
29

Rapid Object Detection

Boosted Cascade Classifiers (*Viola and Jones, 2001*)

31.01.2019

Severin Angerpointner
LMU Munich

General Task

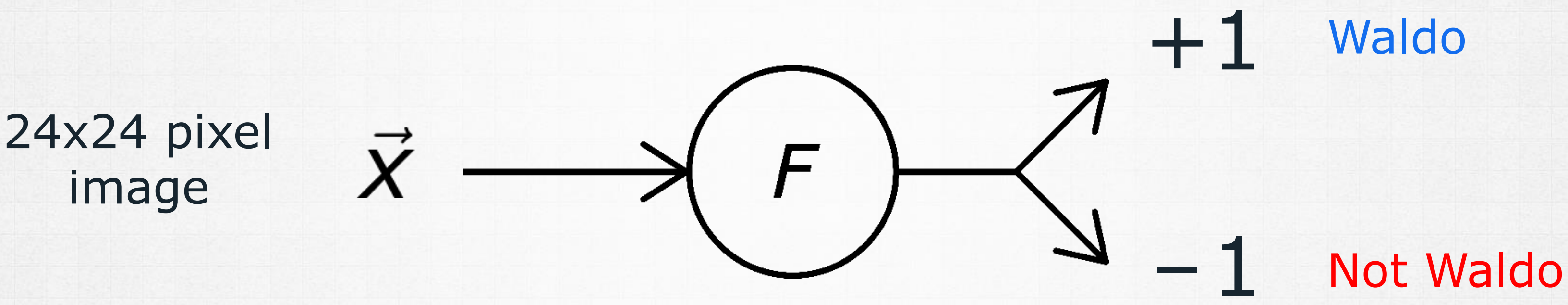
"Find Waldo as fast as possible"

Input:
reference image(s)
of object

Output:
position of object(s)
in any image

Binary Classification Problem

Simplified version:



e.g.:

Perceptron, Adaline, Neural Network, …

Main Concepts

- Features
- Boosting
- Cascading

Reduce Dimensions
(pixels → features)

Increase Performance
(speed, generalization)

Features

- Large input space (e.g. 24x24 pixel intensities)

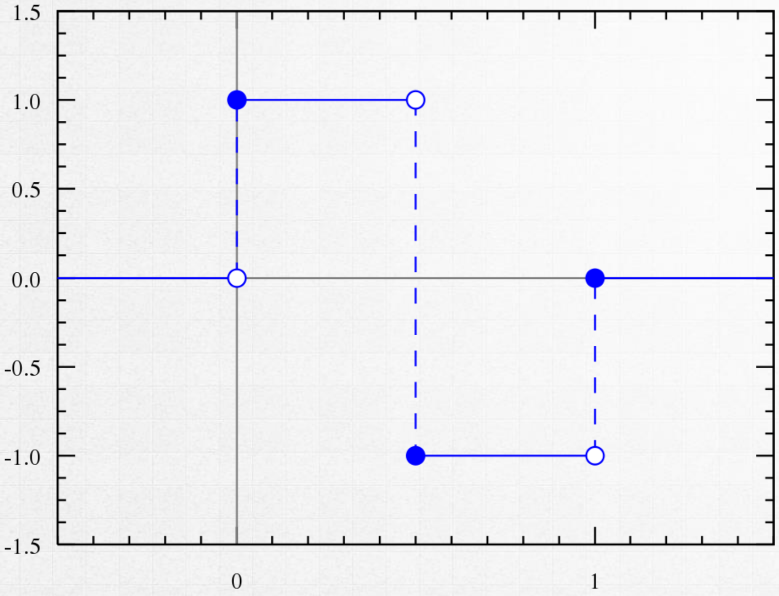
- Objects show certain patterns

→ Reduce amount of input/computation time by selecting features

→ Construct set of features (Haar, LBP, …)

Haar-like Features

Haar Wavelets: complete orthonormal system on $L^2(\mathbb{R})$



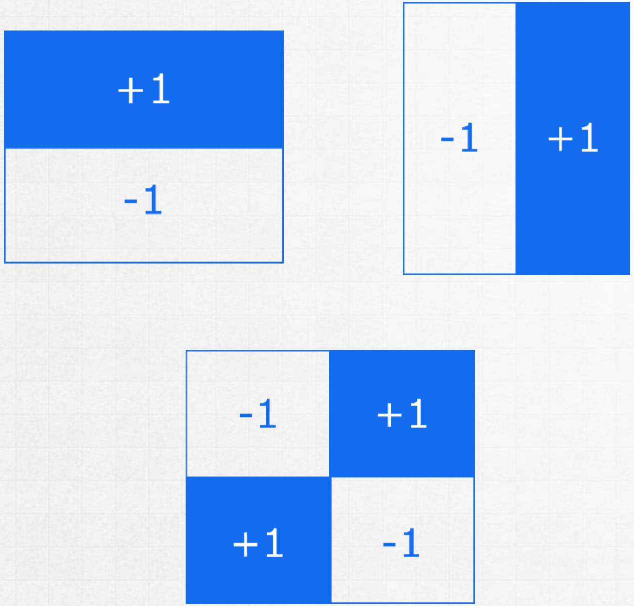
$$\psi(t) = \begin{cases} +1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & otherwise \end{cases}$$

$$\psi_{n,k}(t) = 2^{n/2}\,\psi\left(2^n t - k\right)$$

$$n, k \in \mathbb{Z} \qquad t \in \mathbb{R}$$

Haar-like Features

2D basis set:

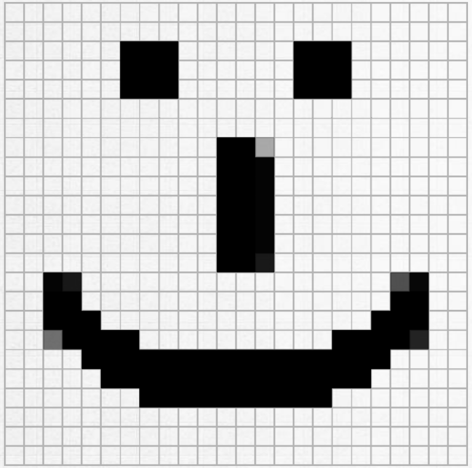


Additional features (Viola/Jones):

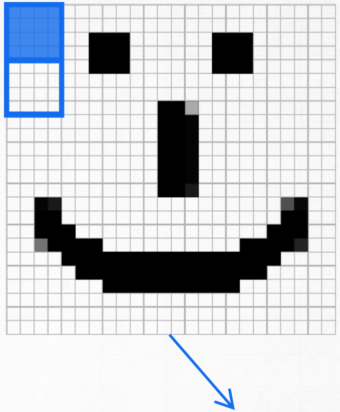
stretching + shifting:

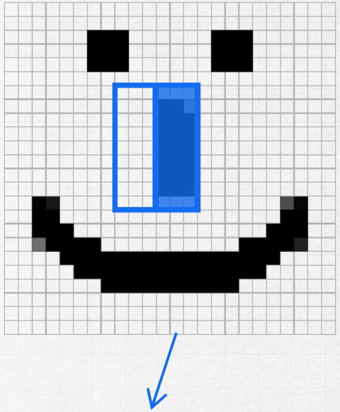
~ 180,000 features
(overcomplete basis)

24x24 Example

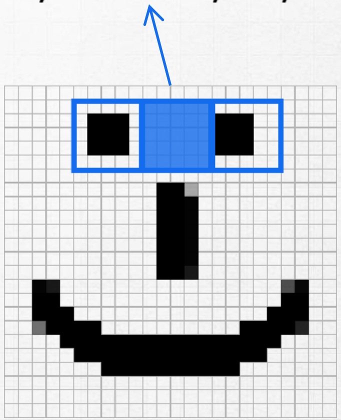


$$(255, 255, \ldots, 0, 0, \ldots) \in \{0, \ldots, 255\}^{576}$$





$$(\ldots, 0, \ldots, -1785, \ldots, -5000, \ldots)$$



Aside: Integral Image

→ Define Row Sum s and Integral Image ii:

$$s(x,y) = \sum_{x' \leq x} i(x',y) \qquad ii(x,y) = \sum_{y' \leq y} s(x,y') \qquad i(x,y)$$

intensity at (x,y)

Compute ii in single pass over image:

$$s(x,y) = s(x-1,y) + i(x,y)$$

$$ii(x,y) = ii(x,y-1) + s(x,y)$$

Feature Selection – Boosting

- Loads of features (more than pixels)
- Idea: construct weak classifiers, each using only one feature
- Select best weak classifiers (i.e. best features) and combine them into strong classifier

→ Adaptive Boosting Algorithm (AdaBoost)

Adaptive Boosting – Notation

- Training data: $(x_m, y_m) \in \mathbb{IM} \times \{-1, +1\}$ $\quad m \in \{1, \ldots, M\}$
- Set of weak classifiers (one for each feature): $f_i : \mathbb{IM} \to \{-1, +1\}$
- Strong classifier: $F_T(x) = sign\left(\sum_{t=1}^{T} \alpha_t \, \tilde{f}_t(x)\right)$

$$\alpha_t = \ln\left(\frac{1 - \tilde{\epsilon}_t}{\tilde{\epsilon}_t}\right)$$

weight of t-th classifier

$$\epsilon_i = \sum_{m=1}^{M} w_m \left|f_i(x_m) - y_m\right|$$

classification error of i-th classifier (weighted with w_m)

Initialize image weights uniformly $w_m = \frac{1}{M} \quad \forall m$

For every *t* in {1,…,*T*}:

1. "Train" every weak classifier *i* and compute error ϵ_i

2. Select classifier $\tilde{f}_t$ with best performance $\tilde{\epsilon}_t = \max_i \left| \epsilon_i - \frac{1}{2} \right|$

3. Compute classifier weight $\alpha_t = \ln \left(\frac{1 - \tilde{\epsilon}_t}{\tilde{\epsilon}_t} \right)$

4. Update image weights $F_t(x) \longleftarrow F_{t-1}(x) + \alpha_t \tilde{f}_t(x)$

5. Update strong classifier $w_m \longleftarrow \frac{1}{Z_t} w_m \exp \left[-\alpha_t \, y_m \, \tilde{f}_t(x_m) \right]$

AdaBoost Performance

Viola and Jones (2001): T = 200

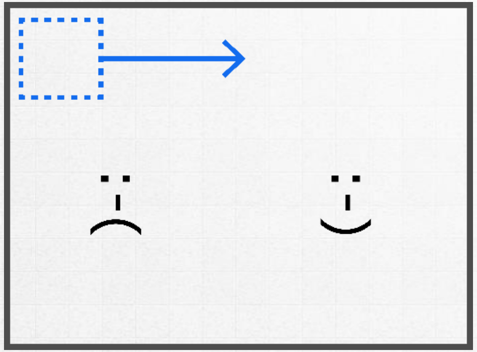
- \+ 95% detection rate
- \+ 1 in 14,084 false positive
- \+ Intuitive features
- – 0.7 seconds for scan of 384x288 pixel image
- – Computation time roughly proportional to T

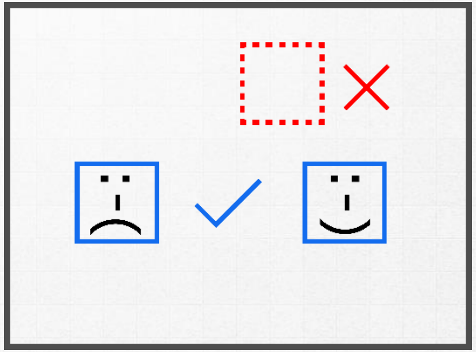
What about larger images? (Waldo)

- AdaBoost detects faces quickly in 24x24 image

- Usual image sizes *much* larger (FullHD: 1920x1080, …)

→ Sweep image with smaller windows and check each for face



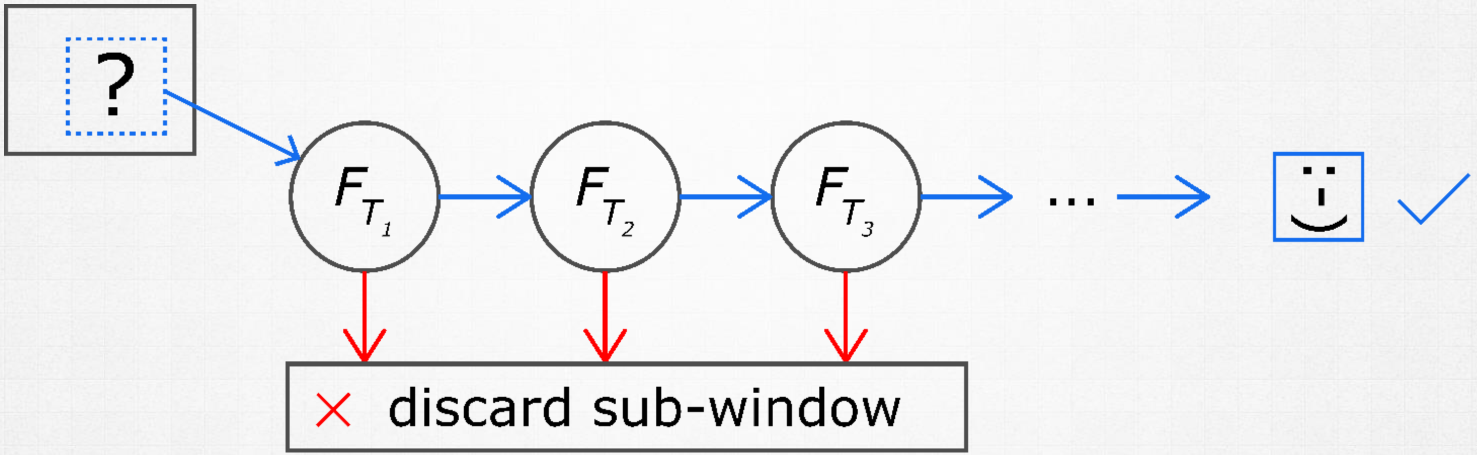


Cascading

- *Many* sub-windows to check
- Most parts of image don‘t contain faces
- Speed up process by quickly discarding “empty” sub-windows
- Modify choice of weak classifiers towards low false negative rates

→ (Attentional) Cascade Algorithm

Cascading



Cascade AdaBoost – “ad hoc”

Fix *general architecture* (no. of stages, desired detection rate, …)

→ Train boosted classifier for each stage

- “Performance” = low false negative rate
- Reduce target false pos. rates stage by stage (better accuracy)
- Test classifiers against separate validation data set
- …

AdaBoost+Cascade Performance

Viola and Jones (2001)

- 38 stages (1, 10, 25, 25, 50, … features – total 6061)
- On average, 10 features evaluated per sub-window
- 0.067 seconds for scan of 384x288 pixel image

Summary

1. Pixel intensities → generalized features (pattern informaion)

2. Best single-feature classifiers → Boosted classifier

3. Cascade of boosted classifiers → Quickly scan large images


29


29

*References

Images:

- https://www.techspot.com/news/75939-ai-powered-facial-recognition-robot-zaps-fun-where.html

- https://www.theridgefieldpress.com/wp-content/uploads/sites/28/2017/06/waldomain-1024x691.png

- https://en.wikipedia.org/wiki/Haar_wavelet#/media/File:Haar_wavelet.svg

Paper:
https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf