Object recognition and computer vision

Vincent BODIN & Thomas MOREAU

December 11, 2013

Abstract

blabla

Contents

Introduction

For

1 Methodology

We decided to implement the paper [1]. Here is frame of how the text detection is performed through the paper:

1. Step 1: Train classifiers for each possible character c_k in $\mathcal{K}$ set of all possible character. The database used was ICDAR 2003[1].
2. Step 2: perform a sliding window detection to detect every possible character. For this we used several scales of windows, computed a feature (HOG) and tested this with our trained classifier. Almost every character should be detected so far - hence we take a very laxist criterion.
3. Step 3: build a graph with all the windows we had. Attribute a certain energy to each sample of this graph, minimize it via the TRW-S algorithm [**?**]. The word detected is the one with minimum energy.

1.1 Training the classifier

The database used to train the classifier is the database ICDAR 2003. There are two different things to learn: first how characters look like, and second how they connect to one another, *i.e.* the lexicon prior.

1.1.1 Training for character detection

We have to learn single characters. They are characterized by their Histogram of Oriented Gradient (HOG) and also by their aspect ratio. The method used in this paper is to train K SVMs with RBF Kernel through a one-versus-all procedure.

1.1.2 Train a prior lexicon for language model

Two methods are described in the paper to learn a prior lexicon.

- Bi-gram: the idea is to count how many times in our dictionary a pair of characters occur. A pairwise energy will then be given according to this prior.
- Node-specific: this method is supposed to perform a bit better than bi-gram since it takes into account the position of occurrence of a pair in a word.

1.2 Sliding window detection

1.3 Graphical model of the language

2 Results

[1] ICDAR 2003 database can be downloaded at:

References

[1] C.V. Jawahar A. Mishra, K. Alahari. Top-down and bottom-up cues for scene text recognition. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2687 – 2694, June 2012.