

Leveraging Location-Based Data to Aid Image Text Recognition

James Lintern
UCSD Department of Computer Science
La Jolla, CA 92093
jlintern@ucsd.edu

Abstract

The city environment is rich with signage containing written language that helps us to define and understand the context of a given location. Any photograph of a city scene will contain numerous segments of text that help the viewer/reader to describe, to navigate, and to interact with their environment. Unfortunately, almost all of this information is unreadable to current text recognition technology. Outside of the limited scope of document OCR, image text recognition is very limited in its ability to read unconstrained text “in the wild” with substantial variation in image lighting, angle, size, text lexicon, and more. The goal of this project is to leverage Google Street View as a way to map between existing high level location information (e.g. street names given on a map) and raw image data containing text (e.g. the digital image of a street sign). Currently, those two types of data can only be connected by human cognition. Text recognition could be made easier by knowing the context of a given environment, and location information can be improved if more trust can be placed in methods that process and map text that is recognized in photographs.

1. Qualifications

I am a fifth-year Computer Science student. I have taken courses in Image Processing (CSE 166), Artificial Intelligence (CSE 150), and Natural Computation (COGS 118A, B). I have also worked as a developer on the Ubiquitous Presenter project at UCSD under Bill Griswold and Beth Simon for several years.

2. Milestones

Week 2 – Collect data from Google Street View. I will use available APIs to gather raw images and location-based information from small geographic city areas.

Week 5 – I will implement existing techniques and test

my implementation on my raw image data to generate baseline results.

Week 6 – I will use data about the surrounding environment from Google Maps, such as street names and business names, as an aide to the lexicon for text recognition to see its effect on the results.

Week 9 – I will incorporate highly specific local data (position and orientation of photograph) into my recognition algorithm. The Street View image set has the advantage that the orientation and position of each photograph are known. Additionally, there will be multiple overlapping photographs of any given sign. Recognition can probably be improved by leveraging the features of the data set rather than treating each image separately.

End of Project – I will finish adjusting my recognition algorithm and compare it to both a basic OCR algorithm as well as an algorithm designed for image text recognition. I will also explore possible uses and possible future effectiveness of an algorithm that could recognize image text with a high degree of confidence.

3. Research Questions

- How effective are current techniques for recognition of text in images of a busy city environment?
- How can map information, such as nearby business and street name information be used to improve recognition of image text?
- How can multiple images of a segment of text from different angles be used to aid recognition?
- What useful purposes could there be for an automated algorithm for recognizing and locating text “in the wild”?

4. Existing Software and Data

I will be implementing this project in Matlab. I will directly borrow techniques and code as much as I can to establish a working baseline for my project before attempting to move forward.

My data will be coming from Google Street View and Google Maps. There are APIs to work with this dataset which I'll use to mine a set of images at the beginning of the project.

5. Related Work

[1] R.G. Casey and E. Lecolinet. A survey of methods and strategies in character segmentation. *IEEE Trans. PAMI*, 18(7):690-706, 1996.

[2] X. Chen and A. L. Yuille. Detecting and reading text in natural scenes. *In CVPR (2)*, pages 366–373, 2004.

[3] T. de Campos, B. Babu, and M. Varma. Character recognition in natural images. *In VISAPP, Feb. 2009*.

[4] V. Vanhoucke and S. B. Gokturk. Reading text in consumer digital photographs. *In SPIE, volume 6500, Jan. 2007*.

[5] Wang, Kai. Reading Text in Images: an Object Recognition Based Approach