October 06, 2017

Computer vision ring

Nicholas Jonas

Barron Webster

Computer vision ring

ABSTRACT

Visually-impaired people have difficulty reading handwritten or printed documents that are not written in Braille. Totally blind people cannot read handwritten or printed documents that are not in Braille. A large fraction of written language encountered in daily life or travel still does not appear in Braille, thereby denying visually impaired and completely blind people access to such information. The present disclosure describes techniques to scan one's surroundings using a low-profile camera, to identify objects and/or written language using an object or character recognition system, and to convey scanned information aurally to a user. Machine learning and inference models are used to automatically perform the tasks of object identification, character recognition, etc. Such a system can be used by visually impaired and totally blind people, and also by individuals without visual impairment — such as travelers who want to translate written text found in their surroundings, e.g., signboards, menu items, etc.

KEYWORDS

- Object recognition
- Assistive device
- Obstacle detection
- Wearable computing

BACKGROUND

A large fraction of written text found in daily life, e.g., signboards, pamphlets, menu cards, etc. is not accompanied by Braille imprint. This is a problem for totally blind or visually-impaired people, as written information from their surroundings is not accessible to them. Furthermore, obstacle detection and avoidance are of key importance in enabling visually-

impaired and completely blind people to smoothly navigate the routines of everyday life. These objectives, of bringing written information from surroundings to blind people, and enabling accurate and efficient navigation of their surroundings, are ideally accomplished using an apparatus that occupies minimal space and is generally of low profile. Such an apparatus need not be useful only to blind people. People without visual impairment requiring translation of foreign-language text face a similar problem. Currently it is possible to use a smartphone to take a photo of foreign-language text and translate it. However, such translation requires several steps, e.g., pulling out the phone, capturing the photo, calling a translation app, etc.

DESCRIPTION

This disclosure describes a low-profile assistive device that enables easy scanning and identification of objects within one's surroundings. Techniques described herein also enable low-profile and quick, e.g., using a minimal number of steps, translation of text found in one's surroundings.

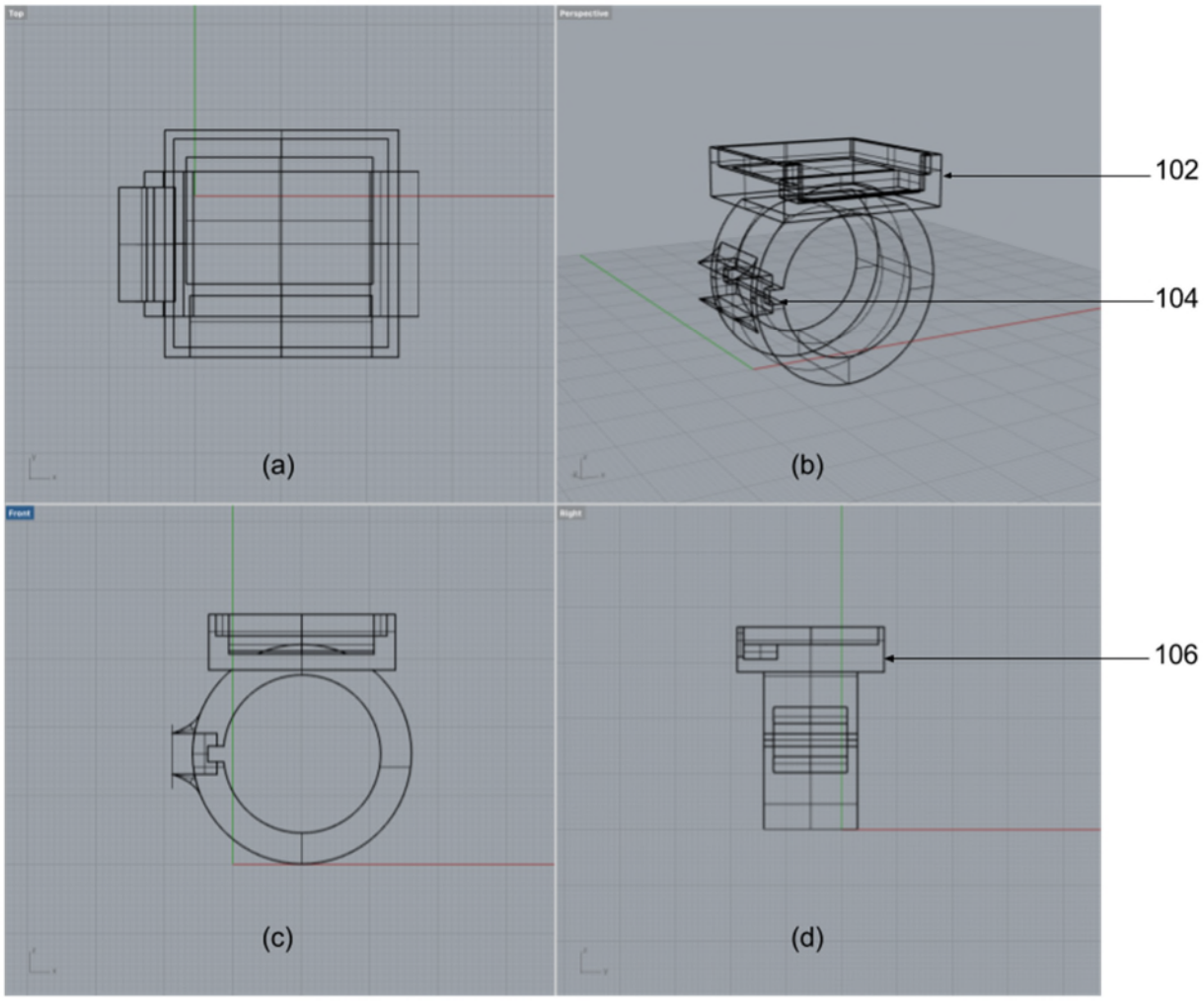

Fig. 1: The computer vision ring (a) top view (b) perspective view (c) front view (d) side view

Fig. 1 shows from top, perspective, front and side views the assistive device, which is a finger ring equipped with camera, haptic sensor, and short-range wireless transceiver. The finger ring is designed to be worn, for example, on the user's index finger between the second and third knuckles. The finger ring is manufacturable through a 3D printing process. A camera (102) is mounted on the top of the ring (pointing away from the palm), such that it faces up (skyward) when the hand is outstretched with palms down. The camera's shutter is triggered by a small button (104) mounted on the side of the ring closest to the thumb.

A haptic motor (106) and a driver embedded at the top of the ring, e.g., below the camera, provide tactile feedback to the user. The ring has the capability of short-range wireless communication, for example via Bluetooth or Wi-Fi. A microcontroller coordinates all components and sensors, and runs any necessary software, e.g., object identification system, optical character recognition (OCR) system, text-to-speech converter, etc. A rechargeable battery, e.g., of lithium polymer (LiPo) variety, provides power.

When the user needs a scan of their surroundings, for example, to determine and avoid obstacles, or to convert-text-to-speech or translate printed text, they press the button using their thumb. This action triggers the camera, which captures an image. The image is fed to an object recognition and/or an OCR inference model.

Detected objects are labeled, and printed text, if any, is translated if necessary. A text-to-speech system is invoked, and labeled objects and (translated) printed text is converted to speech. The generated audio is transmitted over short-range wireless to a paired wireless earpiece, and the original picture is deleted. In this manner, the user receives an aural description of the objects (e.g., tree branch, stairs, etc.) in their immediate surroundings, which they can use to avoid obstacles and navigate safely. Printed text found in the user's surroundings is also provided aurally, translated if necessary to the user's native language.

The device makes use of several machine intelligence technologies, e.g., object identification, OCR, natural language processing and translation, text-to-speech conversion, etc. Fig. 2 depicts a block diagram of an example machine-learned model according to example implementations of the present disclosure. As illustrated in Fig. 2, in some implementations, the machine-learned model is trained to receive input data of one or more types and, in response,

provide output data of one or more types. Fig. 2 illustrates the machine-learned model performing inference.

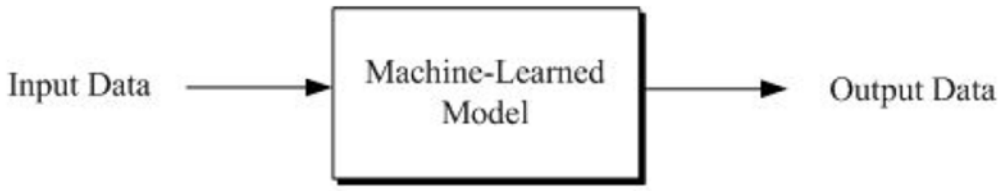

Fig. 2: Block diagram of an example machine-learned model

In some implementations, the input data can include one or more features that are associated with an instance or an example. In some implementations, the one or more features associated with the instance or example can be organized into a feature vector. In some implementations, the output data can include one or more predictions. Predictions can also be referred to as inferences. Thus, given features associated with a particular instance, the machine-learned model can output a prediction for such instance based on the features.

The machine-learned model can be or include one or more of various different types of machine-learned models. In particular, in some implementations, the machine-learned model can perform classification, regression, clustering, anomaly detection, recommendation generation, and/or other tasks.

The machine-learned model can be trained or otherwise configured to receive the input data and, in response, provide the output data. The input data can include different types, forms, or variations of input data. As examples, in various implementations, the input data can include: for object identification, images of the surroundings as captured by the camera on finger ring; for OCR, an image of printed text; for natural language processing and translation, text or speech sample containing a natural language; for text-to-speech conversion, a textual sample.

In some implementations in which the machine-learned model performs classification, the machine-learned model can be trained using supervised learning techniques. For example, the machine-learned model can be trained on a training dataset that includes training examples labeled as belonging (or not belonging) to one or more classes.

In some implementations, the machine-learned model can perform various types of clustering. For example, the machine-learned model can identify one or more previously-defined clusters to which the input data most likely corresponds. As another example, the machine-learned model can identify one or more clusters within the input data. That is, in instances in which the input data includes multiple objects, documents, or other entities, the machine-learned model can sort the multiple entities included in the input data into a number of clusters. In some implementations in which the machine-learned model performs clustering, the machine-learned model can be trained using unsupervised learning techniques.

In response to receipt of the input data, the machine-learned model can provide the output data. The output data can include different types, forms, or variations of output data. As examples, in various implementations, the output data can include: for object identification, labels of objects identified within an image; for OCR, a computer-readable, e.g., ASCII, rendition of printed matter found in an image; for natural language processing and translation, a natural language that is the translated version of the input language; for text-to-speech conversion, speech that simulates a human reading out aloud the input text.

As improved machine-learning inference models for various tasks such as object identification, OCR, natural language processing and translation, text-to-speech conversion, etc. become available, these models are incorporated into the wearable computing device described herein via over-the-air (OTA), e.g., wireless, updates. For example, such models can handle a

larger breadth of classifiable objects or provide improved quality of recognition and/or translation.

Some, none, or all of the machine-intelligence related functionality of the device may be executed at a server. In this case, there are several possible ways by which processes may be divided between server and device. For example, in some implementations, the device provides a photo of the surroundings, which is sent to a server for object and/or text analysis. The server returns labels for the objects and/or translated text, which are then provided aurally to the user via wireless earpiece. In other implementations, the device extracts features of an image it captures of its surroundings and sends the features to the server. The server performs feature analysis, object/text identification, and/or language translation, and sends results of its analysis back to the device for further aural transmittal to user. In still other implementations, the device may perform one or more of object identification, OCR, and natural language translation, while text-to-speech conversion is performed at the server. Other possible configurations are possible that apportion tasks between device and server.

Some versions of the device include a pointer, e.g., a laser, to enable the user to frame the photo. Still other versions of the device include proximity sensors, e.g., based on sonar, to determine distances of objects. Such distances further describe the environment to a completely blind or visually-impaired user, thereby enabling user to navigate more accurately.

In situations in which certain implementations discussed herein may collect or use personal information about users (e.g., user data, data captured by cameras or other sensors controlled by the user, information about a user's social network, user's location and time at the location, user's biometric information, user's activities, user's online history and demographic information), users are provided with one or more opportunities to control

whether information is collected, whether the personal information is stored, whether the personal information is used, and how the information is collected about the user, stored and used. That is, the systems and methods discussed herein collect, store and/or use user personal information specifically upon receiving explicit authorization from the relevant users to do so. For example, a user is provided with control over whether programs or features collect user information about that particular user or other users relevant to the program or feature. Each user for which personal information is to be collected is presented with one or more options to allow control over the information collection relevant to that user, to provide permission or authorization as to whether the information is collected and as to which portions of the information are to be collected. For example, users can be provided with one or more such control options over a communication network. 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. As one example, a user's identity may be treated so that no personally identifiable information can be determined. As another example, a user's geographic location may be generalized to a larger region so that the user's particular location cannot be determined.

CONCLUSION

Techniques of this disclosure provide a low-profile and easy way for blind people to scan their surroundings and listen to what a camera with machine intelligence can recognize. A wearable computing system, comprising camera, microcontroller, short-range wireless transceiver, haptic motor, etc. is embedded into a finger ring. The wearable computer includes capabilities for object recognition, OCR, language translation, text-to-speech conversion, etc., which are enabled by techniques of machine learning and inference. When triggered by the user, the camera captures an image of the surroundings, the object recognizer/OCR labels

objects and text found in the image, the language translator translates text to the user's native language, and the text-to-speech converter converts the resulting text to speech. This speech is provided aurally to the user via a wireless earpiece. In this manner, a blind person can aurally obtain information about their surroundings, enabling them to navigate accurately. In a similar manner, visually-impaired, completely blind, or individuals without visual-impairment can receive aurally information contained within printed matter found in their surroundings, if necessary in their native language.