Face detection: Viola & Jones

*Multiple view-based classifiers* based on simple features that best discriminate faces vs. non-faces

Most discriminating features *learned* from thousands of samples of face and non-face image windows

*Attentional mechanism:* cascade of increasingly discriminating classifiers improves performance
Viola & Jones use simple features

Use simple rectangle features:

\[ \sum I(x, y) \text{ in gray area} - \sum I(x, y) \text{ in white area} \]

within 24 x 24 image sub-windows

- initially consider 160,000 potential features per sub-window!
- features computed very efficiently

**Which features best distinguish face vs. non-face?**

Learn most discriminating features from thousands of samples of face and non-face image windows

---

Learning the best features

Weak classifier using one feature:

\[ h(x, f, p, \theta) = \begin{cases} 1 & \text{if } pf(x) < p\theta \\ 0 & \text{otherwise} \end{cases} \]

\( x = \text{image window} \)
\( f = \text{feature} \)
\( p = +1 \text{ or } -1 \)
\( \theta = \text{threshold} \)

\( (x_1, w_1, 1) \quad (x_n, w_n, 0) \)

\( n \) training samples, equal weights, known classes

\[ C(x) = \begin{cases} 1 & \sum_{i=1}^{T} \alpha_i h_i(x) \geq \tau \\ 0 & \text{otherwise} \end{cases} \]

Final classifier

\(~200 \text{ features yields good results for "monolithic" classifier}~\)
“Attentional cascade” of increasingly discriminating classifiers

Early classifiers use a few highly discriminating features, low threshold

• 1st classifier uses two features, removes 50% non-face windows

• later classifiers distinguish harder examples

- Increases efficiency
- Allows use of many more features

→ Cascade of 38 classifiers, using ~6000 features

Training with normalized faces

5000 faces
many more non-face patches
faces are normalized for scale & rotation
small variation in pose
Viola & Jones results

With additional diagonal features, classifiers were created to handle image rotations and profile views.

Faces everywhere...