Detecting and Reading Text in Natural Scenes

Louka Dlagnekov

October 19, 2004

X. Chen, A. L. Yuille

Outline

- Goals
- Classifiers
- Boosting
- Optimizing
- Binarization
- Questions

Introduction

Given an image of an outdoor scene, goals are to:

- Identify regions where there is text
- Extract the text using OCR
- Convey this information to a blind person

Example

Main Ideas

- Build a strong classifier trained by AdaBoost
- Apply classifier to sub-regions of the image
- Binarize candidate text regions
- Use OCR on binarized candidate text regions

Results

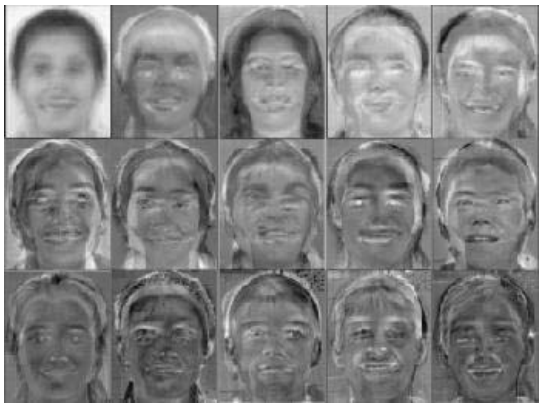
- Reasonably fast – 3 seconds on 3MP image
- 2.8% flase negatives
- 10% false positives
- 93% accuracy of OCR

General Idea

- Find good features of regions containing text that sets them apart from other regions
- Construct classifiers that classify using these features
- Each classifier can be weak (slightly higher than 50% accuracy)

How do we come up with text features in image? [1/3]

- Text has few common features with faces
- PCA Analysis leads to far more non-zero eigenvalues

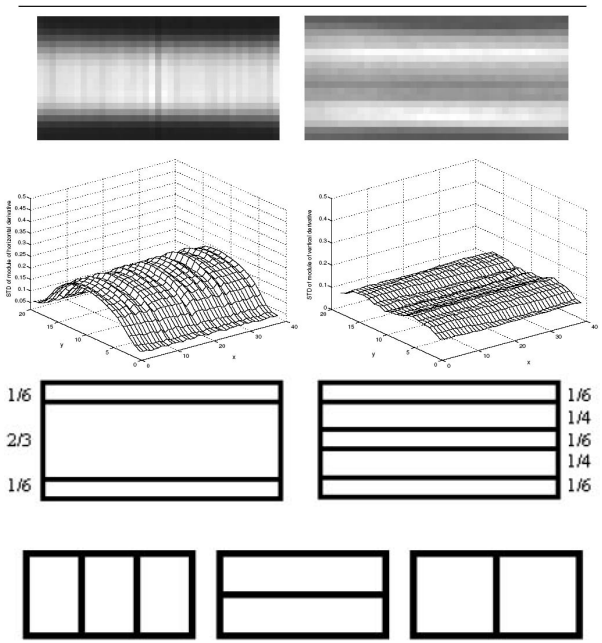


1

[1] http://www.geop.ubc.ca/CDSST/eigenfaces.html

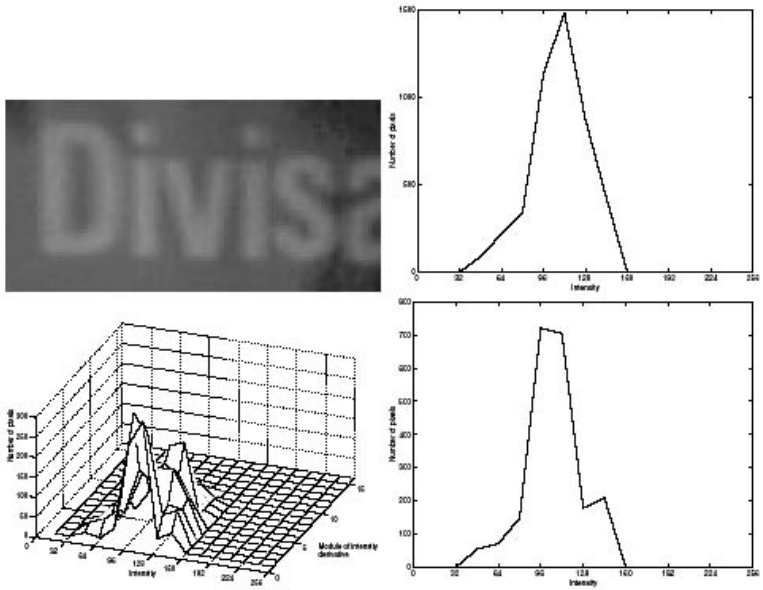
How do we come up with text features in image? [2/3]

- Examine x and y derivatives



How do we come up with text features in image? [3/3]

- Histogram of pixel intensities



- Edge detection ⇒ intensity gradient thresholding ⇒ edge linking

What is boosting?

Start with an example:

- junk email classification

Basic idea:

- finding many rough rules of thumb is easier than finding single highly accurate rule

History on boosting

Gist of AdaBoost

- AdaBoost algorithm is used to combine weak classifiers to make a strong classifier
- Each weak classifier produces a yes/no answer for a particular feature
- Need to come up with lots of features and let AdaBoost decide how to combine them

Algorithm

Given a set of examples $(x_1, y_1), ..., (x_N, y_N)$, where $x_i \in X$, and $y_i \in Y = \{-1, +1\}$:

1. Initialize $D_1(i) = 1/N$
2. For $t = 1, ..., T$:
   1. Train weak classifier using distribution D_t
   2. Obtain weak hypothesis $h_t : X \mapsto \{-1, +1\}$ with error ϵ_t
   3. Choose $\alpha_t = \frac{1}{2} ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$
   4. Update distribution:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

3. Output final hypothesis: $H(x) = sign \left(\sum_{t=1}^{T} \alpha_t h_t(x) \right)$

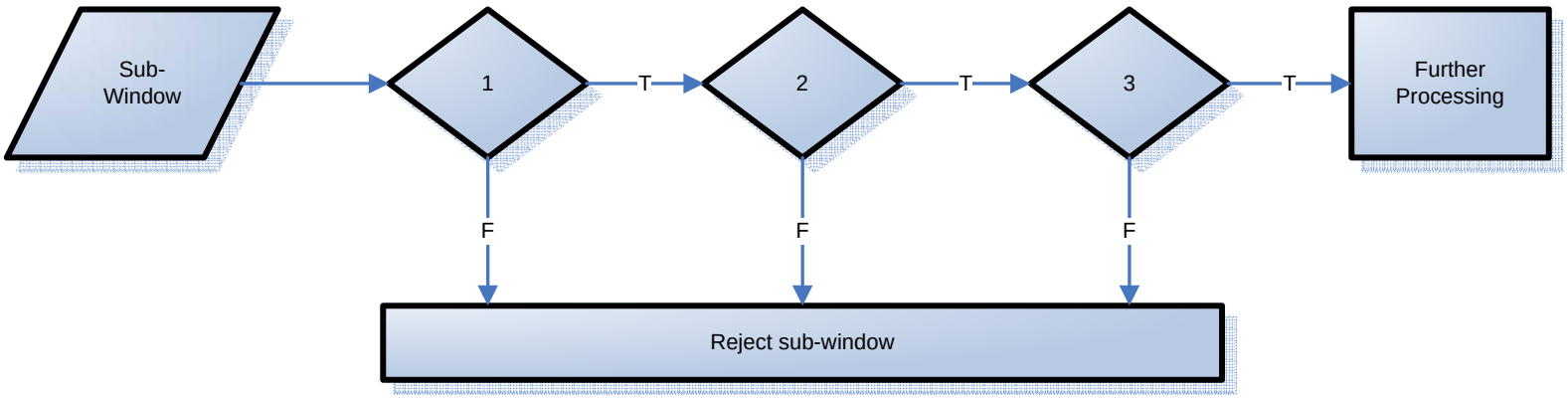
Error

- $Z_t = \sum_i D_t(i) e^{-y_i h_t(x_i)}$
- Upper bound on training error of strong classifier is $\prod_t Z_t$.
- Minimizing $\prod_t Z_t$ is also equivalent to minimizing overall classification error

Problems with AdaBoost

- Minimizes classification error - not number of false negatives.
- To "fix", Viola and Jones propose modifying distribution D_t – give more weight to positive examples.
- They call this modification Asymmetric AdaBoost.

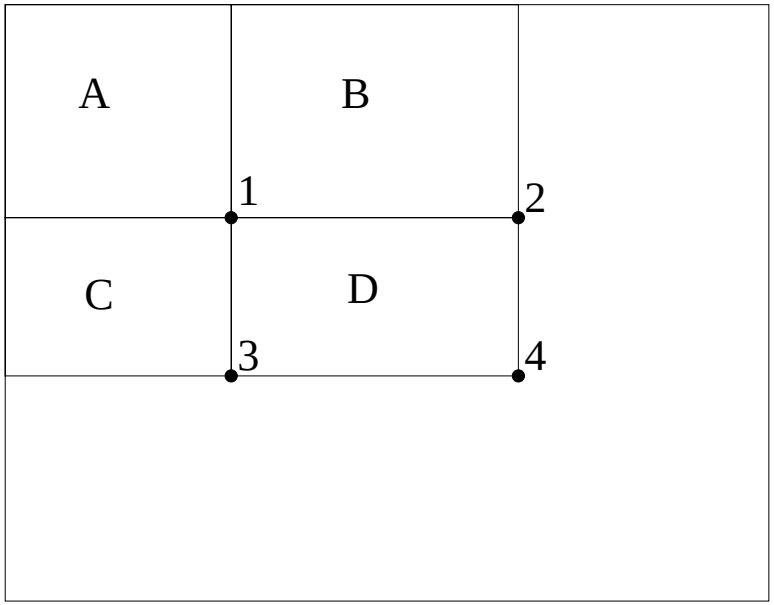
Cascade Classification



Integral Images

- First 3 layers of Chen and Yuille cascade use only mean, STD, and derivative features
- These are easily calculated from *integral images*

Integral Images



- $D = I_4 + I_1 - (I_2 + I_3)$

Binarization

- Use Niblack's adaptive binarization method

$$T_r(x) = \mu_r(x) + k\sigma_r(x)$$

- with a modification:

$$r(x) = min_r(\sigma_r(x) > T_\sigma)$$

Questions [1/3]

- Is AdaBoost the best thing out there for detecting objects? Does it not depend on features used?

Detector \ False detections	10	31	50	65	78	95	110	167	422
Viola-Jones	78.3%	85.2%	88.8%	89.8%	90.1%	90.8%	91.1%	91.8%	93.7%
Rowley-Baluja-Kanade	83.2%	86.0%	-	-	-	89.2%	-	90.1%	89.9%
Schneiderman-Kanade	-	-	-	94.4%	-	-	-	-	-
Roth-Yang-Ahuja	-	-	-	-	(94.8%)	-	-	-	-

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

2

[2] Viola & Jones. Robust Real-time Object Detection.

Questions [2/3]

- Sensitivity of the algorithm to the input data and parameters?
- Modulus of derivative feature?!? What is that?

Questions [3/3]

- Does the order of weak classifiers used matter in AdaBoost?
- What about perspective distortion?