Visual Attention is Attracted by Text Features Even in Scenes without Text

Hsueh-Cheng Wang, Shijian Lu, Joo-Hwee Lim, and Marc Pomplun

Department of Computer Science, University of Massachusetts at Boston, USA

Department of Computer Vision and Image Understanding, Institute for Infocomm

Research, Singapore

Abstract

Previous studies (Cerf, Frady, & Koch, 2008; 2009; Wang and Pomplun, 2011; under revision) suggested that viewers' attention is disproportionately attracted by texts, and one possible reason is that viewers have developed a "text detector" in their visual system during everyday life so that the features of texts are more sensitive to viewers. An approach to verify this hypothesis is to use a visual attention model including text detector. Manually-defined object regions for text, face, and cellphone, and automatic object detectors for car, people, or face have been shown to be able to predict where people look during scene viewing (Cerf, et al., 2009; Judd et al. 2009). It has not been tested if a model including automatic text detector, similar to other object detectors, is able to predict viewers' visual attention better. A model including text-detector (Lu, submitted), saliency (Itti & Koch, 2001), and center-bias (Tatler, 2007) is found to predict viewers' eye fixations better than the model without text detector using correlation and ROC analysis, even in text-absent images. In addition, the text detector designed for English texts predicts the attention for both English- and Chinese-speaking viewers' attention but predicts better for English-speaking viewers. The results support that the ability of detecting texts is trained according to viewers' experience.

Keywords: real-world scenes; text detector; eye movements; visual attention.