MAC 0460 / 5832
Introduction to Machine Learning

21 – Deep learning for Computer Vision

IME/USP (30/06/2021)

Computer Vision deals with image/video data

Often they are referred to as non-structured data
(meaning "non-tabular")

Any 1D signal (speech, music, ECG, etc) as well as "text" are also non-structured

Structured data

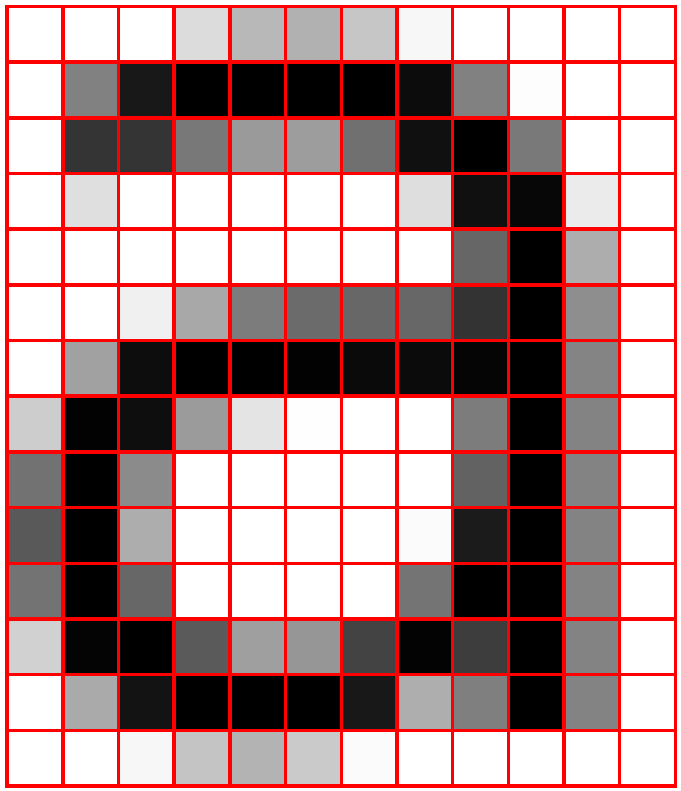
- Tabulated data
- Columns of the table are the "variables" or "features"
- Individual features may be in distinct domains
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

Non-structured data

- raw data, very low-level features
- semantic gap
- spatial/temporal relationships between low-level features
- requires feature extraction (which is strongly dependend on the type of data)

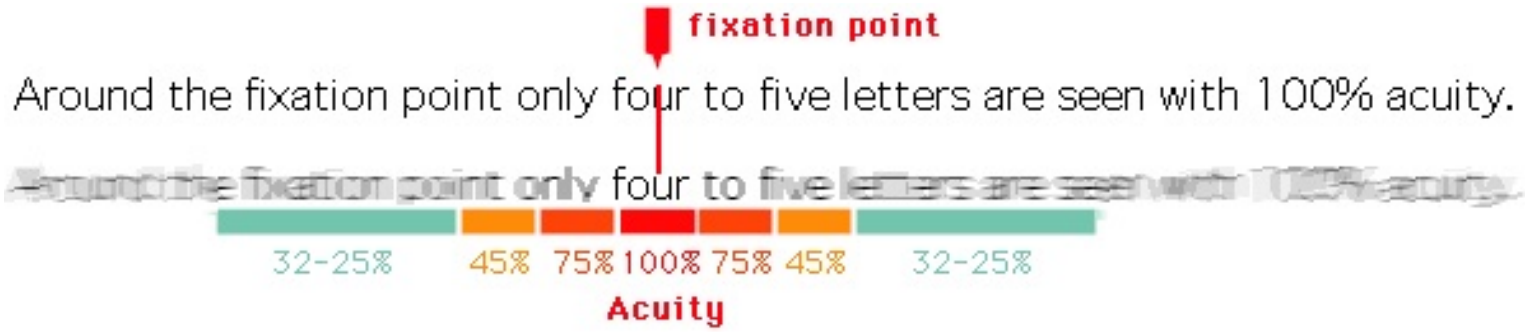
Images – spatial relationship is important

a

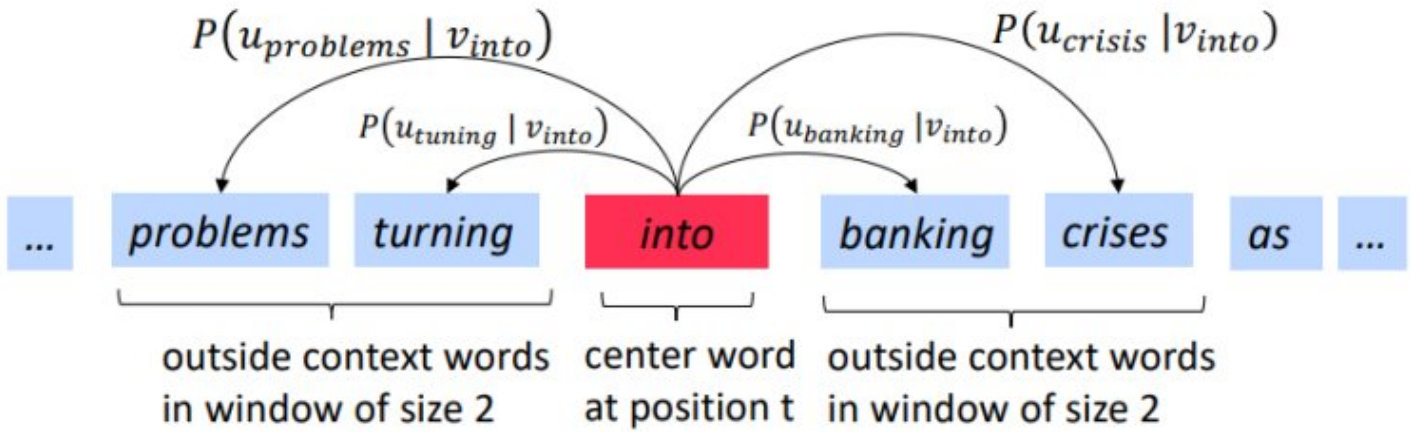


1.0	1.0	1.0	0.9	0.6	0.6	0.6	1.0	1.0	1.0	1.0	1.0
1.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0	1.0	1.0
1.0	0.2	0.2	0.5	0.6	0.6	0.5	0.0	0.0	0.5	1.0	1.0
1.0	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.0	0.0	0.9	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
1.0	1.0	1.0	0.5	0.5	0.5	0.5	0.5	0.4	0.0	0.5	1.0
1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0
0.9	0.0	0.0	0.6	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
0.5	0.0	0.6	1.0	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
0.5	0.0	0.7	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.5	1.0
0.6	0.0	0.6	1.0	1.0	1.0	1.0	0.5	0.0	0.0	0.5	1.0
0.9	0.1	0.0	0.6	0.7	0.7	0.5	0.0	0.5	0.0	0.5	1.0
1.0	0.7	0.1	0.0	0.0	0.0	0.1	0.9	0.8	0.0	0.5	1.0
1.0	1.0	1.0	0.8	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0

NLP – temporal relationship is important



Wikipedia



http://web.stanford.edu/class/cs224n/

Much of the DL advances has been driven by

Computer Vision (photos, videos)

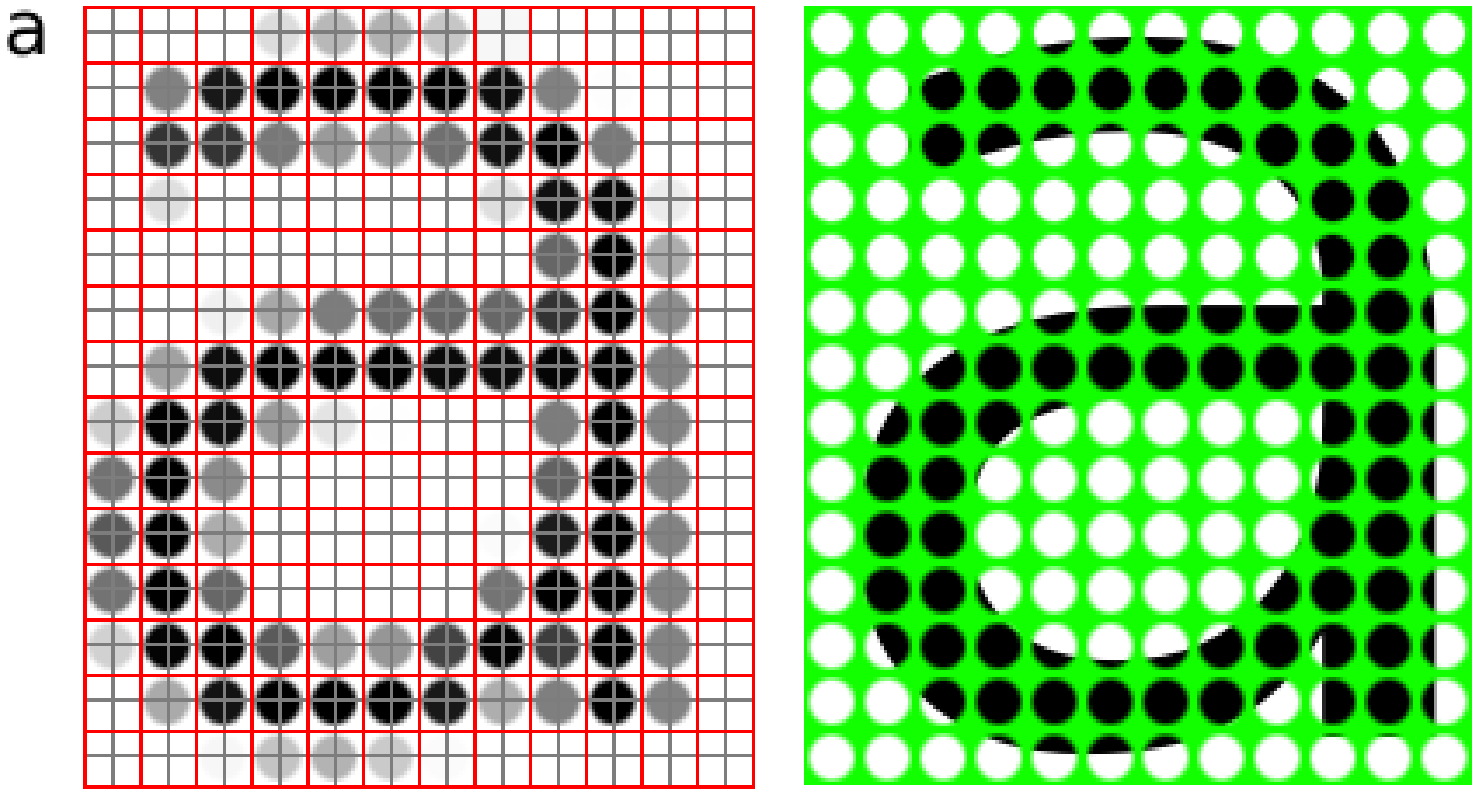
Natural Language Processing (documents, speech)

Traditional ML: given features, learns a prediction model

Deep Learning: given raw data, learns a prediction model including feature extraction

Understanding image data and image processing

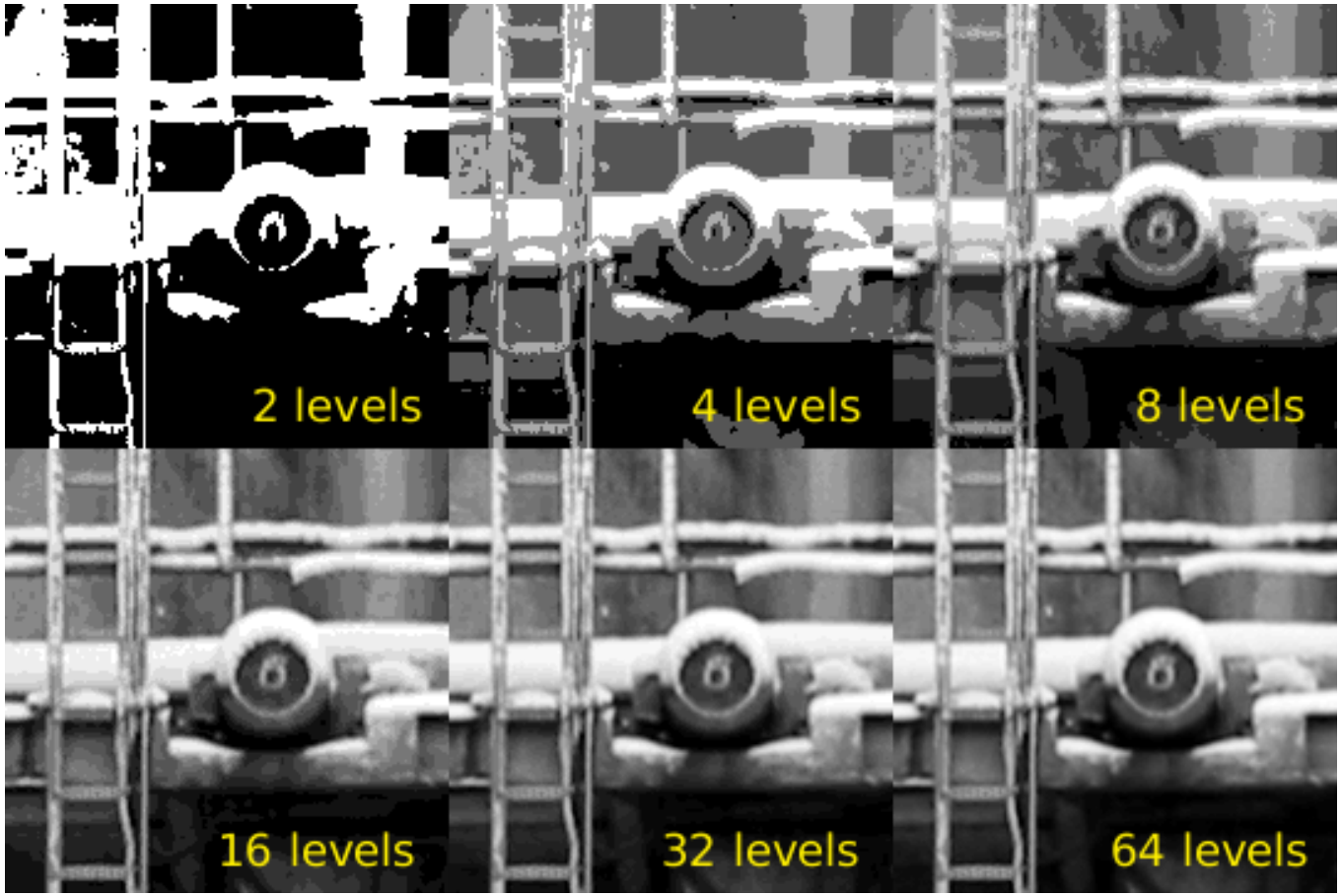
Digital images – spatial sampling



(http://pippin.gimp.org/image_processing)

Resolution: pixels per inch (ppi)

Digital images – quantization



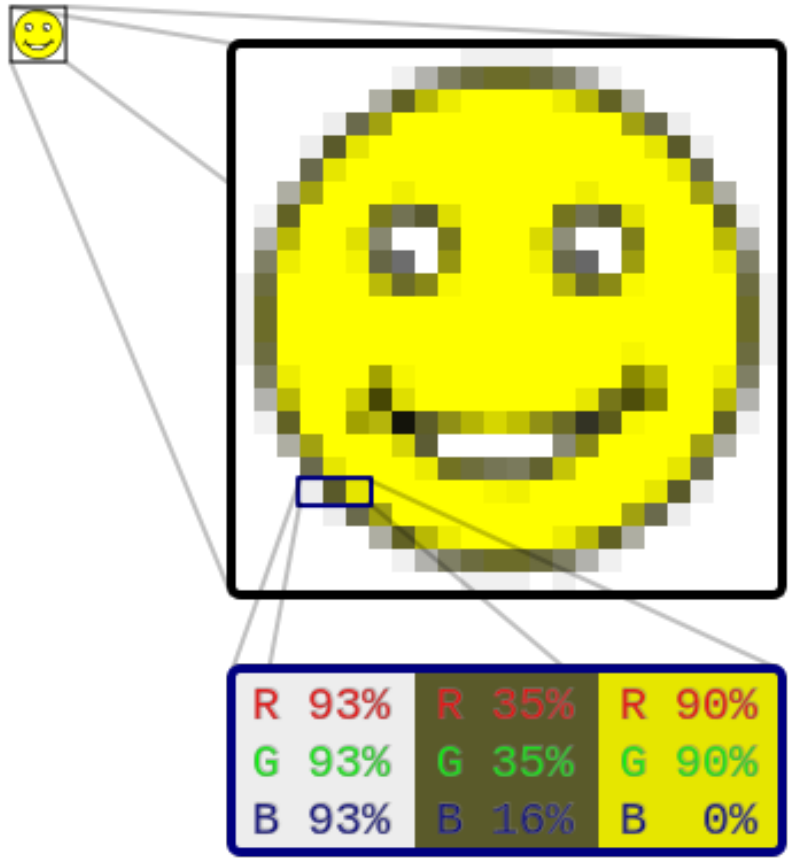
(http://pippin.gimp.org/image_processing)

Number of *bits* per pixel $\longrightarrow$ gray levels

Standard: 8 *bits* $\longrightarrow$ 256 gray levels

Black=0 e white=255

Digital images – Color images

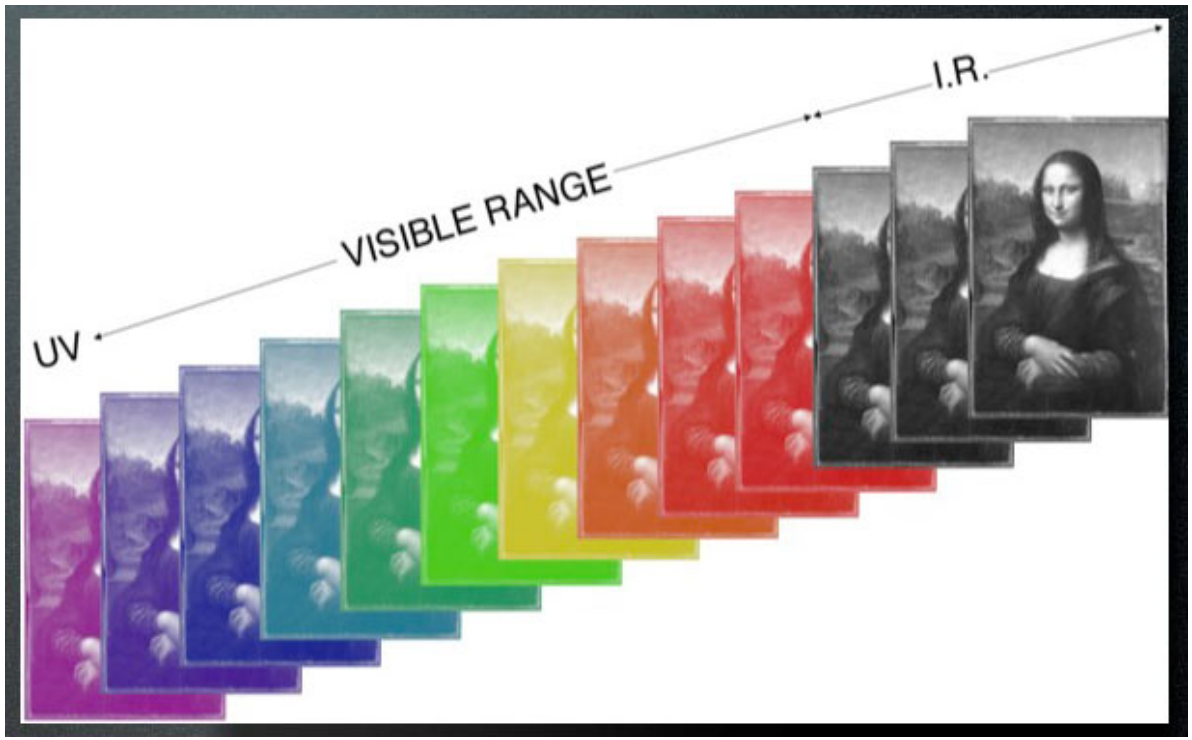


(en.wikipedia.org/wiki/Raster_graphics)

Digital images – Color images

(https://www.photoshopessentials.com/essentials/rgb/)

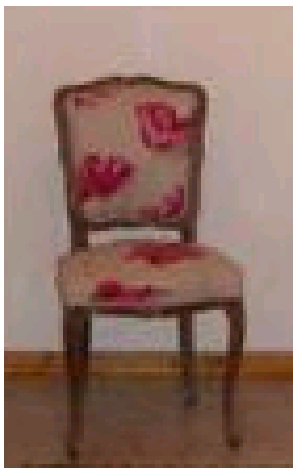
Digital images – Multispectral images



(http://www.lumiere-technology.com/Pages/Services/services3.htm)

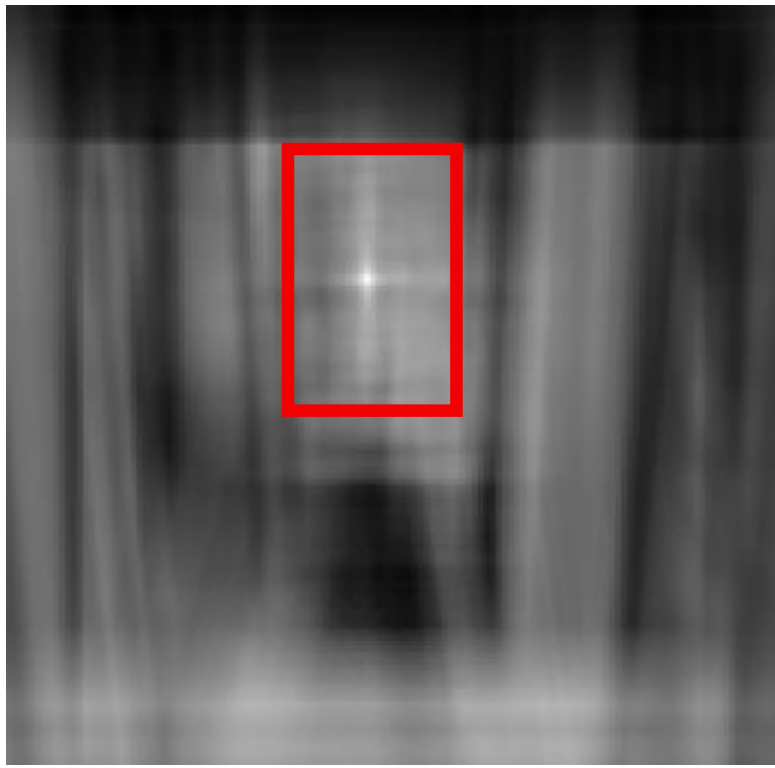
Object recognition
Is it really so hard?

This is a chair

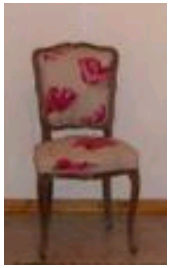


Find the chair in this image

Output of normalized correlation

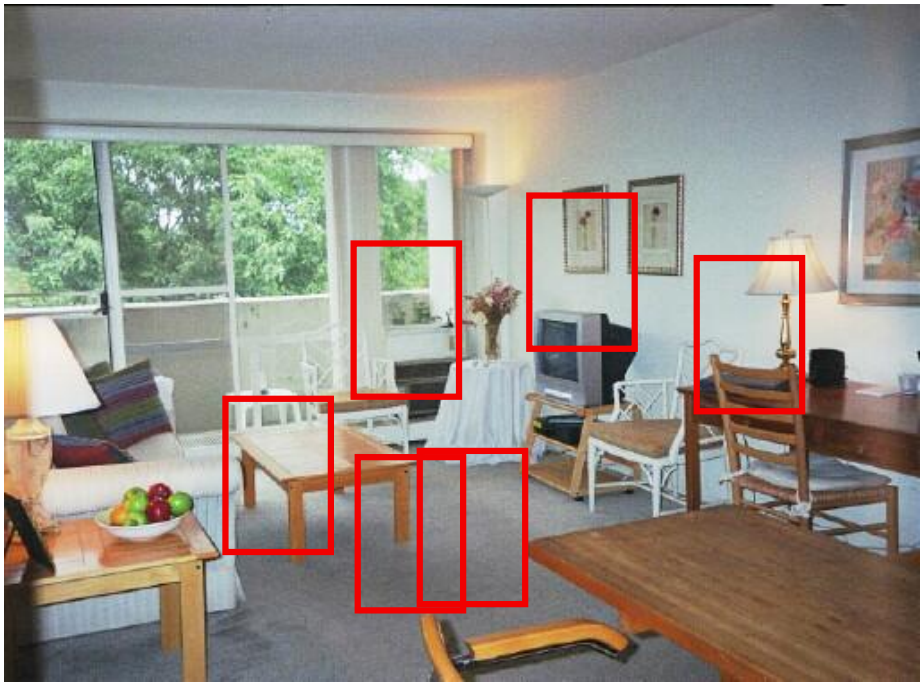


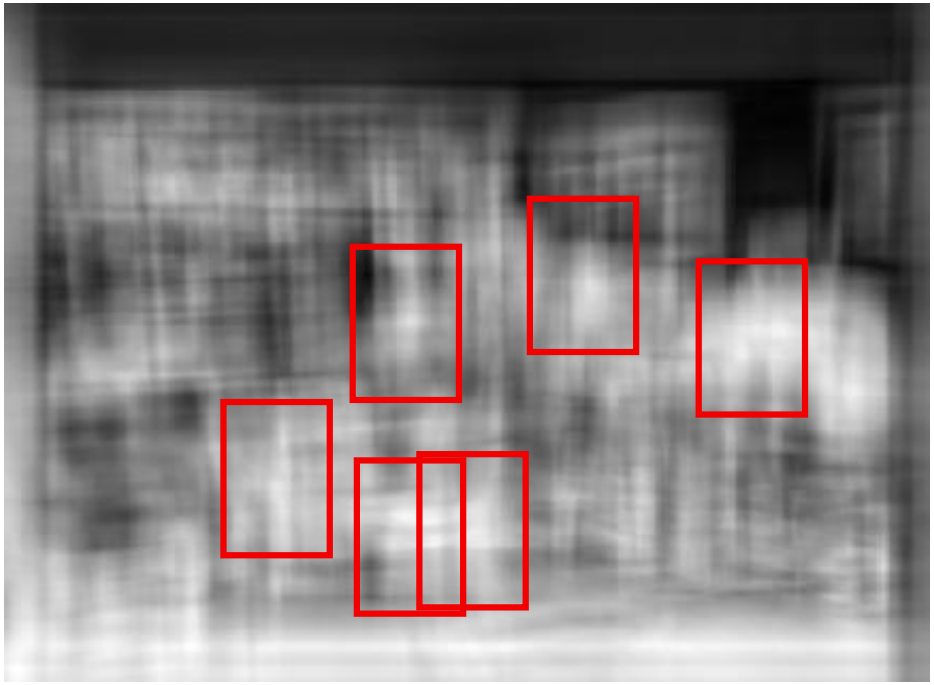
Slide: A. Torralba



Object recognition
Is it really so hard?

Find the chair in this image





Pretty much garbage

Simple template matching is not going to make it

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.

Slide: A. Torralba

Why image processing/analysis is hard ?

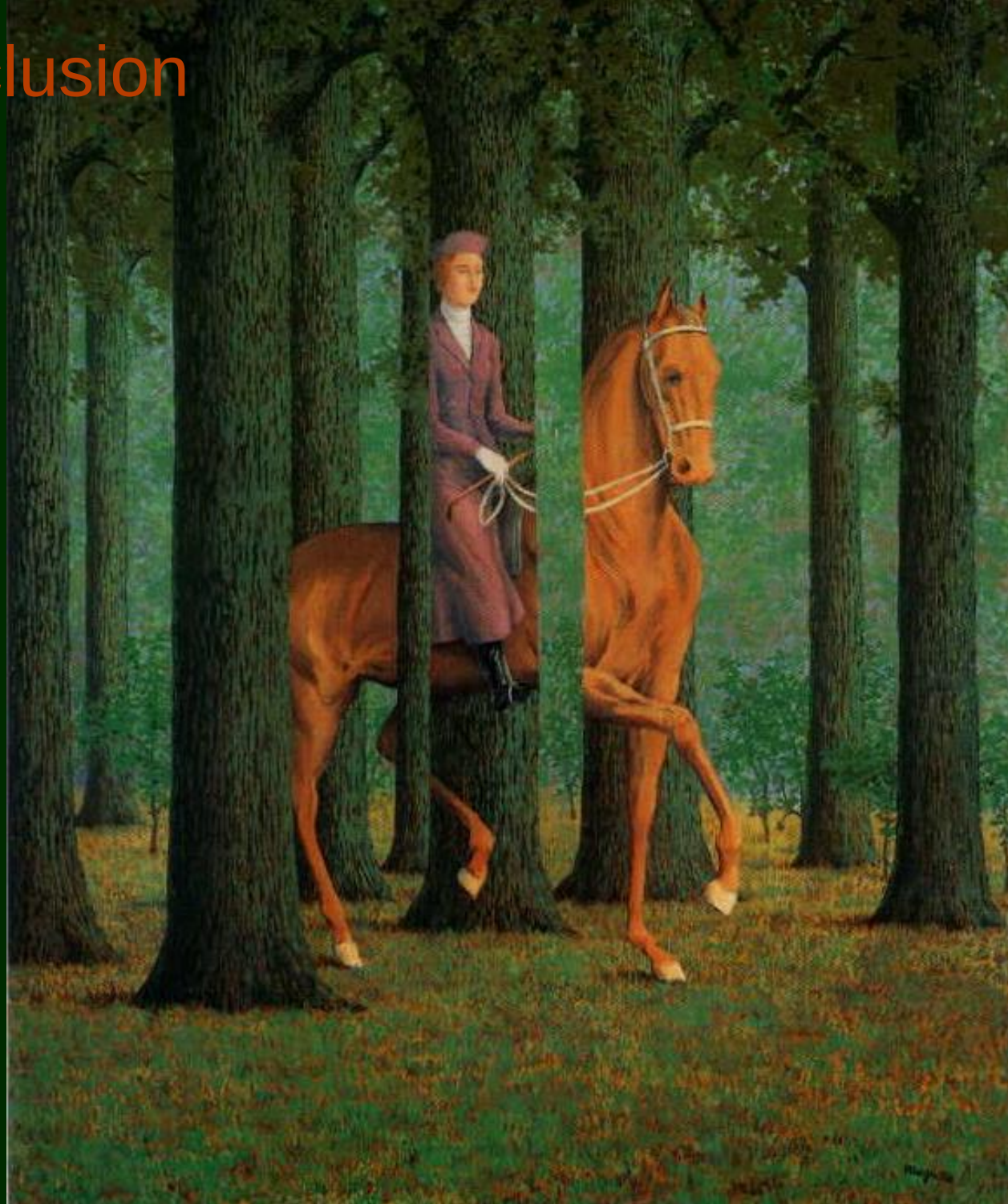
Challenges 1: view point variation

Michelangelo 1475-1564

Challenges 2: illumination

slide credit: S. Ullman

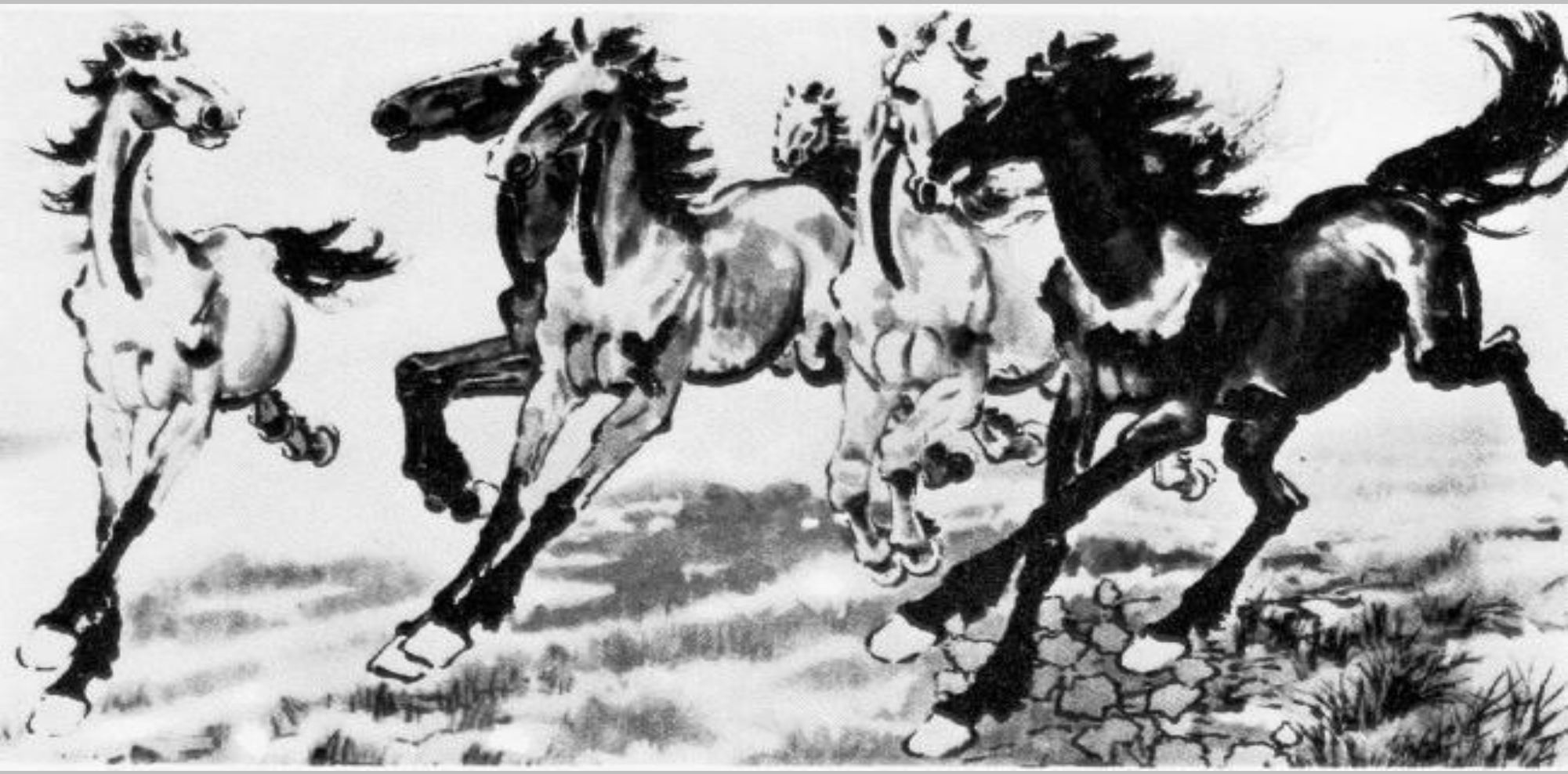
Challenges 3: occlusion



Magritte, 1957

Challenges 4: scale

Challenges 5: deformation



Xu, Beihong 1943

Challenges 6: background clutter

Klimt, 1913

Within-class variations

Image features

Simple features such as an intensity histogram **are not useful**

Intensity histograms may help to separate darker images from lighter ones, but it has **no awareness about geometry/topology** of the image content

Many **local feature extraction algorithms** has been proposed

Many sophisticated methods **combining heuristics and ML** has been developed

Let us look at some solutions of the pre-DL era

- serves as examples of ML application

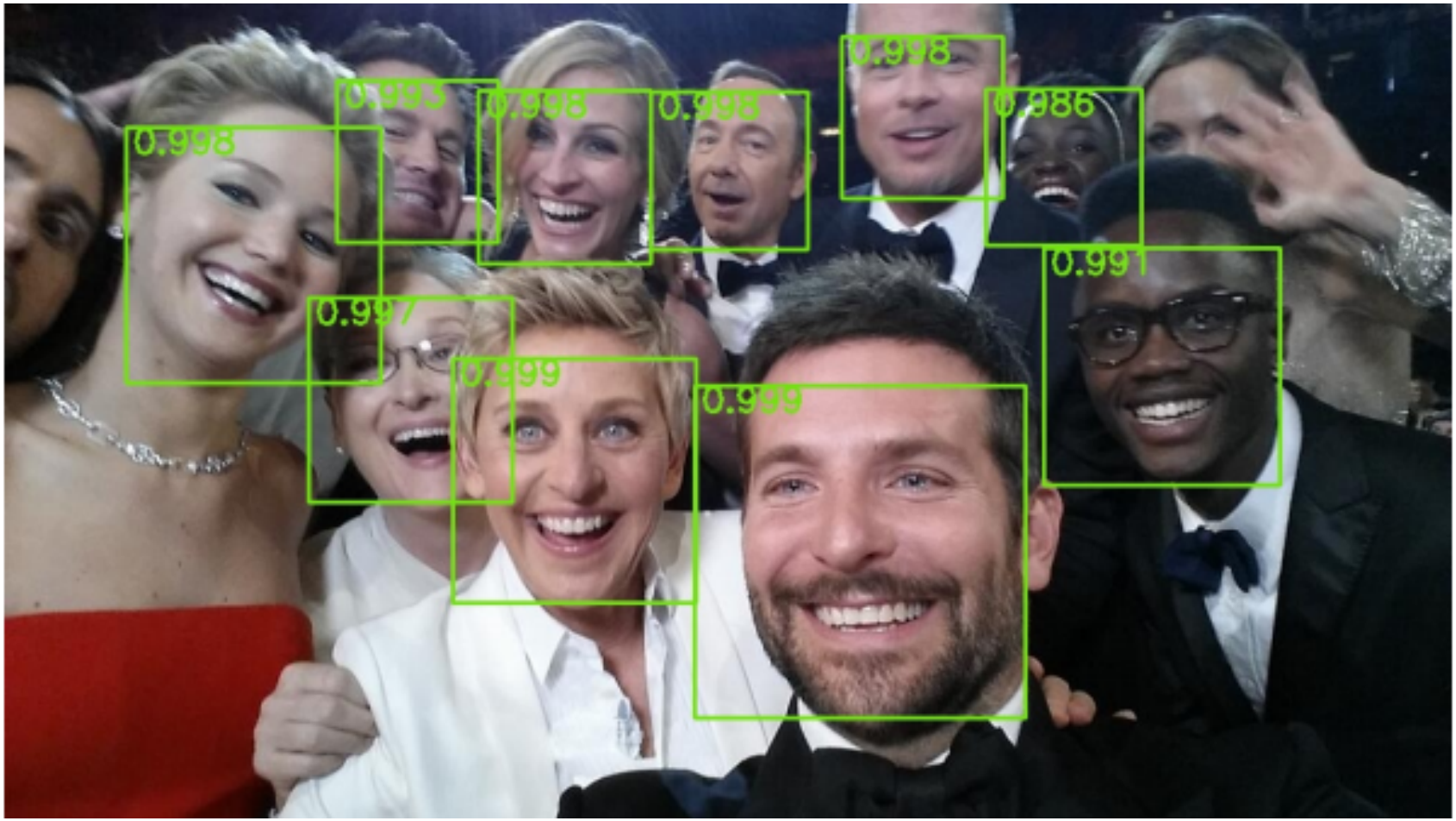
- power of DL methods will became more clear

Object detection

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

Robust Real-time Object Detection, IJCV 2004

The problem of object detection in images



Bounding boxes are used to indicate detections

(this is the outcome of a Deep Learning based solution ...)

Desired characteristics for object detectors

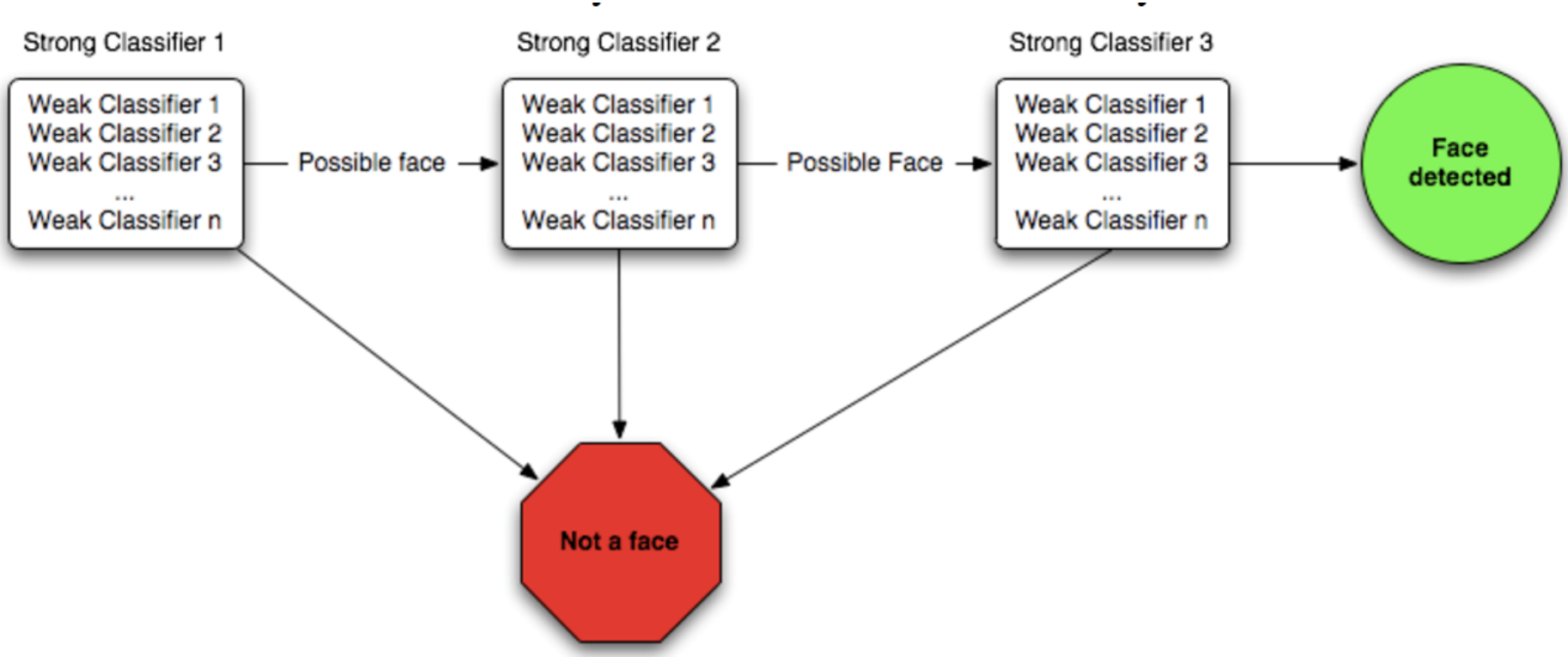
- fast – real-time
- high detection rate
- no false positives
- robust to variations
  - localization inside the image
  - scale
  - pose
  - ilumination
  - partial occlusion

Viola & Jones – Four key ideas

1. **Haar-like** features
2. **Integral image** for fast computation of features
3. **Classifier based on Adaboost**, using Haar-like features, to classify image patches
4. **Cascade of classifiers** for rejecting false *patches* as soon as possible

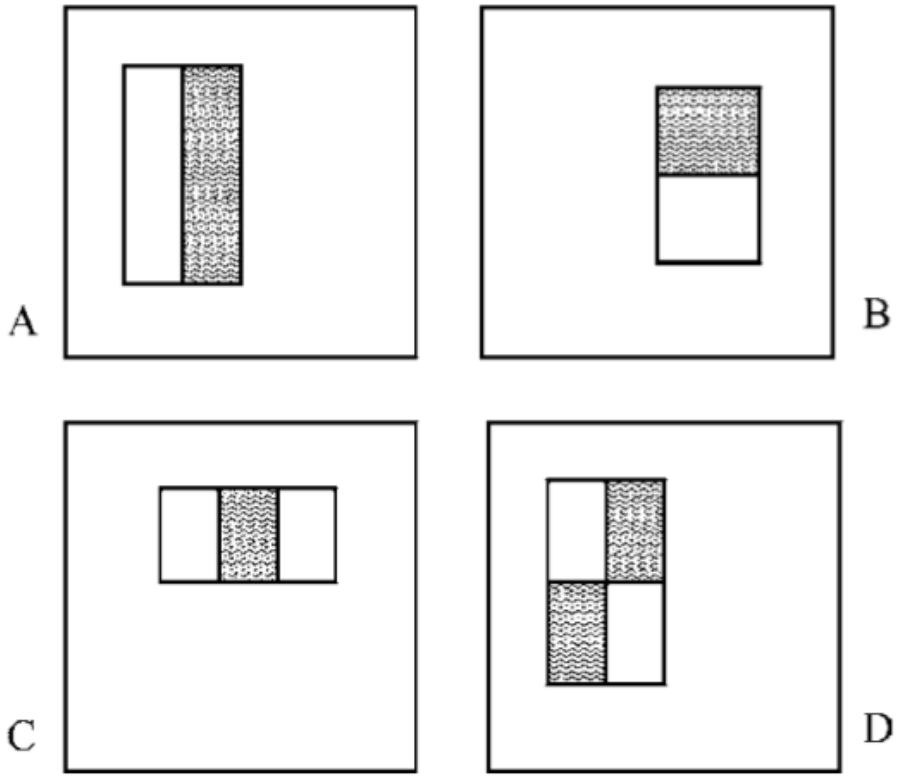
Viola-Jones – based on Adaboost and cascade of classifiers

Image patch classifier structure



(https://www.quora.com/How-can-I-understand-Haar-like-feature-for-face-detection)

Haar-like features



patch size: 24 $\times$ 24

(sum of intensities under white region) - (sum of intensities under dark region)

[or vice-versa]

Integral image

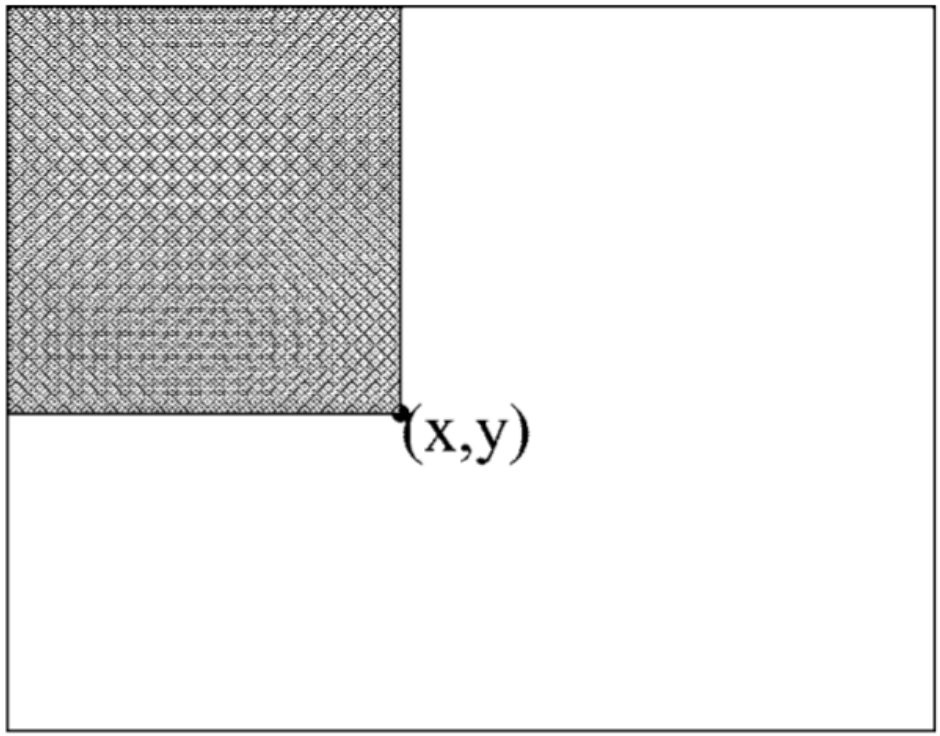


Figure 2. The value of the integral image at point (x, y) is the sum of all the pixels above and to the left.

Feature computation

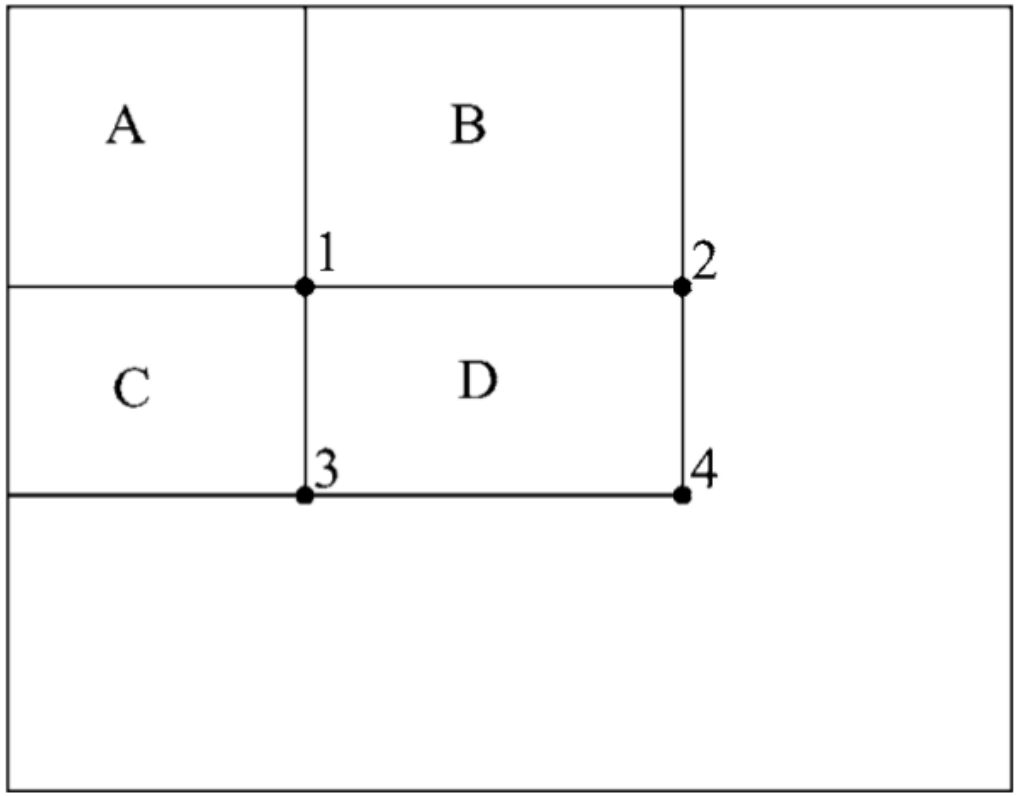
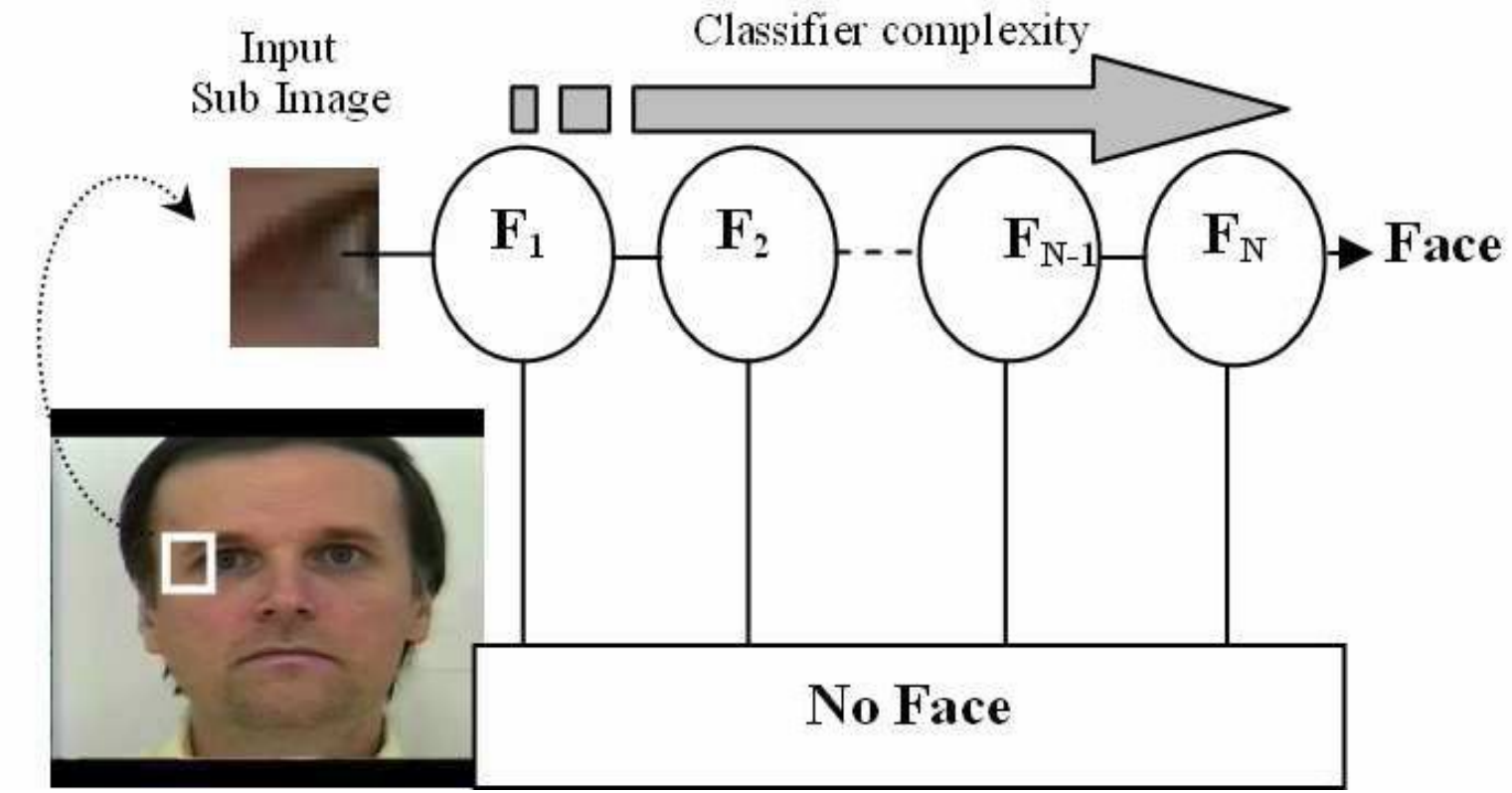


Figure 3. The sum of the pixels within rectangle D can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is $A + B$, at location 3 is $A + C$, and at location 4 is $A + B + C + D$. The sum within D can be computed as $4 + 1 - (2 + 3)$.

Viola-Jones – based on Adaboost and cascade of classifiers



(http://computervisionwithvaibhav.blogspot.com/2015/08/viola-jones-in-nut-shell.html)

Each F_i is an Adaboost-like classifier

Adaboost

Weighted combination of *weak classifiers*

A weak classifier $(h(x, f, p, \theta))$ thus consists of a feature (f), a threshold (θ) and a polarity (p) indicating the direction of the inequality:

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

Here x is a 24×24 pixel sub-window of an image.

Adaboost

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.

- For $t = 1, \ldots, T$:
  1. Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$
  2. Select the best weak classifier with respect to the weighted error

     $$\epsilon_t = \min_{f,p,\theta} \sum_i w_i \, |h(x_i, f, p, \theta) - y_i| \, .$$

     See Section 3.1 for a discussion of an efficient implementation.
  3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where f_t, p_t, and θ_t are the minimizers of ϵ_t.
  4. Update the weights:

     $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

     where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

- The final strong classifier is:

$$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

Training details

4916 faces 24×24 + vertical flipping $\longrightarrow$ 9832
10000 non-face images

At each level of the cascade, 10000 new false positives are collected from images that do not contain faces

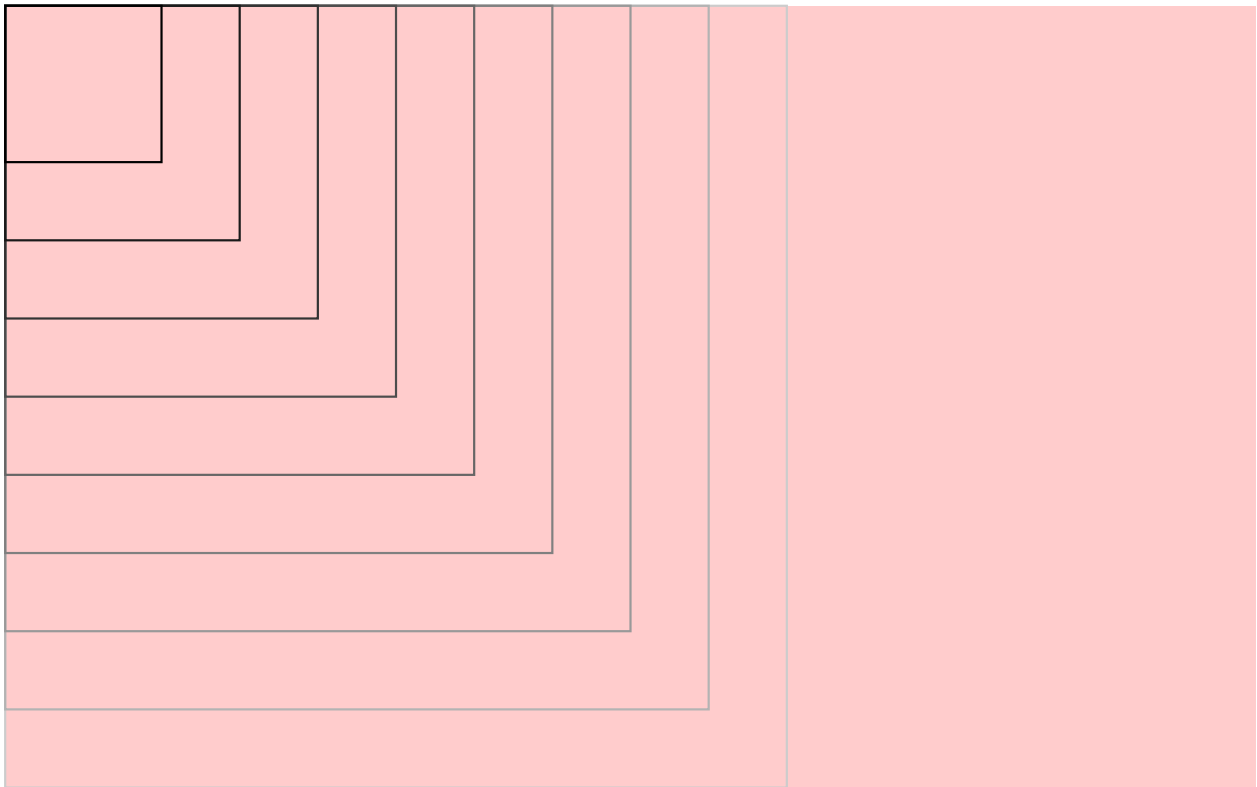
Each classifier in the cascade is trained using the Adaboost algorithm, with threshold adjusted to minimize false-negatives

The closer to the end of the cascade, the harder is the classification task

Final: 38 levels in the cascade, total of 6061 features

Exemples of face images usend in training

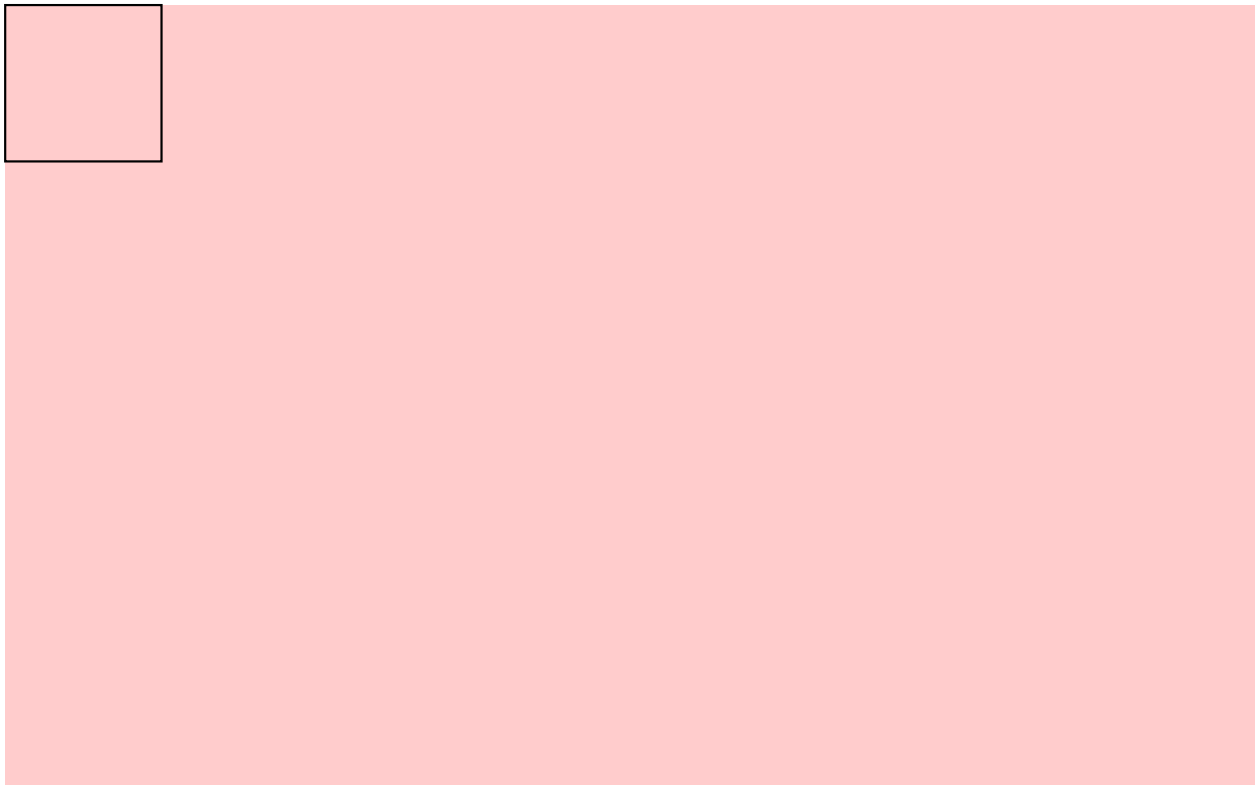
Scale



At each position in the image, the window is scaled by a constant factor

Multiple image patches

Localization



The window is shifted over the image by a step Δ

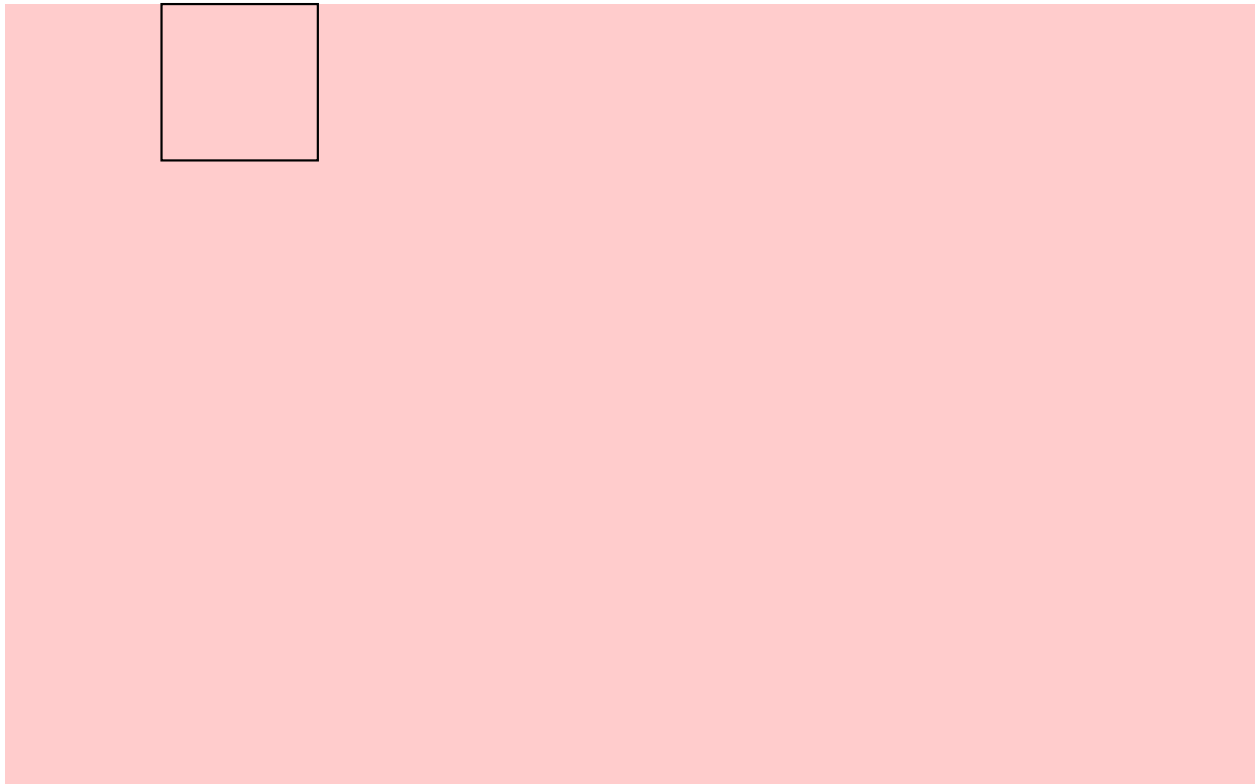
Position invariant

Localization

The window is shifted over the image by a step Δ

Position invariant

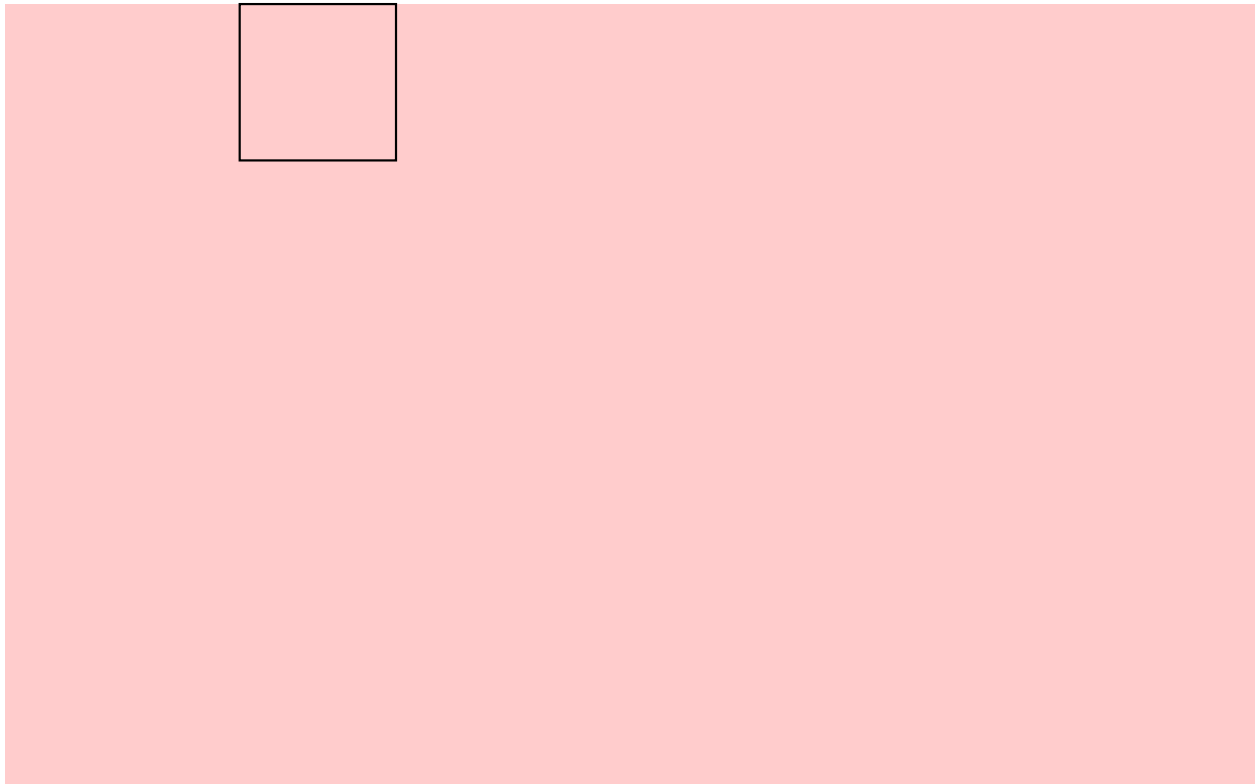
Localization



The window is shifted over the image by a step Δ

Position invariant

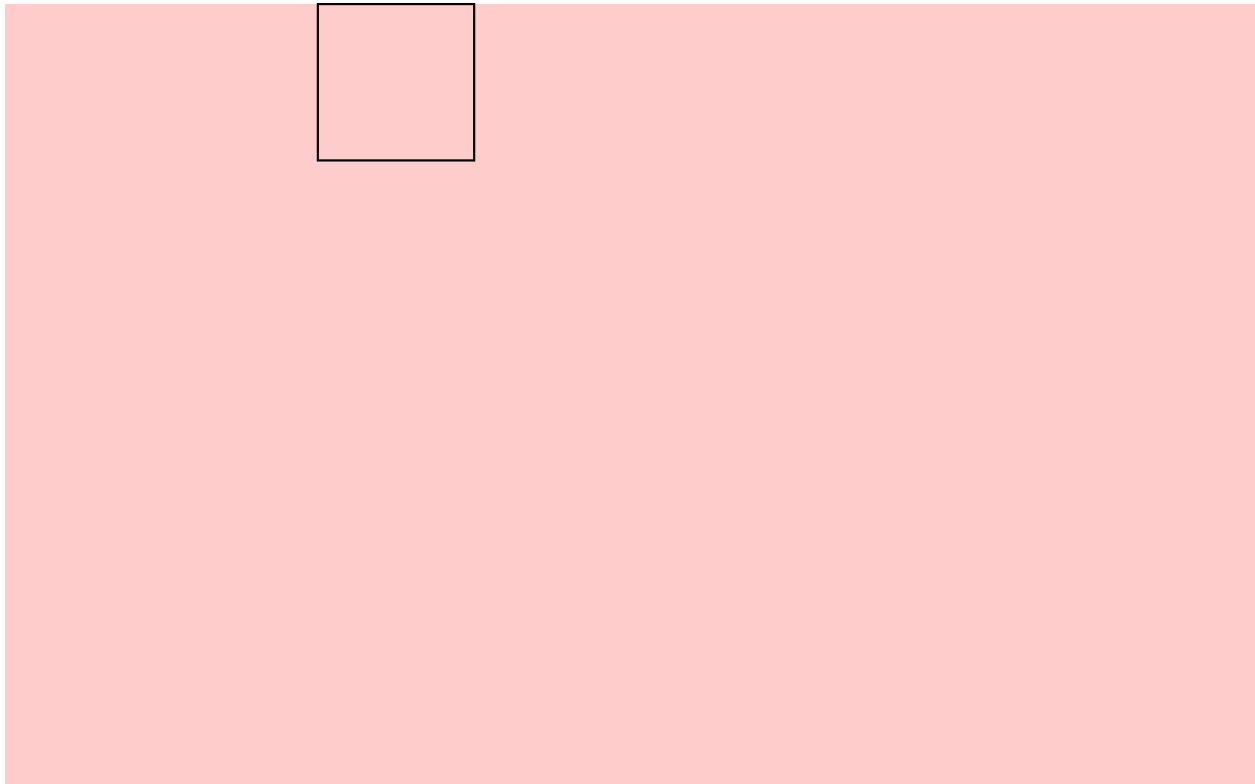
Localization



The window is shifted over the image by a step Δ

Position invariant

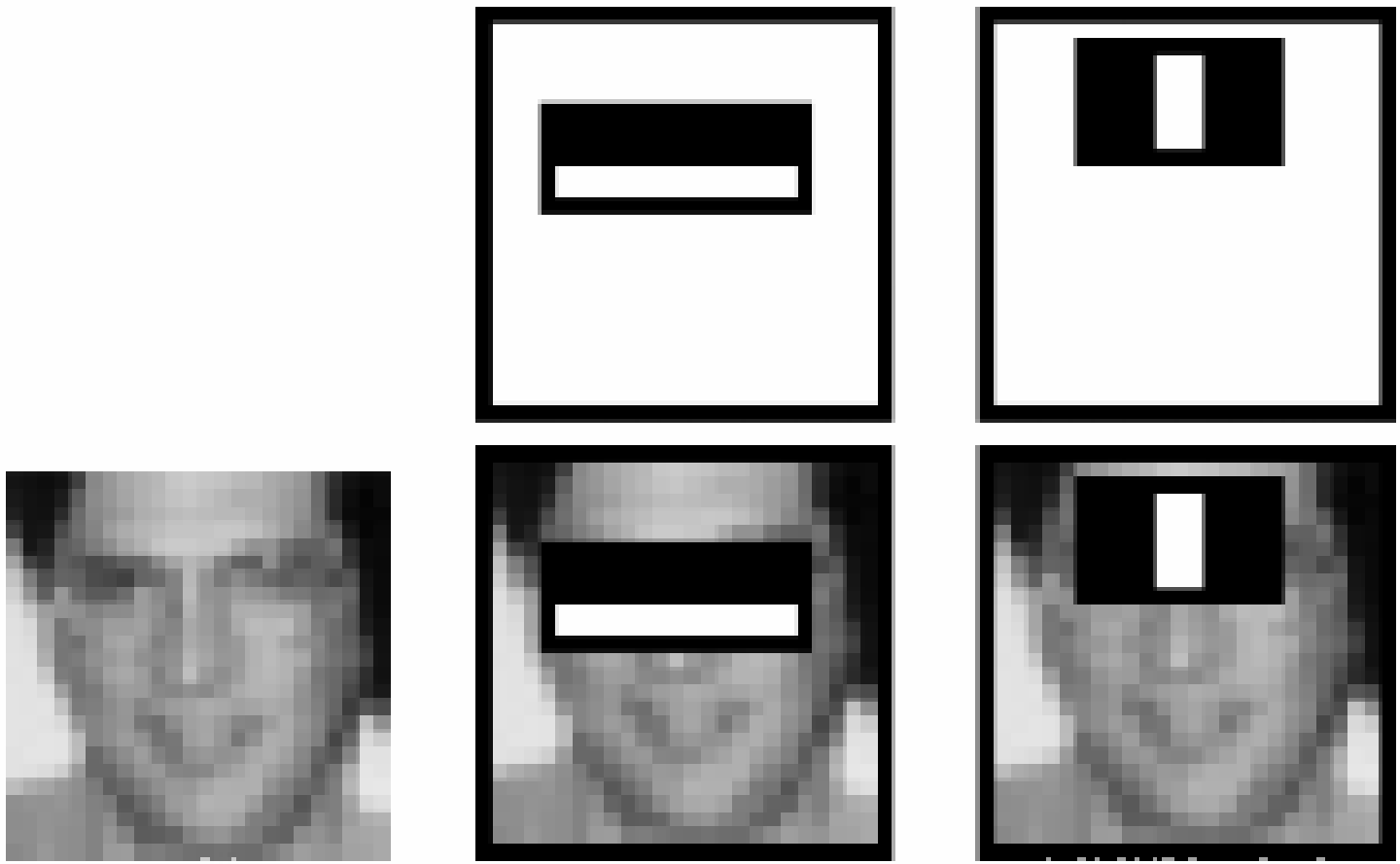
Localization



The window is shifted over the image by a step Δ

Position invariant

Result: The two main features

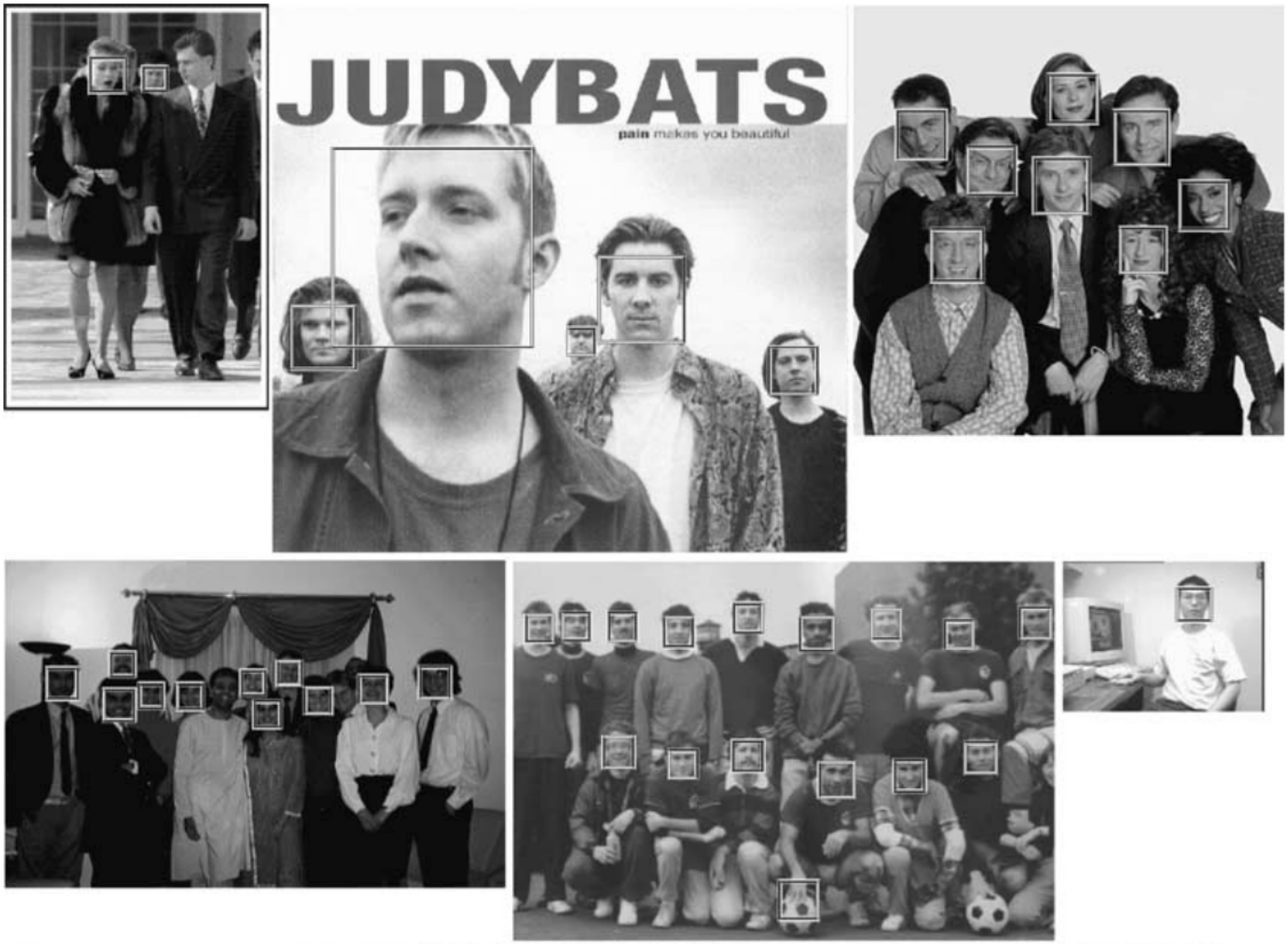


(https://docs.opencv.org/3.4.1/d7/d8b/tutorial_py_face_detection.html)

Results

Table 3. Detection rates for various numbers of false positives on the MIT + CMU test set containing 130 images and 507 faces.

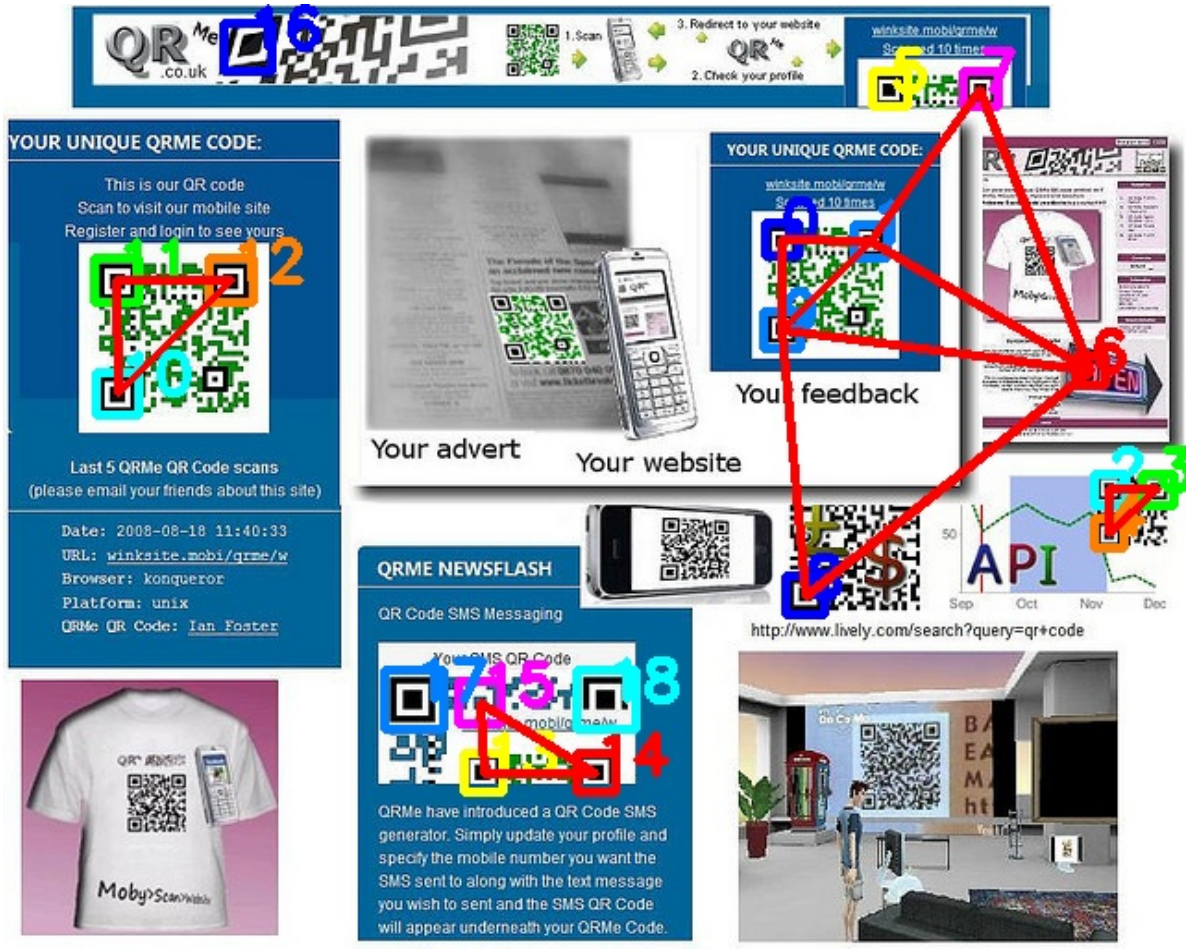
	False detections							
Detector	10	31	50	65	78	95	167	422
Viola-Jones	76.1%	88.4%	91.4%	92.0%	92.1%	92.9%	93.9%	94.1%
Viola-Jones (voting)	81.1%	89.7%	92.1%	93.1%	93.1%	93.2%	93.7%	–
Rowley-Baluja-Kanade	83.2%	86.0%	–	–	–	89.2%	90.1%	89.9%
Schneiderman-Kanade	–	–	–	94.4%	–	–	–	–
Roth-Yang-Ahuja	–	–	–	–	(94.8%)	–	–	–


JUDYBATS
pain makes you beautiful

Other applications – detection of QR codes

Dissertation by Luiz F. F. Belussi


staatstheater darmstadt
20:00
Platz 67
Dienstplätze
www.staatstheater-darmstadt.de
www.staatstheater-darmstadt.de/mobile
Bitte beachten Sie, dass bei beiden Angeboten die jeweiligen Online-Gebühren Ihres Anbieters anfallen. Einen kostenlosen


QR Me .co.uk
1. Scan
2. Check your profile
3. Redirect to your website
winksite.mobi/qrme/w
YOUR UNIQUE QRME CODE:
This is our QR code
Scan to visit our mobile site
Register and login to see yours
Last 5 QRMe QR Code scans
(please email your friends about this site)
Date: 2008-08-18 11:40:33
URL: winksite.mobi/qrme/w
Browser: konqueror
Platform: unix
QRMe QR Code: Ian Foster
Your advert
Your website
YOUR UNIQUE QRME CODE:
winksite.mobi/qrme/w
Your feedback
API
Sep
Oct
Nov
Dec
http://www.lively.com/search?query=qr+code
QRME NEWSFLASH
QR Code SMS Messaging
QRMe have introduced a QR Code SMS generator. Simply update your profile and specify the mobile number you want the SMS sent to along with the text message you wish to sent and the SMS QR Code will appear underneath your QRMe Code.
Moby Scan