November 2019

Object recognition based contextual responses to user queries

Aiko Nakano

Object recognition based contextual responses to user queries

ABSTRACT

This disclosure describes techniques to identify specific objects using image analysis techniques and providing contextual responses to user queries based on the recognized objects. The user query is parsed and matched to available information about the recognized object to provide a contextually appropriate response. In addition to object recognition, text recognition can also be performed to obtain information about the object. The described techniques can provide answers to various users that may otherwise be unable to understand how to use an object, e.g., due to inability to read a language, low vision, or other reasons.

KEYWORDS

- Virtual assistant
- Object recognition
- Contextual response
- User intent
- Vision impairment
- Machine vision
- Natural language processing
- Virtual reality
- Head-mounted display (HMD)

BACKGROUND

Text such as food labels, menus, labels on medicine packages, and other text that is printed at a small size presents difficulties for users with vision impairments such as far sightedness, low vision, etc. Further, users that are illiterate or unable to read the language of the printed text cannot understand such printed text.

While optical character recognition (OCR) combined with text-to-speech (TTS) techniques can provide an audio readout of printed text, such techniques do not work reliably for printed text on curved surfaces, e.g., nutrition labels on bottles. For curved surfaces, portions of the text can be cut off on the sides making the text unintelligible. The task is especially difficult for low vision or blind users whose hands may inadvertently cover essential portions of the printed label. Barcode scanners can be used to query product information.

Further, barcode or OCR-based techniques typically provide a readout of the printed information or all available information about a product, while a user may have a specific query. For example, a user that is trying to figure out the calorie content of a packaged food item may be provided a readout of the item label and other associated information, with the calorie information tucked therein. Further, such text-based techniques do not work for unlabeled objects or objects that are not labeled with text.

DESCRIPTION

This disclosure describes the use of machine vision techniques to recognize an object implemented along with a conversational bot interface, provided as an audio and/or text-based interface to provide relevant information in response to a user query when the user can easily obtain an image of the object about which they seek information. The described techniques can be implemented as part of a head-mounted device, e.g., a virtual reality/augmented reality (VR/AR) headset, a smartphone, or any other device. Image recognition and context detection can be performed on device, or if the user permits, at a server. The techniques can be implemented as part of a virtual assistant application.

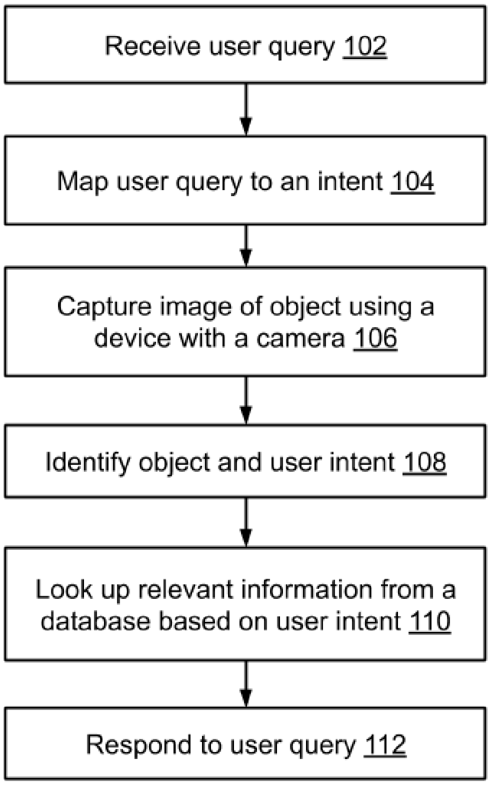
Receive user query 102
Map user query to an intent 104
Capture image of object using a device with a camera 106
Identify object and user intent 108
Look up relevant information from a database based on user intent 110
Respond to user query 112

Fig. 1: Contextual responses based on object recognition and query intent

Fig. 1 illustrates a flowchart of an example method to provide contextual responses based on object recognition and query intent. A user provides a query (102). The user query is mapped to an intent for a conversational bot (304). User intent is identified by use of natural language processing techniques. The conversational bot supports a variety of queries and utilizes a knowledge base to retrieve answers. Along with the query, the user also provides an image (106), e.g., a captured image or a live image of an object. The image can be captured using a built-in camera of a user device such as a smartphone, a smart display, etc. If the user query is received at a device that lacks a camera, e.g., a smart speaker, the image can be captured using a secondary user device that includes a camera.

The image of the object is analyzed, e.g., using a trained machine-learning model, to identify the object (108). Further, if there is text printed on the object, it is recognized, e.g., using optical character recognition. Based on the identified object and the determined query intent, relevant information is obtained from a knowledge base (110). A conversational response is provided to the user (112) with relevant information regarding the detected object. The response can be provided as a spoken response (e.g., by a smart speaker, a smart display, or other device) using text to speech techniques and/or as a text response in a chat conversation between the user and a virtual assistant (e.g., on a smartphone, tablet, or other device with a display screen), etc.

Example of use

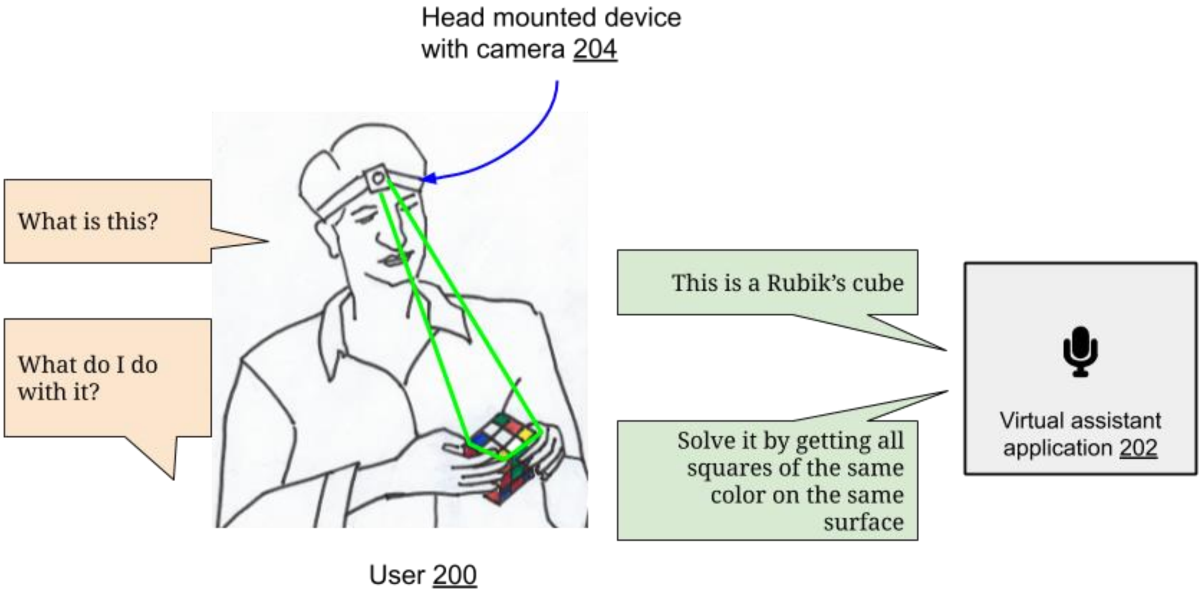
Head mounted device with camera 204
What is this?
What do I do with it?
This is a Rubik's cube
Solve it by getting all squares of the same color on the same surface
Virtual assistant application 202
User 200

Fig. 2: Head mounted device detecting object for user query

Fig. 2 illustrates an example scenario where a user (200) that wears a head-mounted device (HMD) with camera (204) captures an image of an object held in their hand. The user provides a query to a virtual assistant application (202) that is implemented on the HMD, which may be connected to a server. In the example illustrated in Fig. 2, the user is holding a Rubik's cube and asks a question “What is this?” to the virtual assistant.

Using machine vision techniques, the object is recognized as a Rubik’s cube. The virtual assistant therefore responds accordingly with the answer “This is a Rubik’s cube.” Further, the user can ask a follow-up question (“What do I do with it?”) in response to which the virtual assistant provides the guidance “solve it by getting all squares of the same color on the same surface.” Alternatively, the virtual assistant can provide this information along with the object identity, based on the detected intent. In this manner, the user can utilize a conversational interface to obtain relevant information about objects.

While the foregoing example refers to a Rubik’s cube, other types of objects such as toys, electronics, food or medicine packaging, and other types of objects can be recognized, and corresponding information can be provided. For example, a trained machine learning model can be used for object recognition. Object information can be obtained from a knowledge base. The objects can be of any shape, e.g., curved food containers, medicine packages, etc.

The described techniques can be implemented on devices such as augmented reality devices, smartphones, smart displays, etc. The devices are equipped with a camera and either a microphone and speaker or visual display with an input device. The described techniques can also be implemented to allow users to understand how to use an object without prior knowledge of the printed language.

Identification of objects in the image and determination of query intent is performed with specific user permission. Such operations can be performed on the user device, on a server, or a combination, based on user preferences.

CONCLUSION

This disclosure describes techniques to identify specific objects with image analysis techniques to provide relevant information based on user query. Response to queries are provided that are pertinent to user queries about a specific object by combining image detection techniques with natural language processing techniques. The software identifies the object using image detection and natural language processing techniques to parse user intent. Based on the results of image detection and natural language processing, information regarding the specific object is provided by a database to provide a response to user queries. The software architecture provides flexibility to use optical character recognition to read out text captured by a camera using image recognition techniques or provide a description of an object.

REFERENCES

1. Seeing AI, available online at https://www.microsoft.com/en-us/ai/seeing-ai?SilentAuth=1&wa=wsignin1.0
2. Scriptview large print prescription labels, available online at https://www.envisionamerica.com/products/scriptability/scriptview-large-print-labels/
3. Accessible prescription information and medication management, available online at https://www.afb.org/blindness-and-low-vision/using-technology/prescription-health-and-fitness/accessible-prescription