

Pixel-level Features for Scene Text Detection

高, 仁武

<https://doi.org/10.15017/1543996>

出版情報：九州大学, 2015, 博士（学術）, 課程博士
バージョン：
権利関係：全文ファイル公表済



氏 名 : 高 仁武

論 文 名 : Pixel-level Features for Scene Text Detection
(情景内文字検出のための画素レベル特徴)

区 分 : 甲

論 文 内 容 の 要 旨

Scene text recognition has been studied for decades and is still an open challenge. This is mainly because of the difficulty of detecting (locating) texts in natural scenes. The task of detecting text in natural scenes not only suffers from texts having various character appearance, noise and distortion, segmentation, etc., as OCR does, but also suffers from separating texts from the complex and unpredictable background. Consequently, to overcome the above mentioned difficulties, in this thesis, two kinds of pixel-level features (visual saliency and color feature) are studied and utilized to detect scene texts. Unlike the conventional features, such as HOG or SWT, pixel-level features, as the word itself indicates, are features that calculated in the pixel level, therefore are supposed to be independent from text font, distortion and, even to some extent, background. The results show that, pixel-level features provide a new way for the task of detecting text in natural scenes.

In Chapter 2, two image database that are used in the experiments are introduced. To investigate the pixel-level features of texts, a big enough scene image database with manually labeled pixel-level ground-truth is created; to give a comparison with other state-of-the-art methodologies, a public database (ICDAR-2013) is also used.

In Chapter 3, the previous related work in the literature of scene text detection is reviewed according to the feature been used. We categorize features used for scene text detection into three categories: low level, middle level and high level. According to this categorization, the related work is introduced.

In Chapter 4, in order to investigate the visual saliency features (pixel-level feature) of scene texts, five state-of-the-art visual saliency models (Itti *et al.* 's bottom-up visual saliency model, Harel *et al.* 's graph-based visual saliency model, Torralba *et al.* 's top-down visual saliency model, Zhang *et al.* 's model, and Achanta *et al.* 's frequency-tuned model) are utilized. The experimental results show that pixels belonging to texts are more salient than those belonging to non-texts. This means that visual saliency pixel-level feature can be used for the task of scene text detection. The experimental results also show that among the five visual saliency models, Torralba *et al.* 's visual saliency model can retrieve the most texts in terms of pixels. Based on Itti *et al.* 's bottom-up visual saliency model, a new model called hierarchical visual saliency model, which captured more texts in our experiments, is proposed. The proposed model first find the salient region, then focus on the salient region, Itti's model is used again to

find salient regions.

In Chapter 5, another simple but useful pixel-level feature, color feature, is studied. Using the big database, three non-parametric distributions (text color against to non-text color) in HSV color space are estimated. The experimental results show that saturation and value channel distributions are usable for scene text detection. To demonstrate how to use color feature to detect natural scene text, three experiments are done.