Real-life clickable text

Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise

A camera-based interface detects and identifies text in a scene, providing additional relevant information on request.

Researchers have been pursuing the goal of effective character recognition for decades, and tools such as optical-character-reader (OCR) software have influenced everyday life. While flat-bed scanners have been able to capture character images with OCR for a long time, research into capturing characters with cameras is still ongoing. Recently, the research field of camera-based character recognition has grown because of the rapid growth in popularity of portable cameras and mobile phones that incorporate a camera.

Camera-based character recognition has enormous potential for everyday applications. For example, it could enable us to point a camera at text in a foreign language and obtain an instantaneous translation: see Figure 1(a).

This would be helpful when traveling but also for checking dictionaries or translating a foreign-language book. Another possible application is a voice-navigation service for visually disabled people—see Figure 1(b)—who need help to use information provided as text. The service tells the user where texts are located and reads them out. However, it would be annoying to be notified of all texts in a scene, and so it would be preferable to be alerted only to pre-registered words or phrases, at least initially.

To realize such applications, we have developed a quick, robust, and diverse (in terms of fonts) recognition technique for camera-captured characters. Real-time processing is essential for usability. In addition, it is vital to maintain accuracy under real-life conditions such as uneven lighting, occlusion, perspective distortion, and low resolution. A successful system should also be able to decode complex layouts and fonts, such as the rounded text logo of Starbucks coffee. Our system can recognize 220 camera-captured characters per second and has achieved a 95.8% recognition rate for characters captured from a slant angle of 45°. It can recognize 100 fonts and nonstraight, including rounded, texts.

We have developed a prototype system of a new interface that enables us to ‘click’ the physical world to obtain more information, like clicking anchor text on a web browser (see Figure 2).

For instance, when the word ‘Hawk’ is pointed at with a camera, as in Figure 2, it provides the Japanese translation, an image of the bird and an audio file of its cry. The system works on a normal laptop PC in real time.

The detailed process is as follows. First, character regions are extracted from the captured image—see Figure 3(A)—by a simple binarization method. In Figure 3(B) these are colored green, and in Figure 3(C) the recognition results are superimposed on the original characters. The system employs a robust recognition technique to cope with distortion caused by perspective and degradation, and so character images captured at a given range of angles are recognized. In addition, the system can estimate

Continued on next page
Figure 3. Screenshot and process of the prototype system. (A) Character regions are extracted from the scene. (B) These are colored green. (C) Recognized characters are superimposed. (D) Word regions are colored orange or purple. (E) The word under focus is orange. (F) The recognized word is shown in a different window on the computer. (G) The Japanese translation of the word is displayed. (H) A related image may also be displayed.

the deformation of each character and apply the same estimated deformation to the superimposed character.

The system recognizes the captured characters frame by frame. Once each character has been recognized, word regions are identified, extracted by the same binarization method, and regarded as ‘anchor texts.’ Word regions are colored orange and purple: see Figure 3(D). Orange is used to identify the word under focus, that is, the one the cursor is on: see Figure 3(E). Then the recognition result of the focused word is shown in a different window: see Figure 3(F). Our prototype system shows the Japanese translation of the focused word, if the word exists in the dictionary: see Figure 3(G). Finally, it will also display an image related to the focused word as related information: see Figure 3(H).

In summary, the prototype system shows the potential of camera-based character recognition to aid foreign travelers or the visually impaired to make sense of text they find in their ambient environments. We are now working to improve recognition accuracy and adapt the system to other languages.

The authors are grateful for the support of Grant-in-Aid for Scientific Research (KAKENHI) 21700202 and Research for Promoting Technological Seeds of the Japan Science and Technology Agency (2009).

Author Information

Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise
Department of Computer Science and Intelligent Systems
Graduate School of Engineering
Osaka Prefecture University
Sakai, Japan
http://imlab.jp/camocr/

Masakazu Iwamura completed his PhD in 2003 at Tohoku University (Japan). He is currently an assistant professor.

References