SCENE TEXT DETECTION VIA CONNECTED COMPONENT CLUSTERING AND NONTEXT FILTERING

Abstract:

In this paper, we present a new scene text detection algorithm based on two machine learning classifiers: one allows us to generate candidate word regions and the other filters out no text ones. To be precise, we extract connected components (CCs) in images by using the maximally stable external region algorithm. These extracted CCs are partitioned into clusters so that we can generate candidate regions. Unlike conventional methods relying on heuristic rules in clustering, we train an AdaBoost classifier that determines the adjacency relationship and cluster CCs by using their pair wise relations. Then we normalize candidate word regions and determine whether each region contains text or not. Since the scale, skew, and color of each candidate can be estimated from CCs, we develop a text/nontext classifier for normalized images. This classifier is based on multilayer perceptions and we can control recall and precision rates with a single free parameter. Finally, we extend our approach to exploit multichannel information. Experimental results on ICDAR 2005 and 2011 robust reading competition datasets show that our method yields the state-of-the-art performance both in speed and accuracy.

Existing method:

Since mobile devices equipped with high-resolution digital cameras are widely available, research activities using these devices in the field of human computer interaction (hci) have received much attention for the last decades. among them, text detection and recognition in camera captured images have been considered as very important problems in computer vision community. it is because text information is easily recognized by machines and can be used in a variety of applications.
**Proposed method:**
Augmented ground truth and frequently used notations, construction of new ground truth, frequently used notations, candidate generation, candidate normalization, geometric normalization, comparison to other mser-based methods, binarization, multilayer perception learning.

**Merits:**
- One classifier was designed to generate candidates and the other classifier was for the filtering of non-text candidates.

**Demerits:**
- To infer text blocks from CCs, to filter out non-text CCs, to extract text-like CCs.