Hypothesis Preservation Approach to Scene Text Recognition with Weighted Finite-State Transducer

Takafumi Yamazoe, Minoru Etoh, Takeshi Yoshimura, and Kousuke Tsujino
Service & Solution Development Department and Research Laboratories, NTT DOCOMO 3-6, Hikarino-oka, 239-8536 Japan
yamazoet at nttdocomo.com, {etoh, yoshimura.takeshi, tsujino} at nttdocomo.co.jp

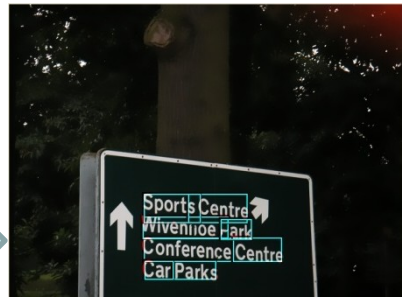
The demonstration contains two mobile applications that shows scene text recognition. The applications, automatic image repository tagging and real-time multilingual menu recognition, use a language model based method which the authors newly propose at ICDAR 2011. The proposed method uses Weighted Finite-State Transducer (WFST) that greatly suppresses large-scale ambiguity in scene text recognition, especially for Japanese Kanji characters. The details appear as the same title in ICDAR 2011 proceedings.

DEMO1: IMAGE TAGGING

This shows a cloud-based prototype service that recognizes photo galleries on a mobile through the recognition of words in the scene images. The system is commercially available on the web for Japanese general public. <http://tangochu.jp/en/>

Over-extracted text regions

Elimination of Incorrect hypothesis

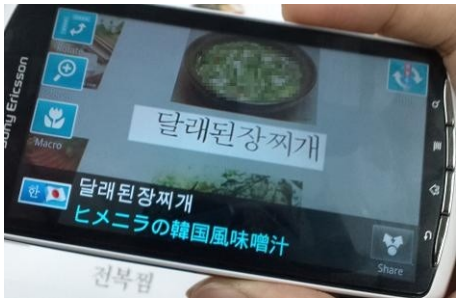


Recognition results

Recognition results (Japanese lexicon)

DEMO2: REAL-TIME MULTILINGUAL MENU RECOGNITION

The developed portable system locally extracts and recognizes multilingual scene texts in real-time. The target language and text types are generic though, the demonstration is tuned to local menu translation for which the system combines multiple language processing based on WFST and an existing OCR engine for traditional Chinese, simplified Chinese, Korean, English, and Japanese.



Application for the food menu translation