Recognition of text in images
Project proposal

Johnny CHEN, Haotian CHU, Marouen CHAFROUDA

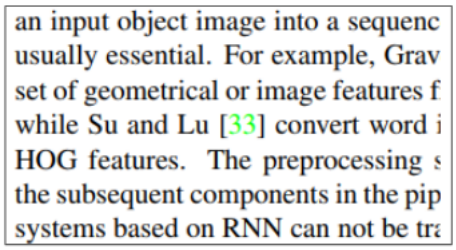
February 10, 2020

1 Problem definition and motivation

The problem is the following: Having as input an image with some text, we want to be able to extract the words written on it, not just detecting the bounding box of the word, but what letters, what words compose the text.

Our first thoughts about the context:

- We could first restrict the language of the text to english, and see later if the language affects the result.
- Different kind of images could be used: wild ones or more structured ones:



- start with simple images: just one word per image first.

Text recognition, or rather OCR for Optical Character Recognition, has been studied from a very long time ago, ever since the 20th century. It has both computer vision focused solutions and deep learning focused solutions.

Text recognition has great use in travelers' everyday lives. For exemple, at the restaurant faced with a lot of unknown foreign words, instead of trying to translate every word, you could just image it and extract all the words, then combine it with Natural language processing such as translation to have a rapid translation. It is even more useful if you are not used to writing foreign words.

There are already mobiles apps developped for that, though most of them have to focus on one word at a time.

We think this could be a good project in order to put into practice our computer vision knowledge learned in class. The idea is not to have an application which has high performance, but rather to understand the concepts of computer vision behind, try to implement those concepts, and see their influence.

2 Methodology

Our principal focus will be to understand the theory behind first, and then try to implement each part of the theory with Python. As there is no available code, we will have to implement everything from scratch.

The papers proposes classical steps in computer vision:

- pre-processing such as de-noising
- edge detection using for exemple sobel operator
- Selective local thresholding
- Text area enhancement
- Coarse-to-fine detection

3 Dataset

For now, we don't have specific datasets, but we could look into datasets such as: SVHN, license plates, MNIST

4 Evaluation

For now, we don't have any metrics besides visualisation.

5 Reference

- "A NEW APPROACH FOR VIDEO TEXT DETECTION" `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1037973&tag=1`
- "A Technical Review on Text Recognition from Images" `\https://www.researchgate.net/publication/274838198_A_Technical_Review_on_Text_Recognition_from_Images`