Face detection

Slides adapted Grauman & Liebe’s tutorial
- http://www.vision.ee.ethz.ch/~bleibe/teaching/tutorial-aaai08/
Also see Paul Viola’s talk (video)

Limitations of Eigenfaces
Eigenfaces are cool.
But they’re not great for face detection.

Chief Limitations
- not very accurate
- not very fast

To make it work on the camera, we need ~30fps, and near-perfect accuracy.

Rectangle filters

Why rectangles?
Answer: very very fast to compute
- Trick: integral images (aka summed-area-tables)

Value at \((x,y)\) is sum of pixels above and to the left of \((x,y)\)

Integral image
Integral images

What's the sum of pixels in the blue rectangle?

input image

integral image

A
B
C
D

integral image

A
B
C
D

Integral images

integral image

A
 integral image

B

integral image

A
B
Integral images

What's the sum of pixels in the rectangle?

Computing sum within a rectangle

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:

$$\text{sum} = A - B - C + D$$

Only 3 additions are required for any size of rectangle!
Filter as a classifier

How to convert the filter into a classifier?

\[ f(x) \rightarrow h(x) = \begin{cases} +1 & \text{if } f(x) > \theta_i \\ -1 & \text{otherwise} \end{cases} \]

Resulting weak classifier:

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

Finding the best filters...

Considering all possible filter parameters: position, scale, and type:

180,000+ possible filters associated with each 24 x 24 window

Which of these filters(s) should we use to determine if a window has a face?

Boosting

Weak Classifier 1

Weights Increased
Boosting

Weights Increased

Final classifier is a combination of weak classifiers

Weak Classifier 2

Weak Classifier 3
Boosting: training

- Initially, weight each training example equally
- In each boosting round:
  - find the weak classifier with lowest weighted training error
  - raise weights of training examples misclassified by current weak classifier
- Final classifier is linear combination of all weak classifiers
  - weight of each learner is directly proportional to its accuracy
- Exact formulas for re-weighting and combining weak learners

Viola-Jones Face Detector: Results

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

Kristen Grauman

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: http://www.intel.com/technology/computing/opencv/
Viola-Jones Face Detector: Results

Detecting profile faces?

Can we use the same detector?

Example using Viola-Jones detector

Frontal faces detected and then tracked, character names inferred with alignment of script and subtitles.

Everingham, M., Sivic, J. and Zisserman, A.
‘Hello! My name is... Buffy’ - Automatic naming of characters in TV video, BMVC 2006. http://www.robots.ox.ac.uk/~vgg/research/face/index.html
Consumer application: iPhoto 2009

Things iPhoto thinks are faces

What other categories are amenable to window-based representation?

Pedestrian detection

- Detecting upright, walking humans also possible using sliding window’s appearance/texture; e.g.,

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
- Objects with less-regular textures not captured well with holistic appearance-based descriptions
- If considering windows in isolation, context is lost
Limitations (continued)

- In practice, often entails large, cropped training set (expensive)
- Requiring good match to a global appearance description can lead to sensitivity to partial occlusions

Image credit: Adam, Rivlin, & Shimshoni

Kristen Grauman