May 02, 2019

Text Scanner and Touch Reader for Visually-Impaired Users

Zhenping Cui

Text Scanner and Touch Reader for Visually-Impaired Users

Abstract:

This publication describes a text scanner and touch reader application software for visually-impaired users (e.g., blind or partially blind) that may help them gain a higher degree of independence and privacy. As the visually-impaired user moves a user equipment (UE) around a scene with text, the application software records the scene with text. To aid the visually-impaired user record the whole text of interest, the application software verbally instructs the visually-impaired user on how to move the UE. Once the application software collects an adequate input, it instructs the visually-impaired user to stop recording. The application software uses machine learning to analyze the captured scene with text to generate a full picture of the scene with text. Then, the application software runs optical character recognition (OCR) over the full picture of the scene with text. The visually-impaired user may choose to have the UE read aloud the whole text or may slide his or her finger across the UE's screen to hear certain parts of the text, similar to reading Braille.

Keywords:

Visually impaired, blind, partial sight, no vision, vision loss, scan, read, read aloud, speech recognition, text-to-speech, blind, partial sight, no vision, vision impairment, vision loss, scan, record, read loudly, optical character recognition, OCR.

Background:

Optical character recognition (OCR) systems are one of the tools that allow people to access printed information. Various professionals (e.g., bankers, human resource administrators, government workers, medical staff, legal professionals, librarians, historians) widely use OCR to digitize information from printed documents. In addition, some OCR systems can read the digitized information as synthesized speech. There are four essential elements to these systems: scanning, storing, optical character recognition, and reading text as synthesized speech.

Since the 1970s there has been extensive research and development on creating paper-to-digital-text conversions, paper-to-speech conversions, and paper-to-Braille conversions. All these advancements have helped visually-impaired people to hear or read in Braille some printed information. The OCR systems can be scanners, handheld pen-shaped scanners, digital cameras, smartphones, and so forth. Often, visually-impaired people do not perform these text conversions because it is challenging for them to properly scan a document, a street sign, a medicine-bottle label, a packaged-food label, and various other texts that people with good eyesight can easily read with no effort. Instead, they often rely on people with good eyesight to perform the text-to-speech conversions for them.

To increase the independence of visually-impaired people, it is desirable to have a device with proper application software that can help them convert text-to-speech with less effort.

Description:

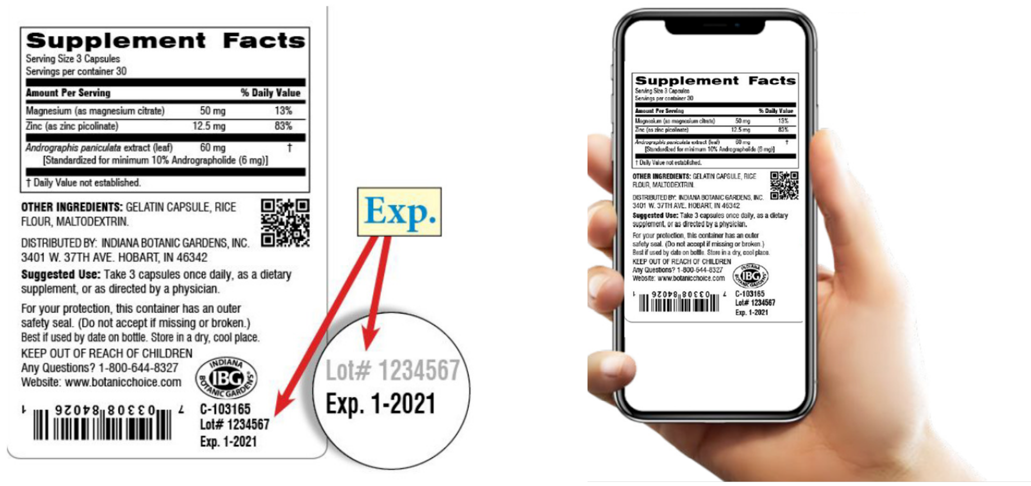
This publication describes an application software of a user equipment (UE), such as a smartphone, which enables a visually-impaired user to record and read aloud texts from documents, books, street signs, medicine-bottle labels, packaged-food labels, and various other texts, without relying on people with good eyesight. The application software leverages existing optical character recognition (OCR) technology to perform text-to-speech conversions. Figure 1 shows the UE with OCR technology.

Figure 1

Figure 1 shows a user who is scanning a document using the UE with OCR technology. A user with good eyesight performs the scanning of the document with relative ease. After the user scans the document, the OCR performs the text-to-speech conversion. A visually-impaired user, however, struggles scanning the whole document because not only he or she cannot see the text, he or she cannot see the document. In some instances, the visually-impaired user may be able to touch the document and that gives him or her some reference where to scan. This, however, is not always possible. For example, the visually-impaired person may be shopping at a pharmacy and wants to find a certain aisle that carries an over-the-counter product that he or she wants to buy.

The visually-impaired user cannot touch the sign to determine where to scan. He or she may resort to asking a pharmacy technician for help, which diminishes his or her independence and privacy.

Even if the visually-impaired user asks the pharmacy technician for directions to the product, he or she cannot locate the product label to perform the scan, as shown in Figure 2.



a) Text of interest **b) Properly scanned text**

Figure 2

Assume the pharmacy technician assists the visually-impaired person finding the desired product. The visually-impaired user cannot see the location of the product label and struggles using his or her UE that supports the OCR technology. The visually-impaired person may face similar challenges even at a familiar location, such as at home. He or she struggles to use the UE to scan for essential information, such as the expiration date of a medicine bottle.

To capture an image similar to Figure 2b, the described application software guides the visually-impaired person by using audible instructions to video record a scene with text. The UE may use a range of frame-frequency (frequency per second) to record the scene with text. Initially, the visually-impaired user starts moving the phone near the area he or she thinks is the text of interest. The application software searches for text and directs the visually-impaired to change the movement to capture the whole text, as illustrated in Figure 3.

a) Part of the text **b) The whole text**

Figure 3

Assume the visually-impaired person has a headache and wants to find a bottle of aspirin in a medicine cabinet. The application software starts reading the impartial text “Aspi”, as shown in Figure 3a. To capture the whole text, the application software instructs the visually-impaired user to “roll the bottle” to record the whole the text, as shown in Figure 3b. The application software video-records the scene, stores the video frames, and instructs the visually-impaired user to stop recording when the UE records an adequate input.

The application software analyzes the video-recorded input using a machine-learning model, such as a neural network, as shown in Figure 4.

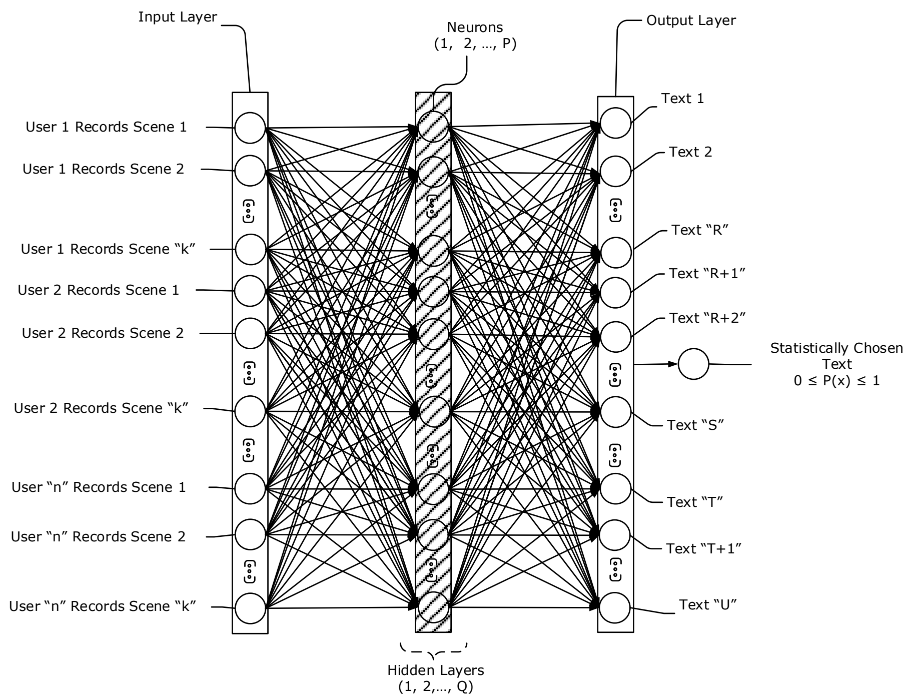


Figure 4

Figure 4 demonstrates a neural network, which is used to analyze a recorded scene with possible text and determines the whole text of that recorded scene. The neural network illustrates an input layer, several hidden layers, and an output layer. The input layer includes "k" recorded scenes from "n" number of users, which are captured by the UE's cameras. The recorded scene of the input layer may represent a recorded scene of a document, a book, a street sign, a medicine-bottle label, a packaged-food label, and various other scenes with text. For example, a visually-impaired user may have already recorded Herman Melville's "Moby-Dick". If another visually-impaired user starts video-recording his or her copy of "Moby-Dick", the application software recognizes that this work has already been text-to-speech converted and instructs the visually-impaired user to stop recording and start listening to the audio of the book. There are "Q" number of hidden layers with up to “P” number of neurons in each layer. There can be a different quantity of neurons in each hidden layer. The output layer includes “U” number of bins with different probabilities on the combined selected frames used to generate a full picture of the scene with text, where U > T > S > R > 0. The neural network interprets as the correct output the bin with the closest probability to one (1). Given the large computational power that machine learning uses to train a model to analyze so many scenes with text, the model training may be performed on a cloud, server, or other capable computing device or system.

Once the machine-learning model has generated a full picture of the scene with text, the application software runs OCR and recognizes the text in the scene. At this stage, the visually-impaired user may instruct the application software to start reading the whole text or may choose to use his or her finger to instruct the UE to read aloud as his or her finger moves across the UE’s screen, similar to reading Braille. Figure 5 shows a visually-impaired user using his or her finger to read part of the text.

Figure 5

In Figure 5, assume the visually-impaired user has already determined what medicine-bottle he or she is holding. The visually-impaired user may want to find out whether the medicine has expired. Instead of listening to the whole text, sliding a finger across the UE's screen, the visually-impaired person can listen for the expiration date.

Additionally, a visually-impaired user may be provided with controls allowing the visually-impaired user to make an election as to both if and when systems, programs, or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the visually-impaired user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a visually-impaired user's identity may be treated so that no personally identifiable information can be determined for the visually-impaired user, or a visually-impaired user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a visually-impaired user cannot be determined. Thus, the visually-impaired user may have control over what information is collected about the visually-impaired user, how that information is used, and what information is provided to the visually-impaired user. The visually-impaired user may instruct the UE on his or her preferences by speaking to the UE that supports the described application software.

In summary, the text scanner and touch reader application software for visually-impaired users may help them gain a higher degree of independence and privacy.