Text Detection and Localization from Complex Background

Mingmin Zhao, Yisong Chen

Key Lab. of Machine Perception (MoE), Peking University, Beijing, China

zhaomingmin@pku.edu.cn

Overview

- Text in images and video frames is one of the most powerful sources of high-level semantics once the text can be detected, localized and recognized automatically.
- In this paper, we develop a two-stage method to detect and locate text lines from complex background.
- The experiment on challenging datasets from ICDAR shows that this is a fast and robust method to detect and locate text lines in complex background.

Our Method

- A coarse-to-fine two-stage architecture
- At the coarse stage, our method use several basic features of text to select candidate text lines.
- At the fine stage, using candidates from the first stage as input, an SVM classifier based on HOG feature and co-occurrence matrix features identify true text lines from candidates.

Properties of Text

- Edge:
  - Most texts are designed to be striking and easily read.
  - Thereby resulting in strong edges at the boundaries of text and background.
- Color:
  - Every single character tends to have the same or similar color.
- Edge:
  - Texts appear in clusters.
  - Characters within a text line tend to have similar size and are always aligned.

Pipeline of Coarse Detection

The goal of coarse detection is to find text-like textures from original image.
We developed a method to find text-like textures as follow:

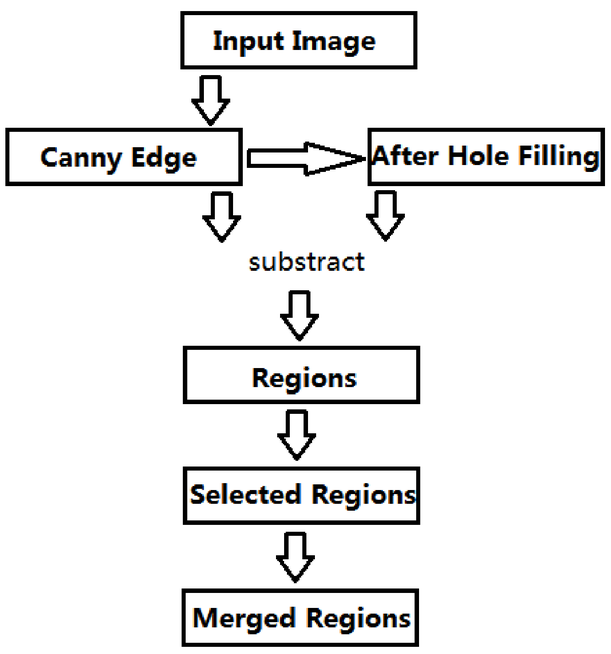

Fig. 1: Pipeline of coarse detection

Coarse Detection

Fig. 2: An example of coarse detection

Features for Fine Detection

- HOG features after splitting:

Fig. 3: An example of splitting

- Co-occurrence Matrix features:
  - Statistics:

$$Contrast : \sum_{i,j} |i-j|^2 p(i,j) \quad (1)$$

$$Correlation : \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \quad (2)$$

$$Energy : \sum_{i,j} p(i,j)^2 \quad (3)$$

$$Homogeneity : \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (4)$$

  - Different offsets are used:

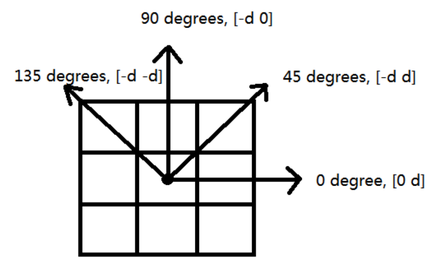

Fig. 4: Offsets used for different matrix

Accuracy of Classification

	Splitting	No Splitting
accuracy	96.0%	94.9%

Table 1: Accuracy of SVM based on HOG features

step	1	2	3	4	5	6
accr	.86	.87	.882	.897	.899	.90

Table 2: Accuracy of SVM based on co-occurrence matrix features

Experiment

We use dataset with about 1000 pictures from International Conference on Document Analyze and Recognition (ICDAR) to train and test.

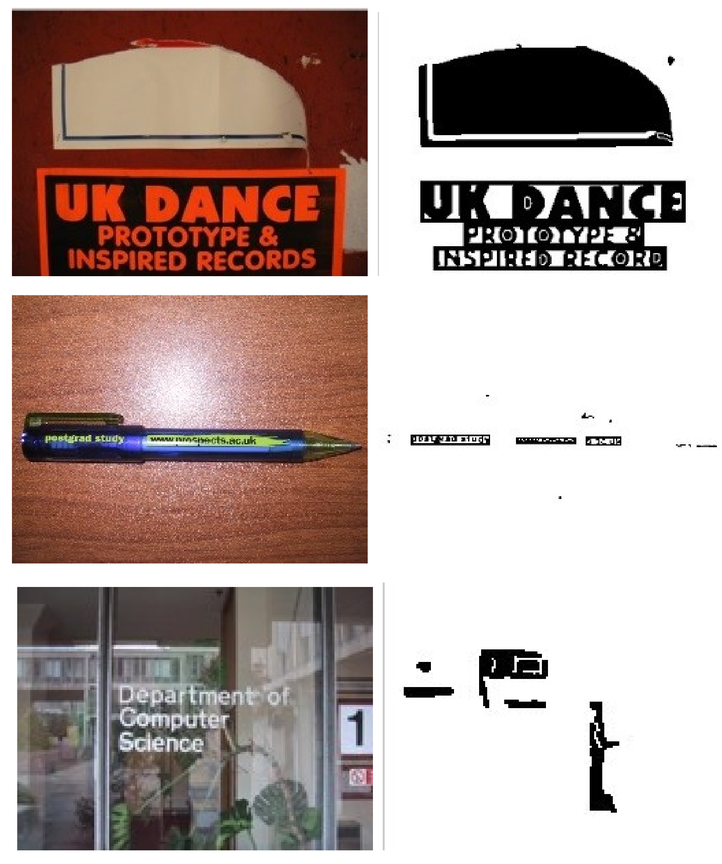

Fig. 5: Good and bad examples

Future Work

- Unsupervised Feature Learning also known as Deep Learning gives a promising way to find intrinsic feature of text pattern.
- Try to use a more efficient detection strategy like classifier cascade.

Acknowledgements

The author acknowledge resourceful discussion and support from professors and students from Key Lab. of Machine Perception (MoE),without whom this work would never have been started and finished.