CMPS 242 Project
Applying AdaBoost for text detection in natural scenes

Henry Lin
hhlin@ucsc.edu

Fall 09
Viola-Jones’ cascading face detection algorithm is fast, but uses **Haar basis features in its experts**.

The Haar functions were chosen for their computational speed and do not effectively model objects with high spatial variation (i.e. text).

A large number of weak, Haar basis learners are needed to accurately detect text and the real-time property is lost.

By using a different, heterogeneous collection of experts, we can find text in natural scenes using far less experts than using Harr-based features alone.
Gradient edge response

- Gradients are useful for detecting edges, and these experts try to find useful patterns.
Means of derivatives’ absolute values

(Top) $X$ derivatives (Bottom) $Y$ derivatives
(a) one sample (b) 100 samples (c) 1000 samples
Observe that:

- The y derivative of text regions has twin peaks.
- Stronger variance in the x derivative.

Solution: Use features on the gradient! (Chen and Yuelle)
Chen and Yuelle (2004)

Similar to Viola-Jones’ algorithm. Uses cascading.

- Each cascade layer uses a strong learner to prune regions of the image.
- Uses AdaBoost to find 4 cascade layers:
  - First three based on mean, variance, and modulus of derivatives. Features chosen with biases against false negatives.
  - Last contains computationally-expensive edge linking. (Not in the paper)
- Also uses block features to give the experts a ‘push’ in the right direction:
  - Analyzes mean derivative within each block.
My implementation

- Implemented block features, variance, derivative features. Did not implement edge linking.
- Used SVMs to generate experts w/ respect to distribution \( w \) at the start of each AdaBoost iteration.
- Implemented cascading.
- Does well on classifying text regions from non-text regions. With sliding window, not so much.
Shen and Coughlan (2006)

- Uses edge detection (based on the gradient) to heuristically locate common features of the latin alphabet.
  - For example, an upper-left corner might indicate the start of a horizontal stroke.
- Draws sticks between compatible features to form candidate regions. Then weeds out false positives using affinity functions that quantify how compatible features are in relation to each other.
- Advantages: Low false negative rate. Doesn’t need subwindows.
Final Process

- Use Shen-Coughlan segmentation algorithm to find candidate text regions.
- Use AdaBoost to categorize each segment.