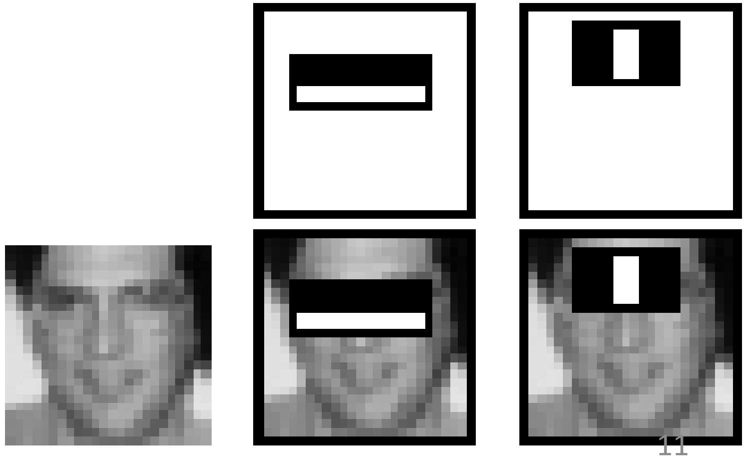




How many features?

- Consider a 24×24 patch
- 160k possible features!
- Huge number of combinations: an exhaustive analysis is not feasible





Viola & Jones face detector: overview

- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

Viola & Jones face detector: overview

- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

Weak learners

- A weak learner works on the number evaluated by a Haar feature f_j
- Sets a threshold on a single feature
- Separates positive and negative examples (slightly) better than casual guess

Weak learners

- A weak learner works on the number evaluated by a Haar feature f_j
- It is evaluated as:

$$h_j(\boldsymbol{x}) = \begin{cases} 1 & if\ p_j f_j(\boldsymbol{x}) > p_j \theta_j \\ -1 & \text{otherwise} \end{cases}$$

Weak learners

- A weak learner works on the number evaluated by a Haar feature f_j
- It is evaluated as:

Threshold

$$h_j(\boldsymbol{x}) = \begin{cases} 1 & if\ p_j f_j(\boldsymbol{x}) > p_j \theta_j \\ -1 & \text{otherwise} \end{cases}$$

Window of the Haar feature

Parity term $\in \{-1,1\}$
(may change inequality)

Value of the Haar feature

Viola & Jones face detector: overview

- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

Viola & Jones face detector: overview

- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

Boosting

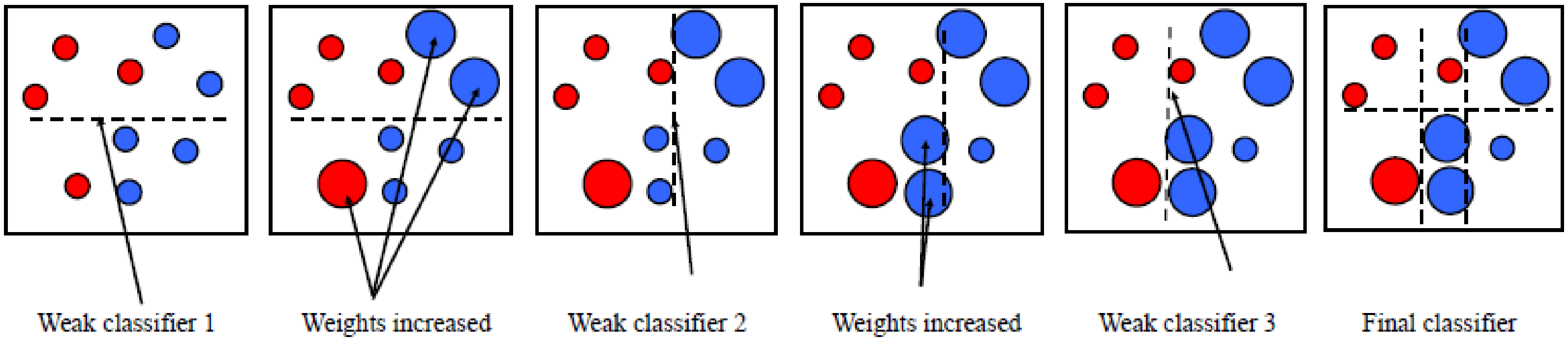
- Build a strong classifier combining several weak classifiers
- Weighted sum of simple weak learners

$$h(\boldsymbol{x}) = sign\left[\sum_{j=0}^{m-1} a_j h_j(\boldsymbol{x})\right]$$

- Weights selected depending on the classifier accuracies
- AdaBoost used to select the features and train the classifier

Adaboost

- Initial condition: equal weights
- For each round of boosting:
  - Evaluate each rectangle filter on each example
  - Select the best threshold for each filter
  - Select the best filter/threshold combination
  - Reweight examples



Adaboost

- At each round, the best weak classifier is found
  - The best classifier after the effect of the previously selected classifiers
- Computational complexity: O(MNK)
  - M rounds, N examples, K features

Viola & Jones face detector: overview

- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

Viola & Jones face detector: overview

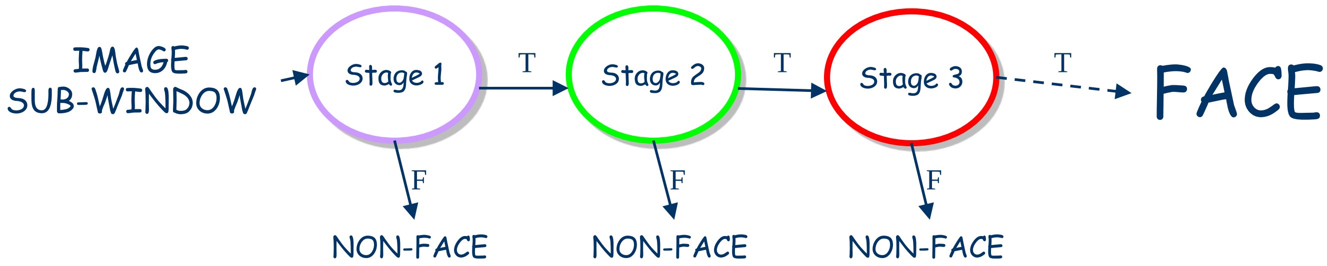
- Key elements:
  - Haar features
  - Weak learners working on Haar features
  - Boosting
  - Cascading for organizing classifiers and fast rejection of non-matching windows

Cascade of classifiers

- The classifier is a combination of weak learners
- The classifier becomes more efficient if weak learners are divided into different stages
  - Stages are applied in sequence (*cascade*)
  - Each stage acts as a filter
  - Each stage makes use of several weak learners

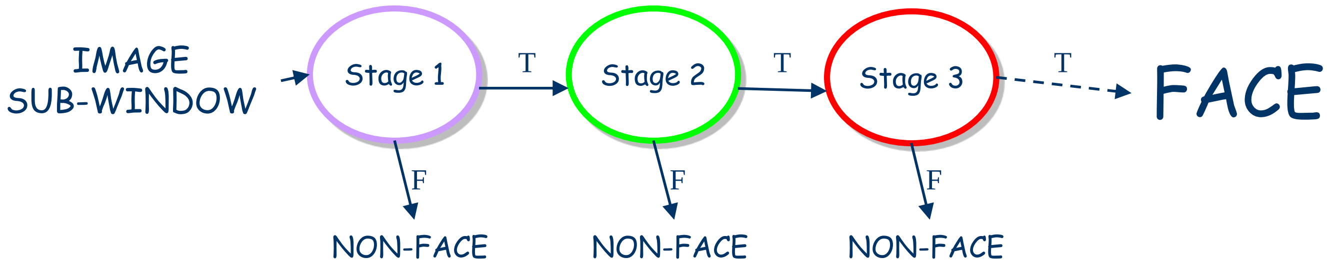
Cascade of classifiers

- Each stage acts as a filter
  - The first stage that discards a sample prevents the subsequent stages to work
  - False negatives cause a failure
  - False positives are acceptable at a high rate
- First filters shall be as fast as possible



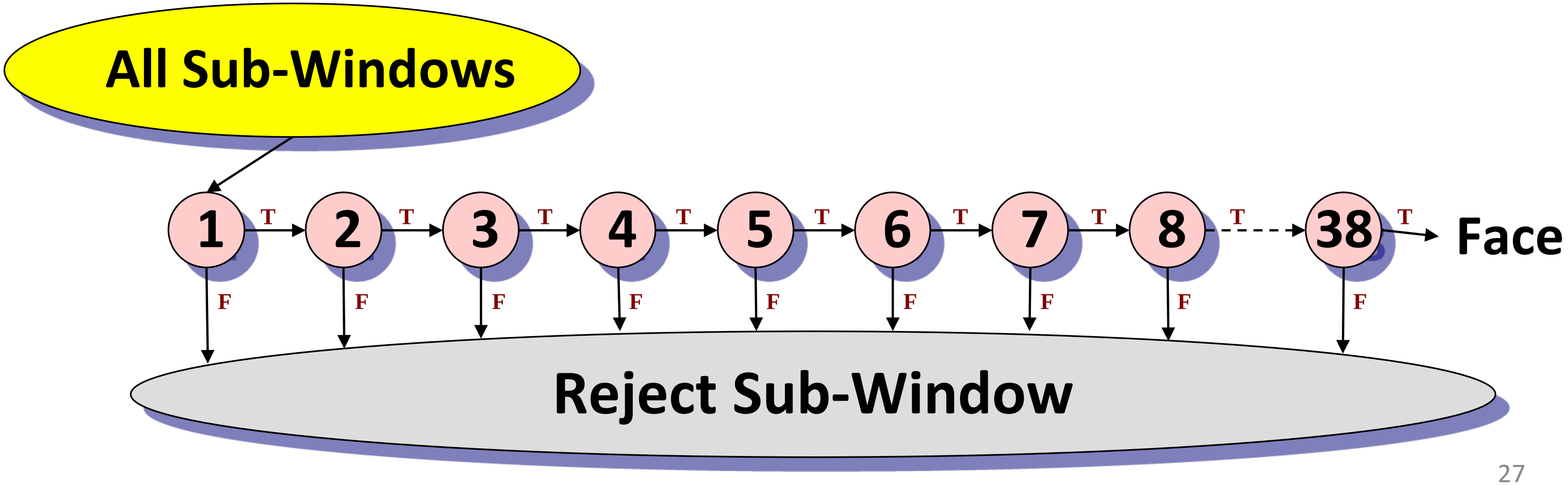
Cascade of classifiers

- Train several classifiers with an increasing number of features until the target decision rate is reached
- Re-weight training examples after each stage giving an higher weight to samples wrongly classified in previous stages
- Classifiers are progressively more complex and have lower false positive rates

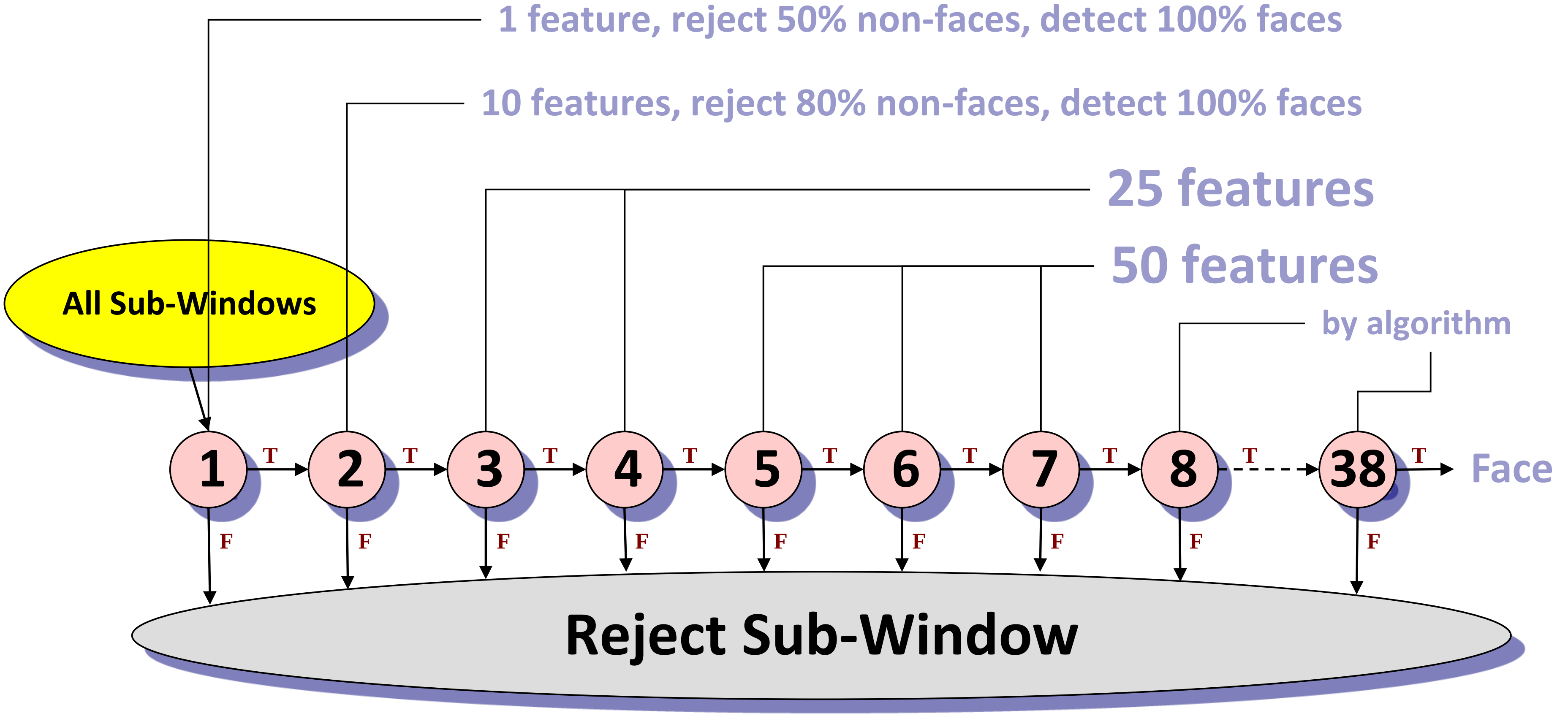


Structure of the cascade

- Structure used by Viola and Jones
  - 38 stages / layers
  - 6061 features



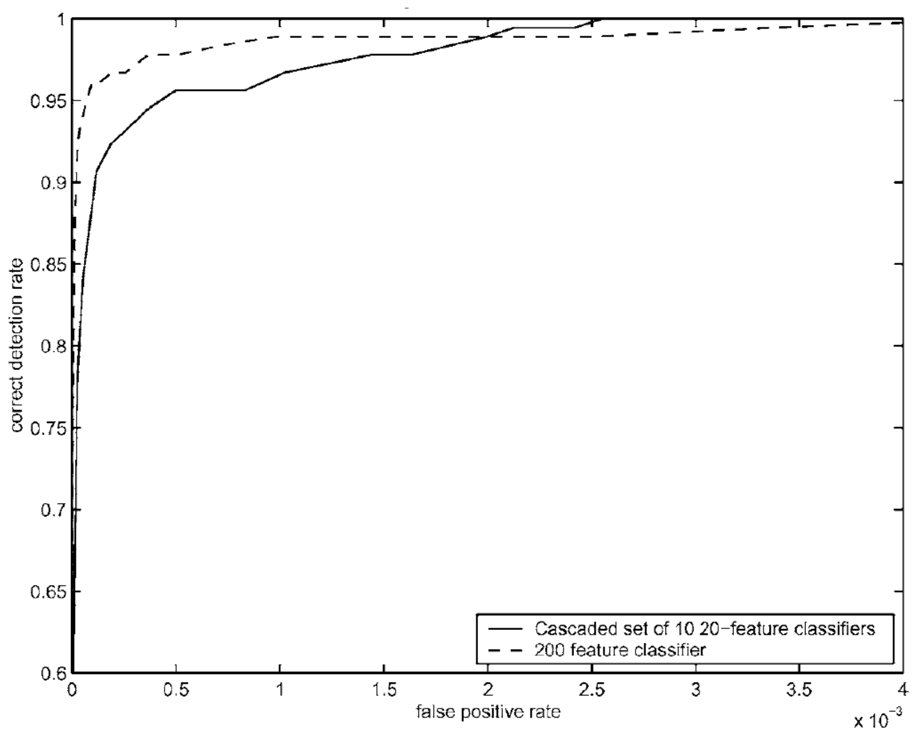
Structure of the cascade



Cascaded vs monolithic classifier

- Similar accuracy
- 10× difference in processing time!

ROC di dl, sensitivity su 1-specificity



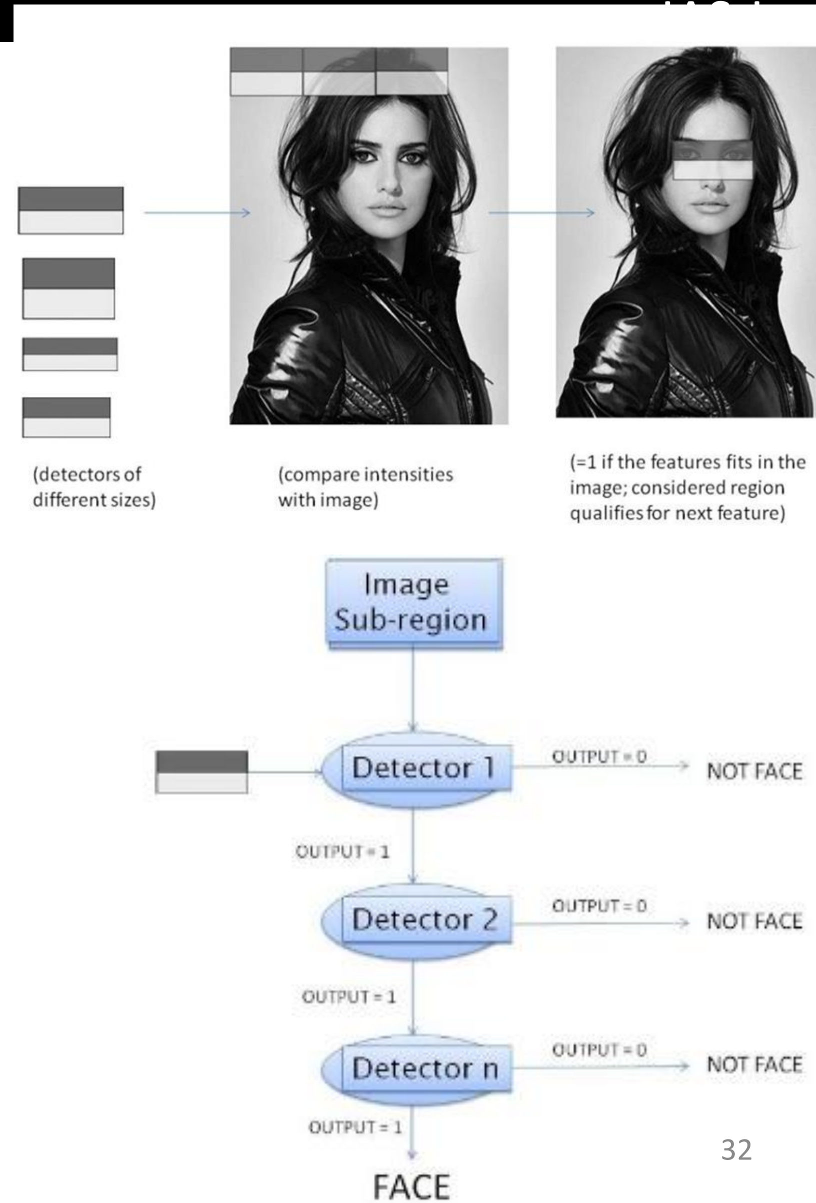
Dataset

- Work by Viola-Jones used the following training data:
  - 5k faces
    - All frontal, rescaled to 24×24 pixels
  - 10k non-faces (found in a collection of 350M)
- Variation factors:
  - Individuals
  - Illumination
  - Pose

Examples

Università degli Studi di Padova

Example



Feature selection example

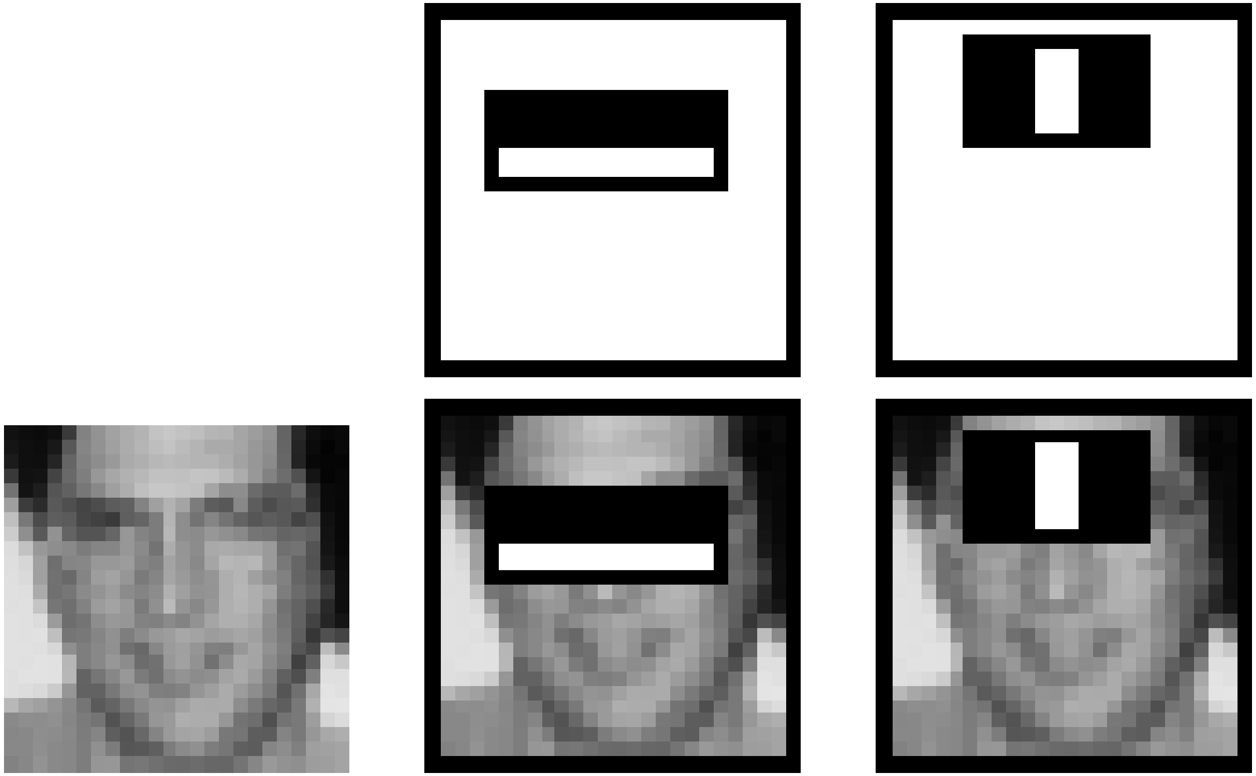
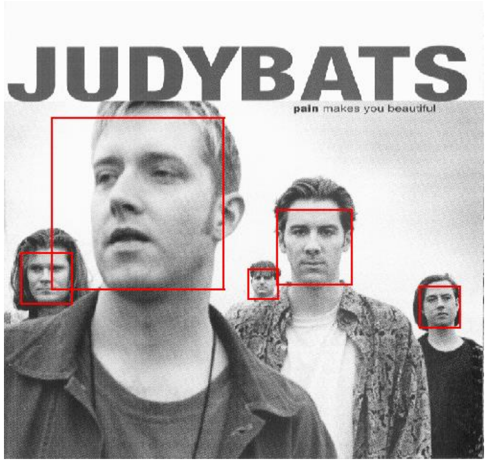
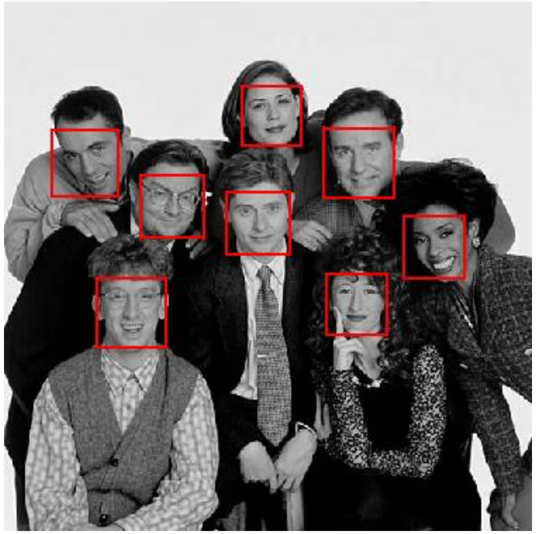


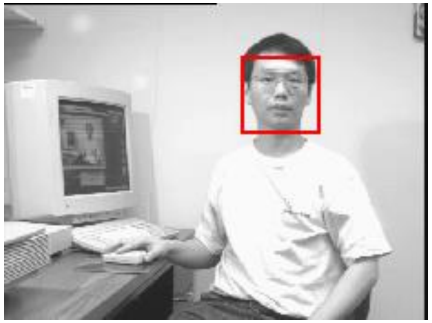
Figure 5: The first and second features selected by AdaBoost. The two features are shown in the top row and then overlayed on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.

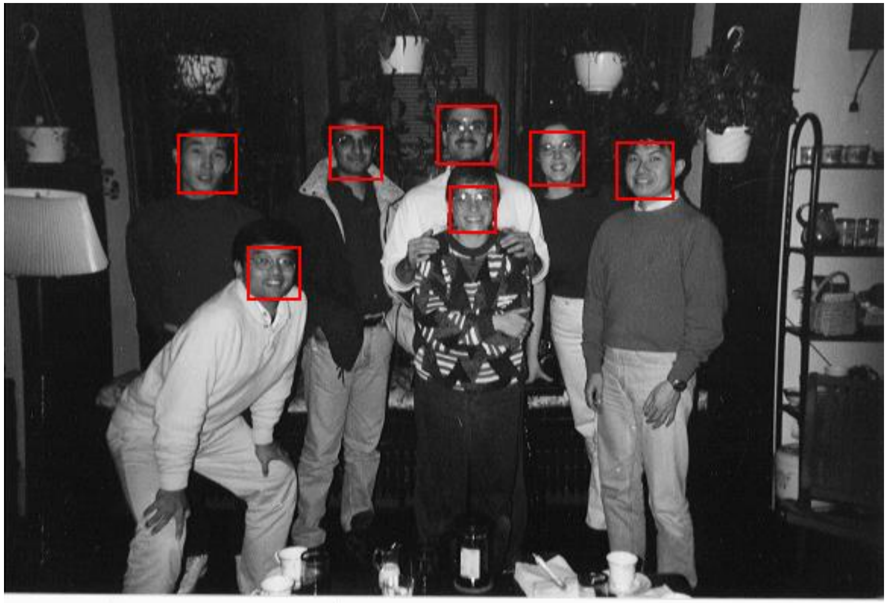
Viola and Jones: examples

(profile face)







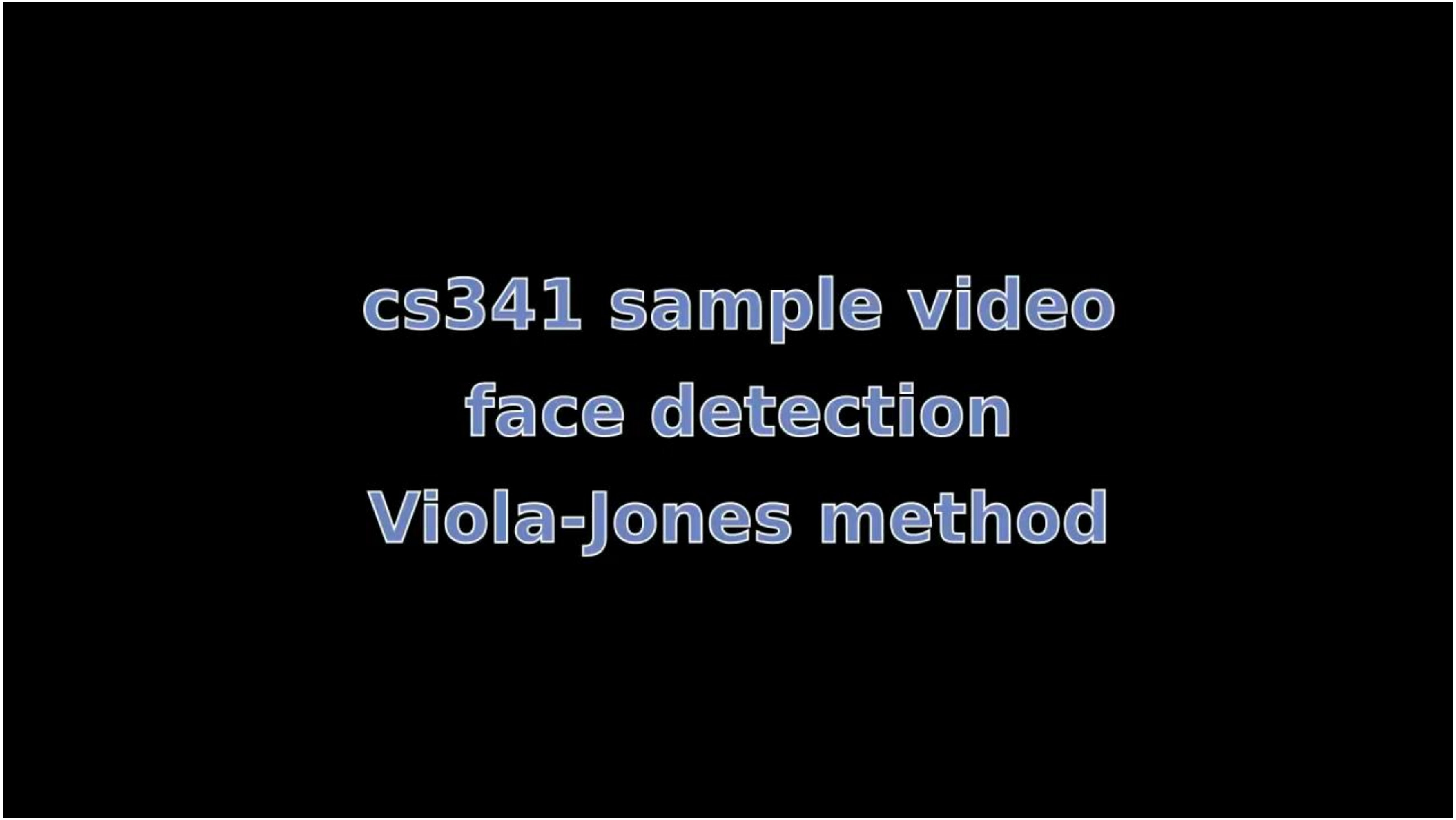


Viola and Jones: examples

Missed faces (false negatives)

Fake faces (false positives)

Viola and Jones: examples



UNIVERSITÀ DEGLI STUDI DI PADOVA

Face detection: the Viola-Jones approach

Stefano Ghidoni