Con-Text : Text Detection Using Background Connectivity for Fine-Grained Object Classification

Sezer Karaoglu, Jan C. van Gemert and Theo Gevers

Intelligent Systems Lab. Amsterdam (ISLA), University of Amsterdam, The Netherlands

OVERVIEW

Goal

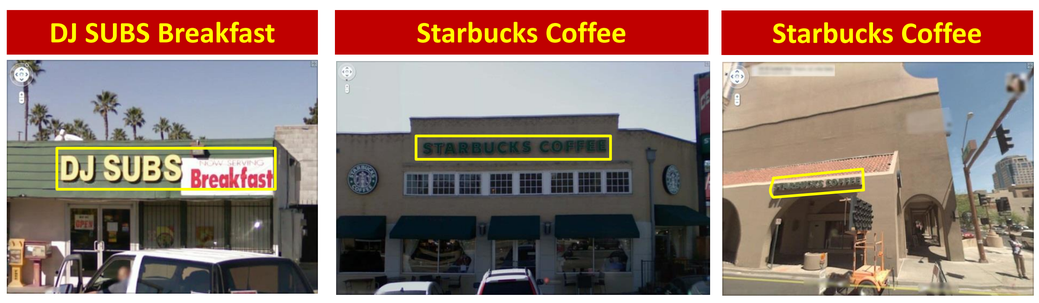
Exploit hidden details by text in the scene to improve visual classification of very similar instances.

Approach

- A novel text detection algorithm using background connectivity.
- Additional semantics using the scene text.

Key Idea

- When text is present in natural scenes, it is typically there to give semantic meaning beyond what is obvious from exclusively visual cues.
- Rather than trying to detect all variations in text appearance, we propose to detect the background.

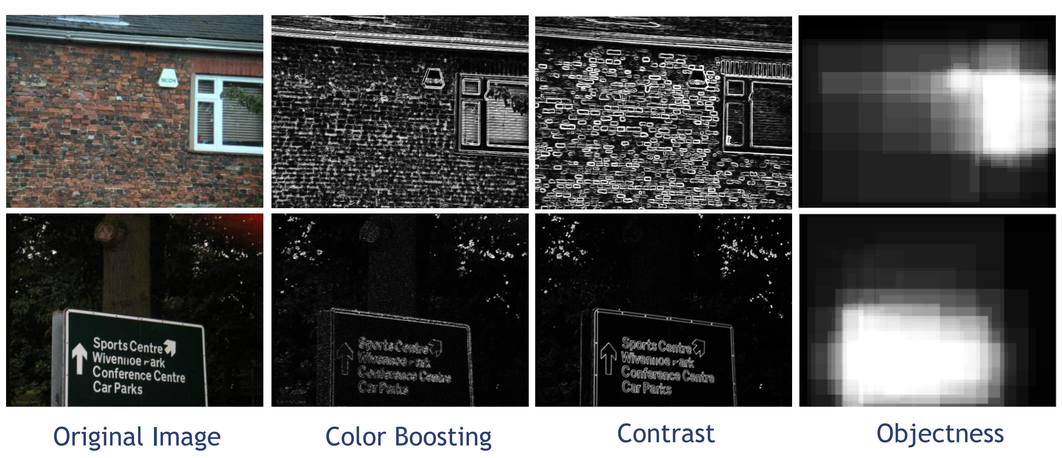


Motivation to Remove Background for Text Detection

- To reduce majority of image regions for further processes.
- To reduce false positives caused by text like image regions (fences, bricks, windows, and vegetation).
- To reduce dependency on text style.

METHODOLOGY

Automatic BG seed selection



- Color, contrast and objectness cues are used in combination with Random Forest classifier to detect background pixels.

BG Reconstruction

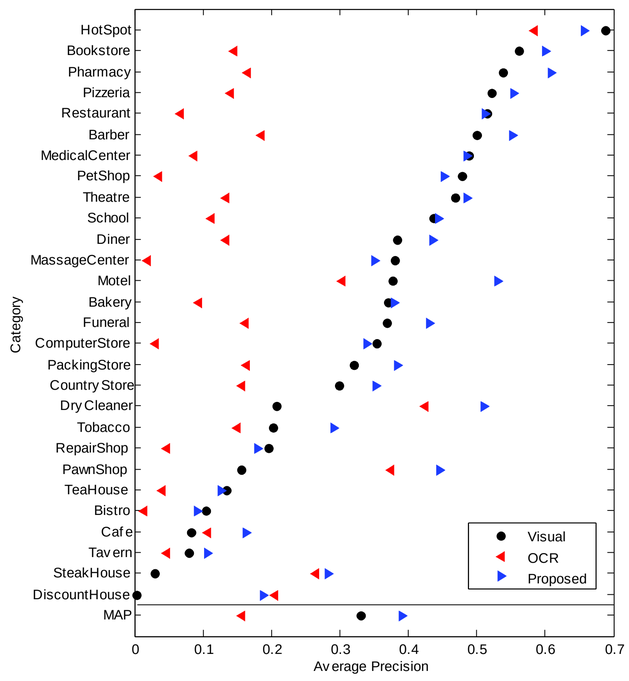
- Intensity-level pixel connectivity with conditional dilation is used to reconstruct the background.

EVALUATION 1: ICDAR 2003 DATASET

- ICDAR03 dataset contains 249 images with 5370 annotated characters.
- Improved ABBYY character recognition from 36% to 63%.
- 87% of the non-text regions are removed where on average 91% of the test set contains non-text regions. It retains approximately 98% of text regions.

EVALUATION 2: ImageNet DATASET

- ImageNet *building* and *place of business* dataset contains 24255 images with 28 different classes.
- The dataset is the largest ever used for scene text recognition.
- **Visual features :** 4000 visual words, standard gray SIFT only.
- **Text features:** Bag-of-bigrams , ocr results obtained for each image in the dataset.



TEXT (ocr)	VISUAL (BOW)	FUSION
15.6 ± 0.4	32.9 ± 1.7	39.0 ± 2.6

CONCLUSION

- Background removal is a suitable approach for scene text detection.
- Color, curvature and objectness prove valuable cues for background modeling.
- A new fine-grained classification problem is introduced based on ImageNet subcategories and a baseline for further research is build.
- We have shown that multimodal information fusion of visual and textual cues improves fine-grained classification on this dataset by 6%.

ACKNOWLEDGEMENT

This publication supported by the Dutch National program COMMIT.

CONTACTS

E-mail : s.karaoglu@uva.nl

Website : http://staff.science.uva.nl/~sezerk/

UNIVERSITY OF AMSTERDAM

COMMIT/