Assistive Mobile Application for the Blind

Ismail Sahak[1], Ong Huey Fang[2*], Syuhada Abdul Rahman[1]
[1]Faculty of Computing, University Malaysia of Computer Science & Engineering, Malaysia
[2]School of Information Technology, Monash University Malaysia, Malaysia
ong.hueyfang@monash.edu

Abstract

One of the challenges faced by blind people is the difficulty in identifying objects with concise information. They could only rely on the senses of hearing, smell, taste or touch to engage and get some perspectives of objects. Hence, this paper presents a mobile application called Iris to aid blind people in "visualising" their surroundings with descriptive objects. Iris combines the multiple object detection and optical character recognition capabilities of Microsoft Computer Vision API to turns smartphones into assistive devices for the blind to use in their daily activities.

1 Introduction

According to a study in 2010, visual impairment is a major global health issue affecting 285 million people in six World Health Organization regions. Approximately 39 million of them are blind, and 246 million with decreased visual acuity (Pascolini & Mariotti, 2012). Another study reported that adults with visual impairment having difficulty performing their daily tasks and need assistance in activities such as reading, writing, shopping, driving and using the computer (Riddering, 2016). Along with that, the use of smartphones is prominent among the visually impaired population. The advent of computer vision technologies, such as object detection and optical character recognition (OCR), is also promising for creating more effective mobile applications to aid blind people dealing with problems of identifying objects, texts, and spatial locations without having to engage them (Ramkishor & Rajesh, 2017).

Among the issues of current mobile applications for blind people is it could only detect a single object at a single frame. In a real-life environment, there is a tendency for objects to be close to each other. Therefore, it is crucial to develop a mobile application that could detect not only one but multiple objects in a single camera frame. Another issue is that most applications (e.g. BlindTool and Aipoly) respond without additional contexts or descriptions of objects with their surrounding environment. For instance, if an apple is detected, the application speaks out the word "apple" to the user. Little does the user know, the apple may be on top of a table. Moreover, printed text is everywhere in our daily life, such as on reports, receipts, bank statements, product packages and medicine bottles. It is troublesome for blind people if they are unable to read these texts.

Therefore, this paper developed a mobile application called Iris for the utility of blind people. The user can tap on the screen to detect descriptive objects. A captured image is sent to Microsoft Computer Vision API as a parameter to retrieve values of detected objects and transcribed text in the image. The mobile application then sorts the retrieved object descriptions and text transcriptions based on confidence to form a sentence. Finally, the application speaks out the sentence to the user in a natural language.

The rest of the paper is organised as follows. First, section 2 gives an overview and discusses some of the related works. Section 3 provides an overview of Iris's design and its main components. Then, the proposed Iris's design is presented in section 4. Subsequently, section 5 concludes this paper.

2 Related Works

In 1996, Malaysia's National Eye survey found that among 18,027 residents examined, the age-adjusted prevalence of blindness and low vision were 0.29% and 2.44% respectively. Females had a higher age-adjusted prevalence of low vision compared to males (Zainal et al., 2002). The authors also highlighted that there is a need to evaluate the accessibility and availability of eye care services and the barriers to eye care utilisation in the country.

With higher computational and storage capacity of mobile devices, as well as growing speeds and coverage of mobile internet, provide unique possibilities for the use of smartphones as universal assistive devices (Punchoojit & Hongwarittorrn, 2017). The promising development in computer vision, such as in optical character recognition (OCR), makes it possible to create assistive devices with camera-based products systems (Dharmale, & Ingole, 2015). Text is widely used in our daily life and an important form

of communication. For example, different signboards with directions and shop names contain important textual or symbolic information to facilitate human's knowledge and perception of the environment and in performing activities, such as for navigation. The need to read textual or symbolic information is essential in the case of blind or visually challenged persons. Having the ability to determine what objects precisely are in front of them, along with any additional information is indeed helpful for the blind (Brady, Morris, Zhong, White, & Bigham, 2013).

This study had benefitted from the use of existing accessibilities technologies. People with visual impairment can easily browse and navigate through their smartphones with accessibility features to use the proposed mobile application. Table 1 shows some of the accessibility features for blind and low vision people, which are available in mobile devices using iOS (Apple, 2018) and Android (Google, 2017) operating systems.

iOS	Android
VoiceOver	TalkBack
Speak Screen	Select to Speak
Captioning & audio descriptions	Audio & on-screen text
Dark mode and smart invert colours	Contrast and colour options
Zoom and font adjustment	Change display and font size
Magnifier	Magnification
Accessibility Shortcuts	Interaction controls
Dictation	Voice dictation
Braille entry & display	BrailleBack

Table 1: Accessibility features on mobile operating systems

Object recognition brings forth a multitude of possibilities in the modern world. This study had also implemented the use of multiple object detection technology. An object detection algorithm typically creates a bounding box around the object of interest to locate it within the image. However, the algorithm might not necessarily draw just one bounding box in an object detection case, there could be many bounding boxes representing different objects of interest within the image, and it would not know how many beforehand (Khurana & Awasthi, 2013).

Faster R-CNN (Region Convolutional Neural Network) was developed by researchers at Microsoft, which is based on R-CNN with a multi-phased approach to object detection. R-CNN used selective search to determine region proposals, pushed these through a classification network and then used a Support Vector Machine (SVM) to classify the different regions (Hulstaert, 2018). However, a selective search is a slow and time-consuming process affecting the performance of the network. Therefore, Faster R-CNN algorithm elimi-

nates the selective search algorithm and lets the network learn the region proposals (Kawazoe et al., 2018). This object recognition technology is provided as an API service by Microsoft, known as the Computer Vision API. One can use Computer Vision in their application by using either a native SDK or invoking the REST API directly (Microsoft Docs, 2020).

Application/Research	BlindTool	AiPoly Vision	Iris
Made for use of blind people	Yes	Yes	Yes
Detection model used	ImageNet	SSD_MobileNet by Cloudsight API	R-CNN by Microsoft Computer Vision API
Number of objects able to be detected	1,000 objects	1,000 objects	Dependent on Microsoft Computer Vision API
CPU use	High	Low	Low
Singe/Multiple Object Recognition	Single	Single	Multiple
Output to the user	Name of an object detected.	Name of an object detected.	Description of multiple objects detected along with text if it is present.
Additional feature	Vibrates based on how confident it is.	Detects colors.	Reads to the user if there are texts available
OCR integration with object recognition	None	None	Yes
Mobile Platform	Android	Android & iOS	iOS

Table 2: Accessibility features on mobile operating systems

Some of the existing mobile applications based on object detection for the blind are such as BlindTool (Cohen, 2015) and Aipoly Vision (Aipoly, 2018). Table 2 presents a comparison of these mobile applications with the proposed. Iris is seen to be better than others in terms of its output to the users. Iris serves to give a better insight of the objects detected by describing them. For instance, if in the captured picture there is a mug on a table, Iris would describe it as "Mug on a table" instead of just saying "Mug". In addition to that functionality, Iris can read any text that is present with an object. If per se there are two drink cans, which could be similar in dimension but different in brand, Iris would come useful because it could tell the user what they

are holding, and what texts on the products using the OCR functionality. Ultimately, the proposed mobile application helps in giving blind people a much better perspective of objects around them and is useful for their day-to-day activities.

3 Requirement Analysis

Requirement analysis is done to understand what will be built, why it should be built, and in what order it should be built. This section explains in detail about the needs of the target users of this application, which are blind people. Interviews were conducted with three voluntary respondents, courtesy of Malaysian Association for the Blind in Kuala Lumpur. Two of them are males, and one is female. They are aged between 24 and 31 years old. The purpose of the interviews was to understand the challenges that blind people face in identifying objects, to get their perception regarding mobile applications, and get their inputs on the development of Iris. The following subsections discussed the results of the interviews.

3.1 Challenges in Identifying Objects

Most of them have problems identifying new objects even by touching them. It means that when a product that they have been buying had changed the packaging, then they would have difficulty to identify the product without someone telling them what it is. Problems also arise when they are unable to touch something, and they could not determine what objects are present in their proximity, or even the environment or location they are situated in. They would need to depend on other senses, such as sound and smell, which could be tricky if it is a new environment. The respondents agreed that they prefer to know what environment they are in without having to touch around.

Besides, the respondents have trouble distinguishing two physically similar objects. They are not able to tell one object from the other when the objects have the same textures or properties when touched. Hence, they need the ability to differentiate objects. For example, they might want to differentiate between a can of tomato soup from a can of evaporated milk. This problem also relates to their inability to read texts if they are not in braille.

3.2 Perceptions of Using Mobile Applications

All of the respondents have their personal smartphones. However, their primary uses of a smartphone merely to text and call. They use assistance features that come with their phones to navigate and interact with the phone's interfaces. The respondents did not use any assistance applications with their smartphones, but they surely welcome any applications that could support their visual needs. One of the respondents shared that they wanted a mobile application that could tell if their hair or cloth is messy. Another respondent shared that he wanted an application that would tell him if there are things in front of him and to tell him the description of cer-

tain items. The responded also added that he wanted an application to manage his medications because it could be confusing sometime.

3.3 Features in Iris

Due to usability reasons, all of the respondents agreed that tapping the screen would be the best way to take a picture. It presumably because it is easier to just tap anywhere on the screen rather than having to locate a button. One key feature suggested by a respondent was an auto-flash feature. This feature proves to be useful to blind people because they would not know if the scene is dark. Hence, having the auto-flash feature helps with the overall usability of the application. Another key feature suggested was repeating the application's instructions. This feature was suggested to allow repetition of instructions for new users and without having to go through the process of retaking the picture. This part of the interview had helped in designing the application to be more user-friendly to blind people.

4 The Proposed Mobile Application

4.1 Architecture

Figure 1 shows the overall system architecture of Iris. It interfaces with Microsoft Cognitive Service and Android OS APIs in order to take advantage of the service that they offer in complement with the application's features. The architecture of the application follows the design of Mobile-View-Controller.
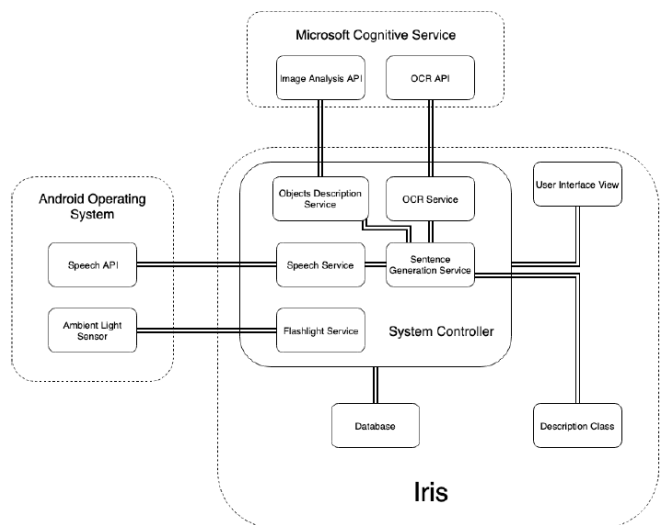

Figure 1: The system architecture of Iris

After uploading an image, Computer Vision API's output tags based on the objects, living beings, and actions identified in the image. Tagging is not limited to the main subject, such as a person in the foreground, but also includes the setting (indoor or outdoor), furniture, tools, plants, animals, accessories, gadgets and others. Computer Vision API's algorithms analyse the content in an image. This analysis

forms the foundation for a "descriptive" displayed in complete sentences. Computer Vision API's algorithms generate various descriptions based on the objects identified in the image. The descriptions are each evaluated and a confidence score generated. An ordered list is then generated from the highest confidence score to the lowest.

OCR technology detects text content in an image and extracts the identified text into a machine-readable character stream (Ramkishor & Rajesh, 2017). This technology can be used for search and numerous other purposes like medical records, security and banking. It automatically detects the language. OCR supports 25 languages, and the accuracy of text recognition depends on the quality of the image. Inaccurate detections may be caused by blurry images, handwritten or cursive text, artistic font styles, small text size, complex backgrounds, shadows, or glare over text or perspective distortion, oversized or missing capital letters at the beginnings of words, subscript, superscript, or strikethrough text.

4.2   User Interfaces

Figure 3 (a) shows the main interface for Iris. The main interface serves as the only interaction screen of the application to make it easy to use for users. It is applied to reduce short-term memory load, especially for blind people. The interface consists mostly of the camera view, followed by the result panel below it; where results are shown in text, along with the spoken output that the application produces. Users have to tap anywhere on the screen to quickly take a picture for it to produce descriptions of objects and texts in the picture. To cancel a request, the user would only have to swipe the screen while the application is loading a request or speaking an output. Another feature is the users can tap and hold anywhere on the screen for the application to repeat a previous description. The control conventions were utilised based on the ease-of-use for blind people, and at the same time reducing the short-term memory load that was mentioned.

Figure 3: (a) Iris's main interface (b) Describing multiple objects

Figure 3 (b) shows a sample output in the case that there are multiple objects detected. It is worth noting that the output

is also spoken to the user. As previously discussed, in this specific case, the API had not been able to return a result in descriptions. Hence, the tags from the result's JSON are accessed and mapped; objects to their characteristics. It makes it possible for the application to output a result even when the Computer Vision API could not construct a description from the picture.

Figure 4 (a) shows an output when there are texts associated with the object detected. It can be seen that the outputs from two of the APIs used are combined within the application to create an output that is intuitive to the user. While Figure 4 (b) shows an exception case when Iris is opened in a poorly-lit environment. The figure depicts that the light source is coming from the smartphone's flashlight. In this case, the application detected that the ambience value in the dark-room is too low. To counter that issue, Iris automatically toggles the smartphone's flashlight on. With this feature, it is more accessible to blind people in the sense that they could not tell if they are in a dark environment while using Iris. Therefore, it is a feature that helps illuminate objects while being in the dark.
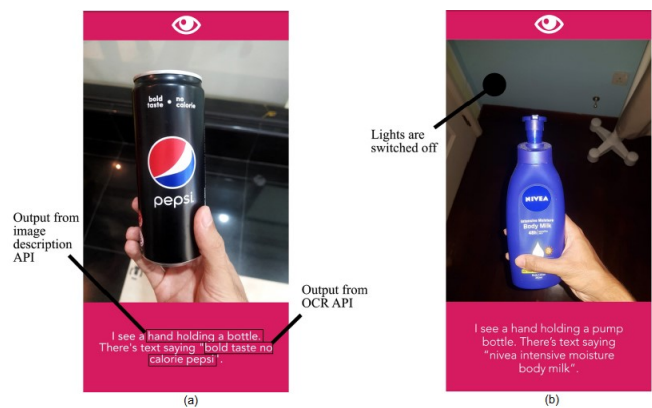

Figure 4: (a) Iris's main interface (b) Describing multiple objects

5   Conclusion

In a nutshell, this paper adopted multiple object recognition and OCR technologies to develop an assistive mobile application for blind people. Iris helps to identify objects and produce descriptive texts from a picture taken by a camera. It serves to describe and distinguish objects so that the blind can have a better insight into the objects in their surroundings. The overall outcome was satisfactory, considering that blind people can make use of the proposed mobile application to tackle problems in their daily activities, hence, aiding them towards independence. However, there are plenty of features that could be implemented to improve the application, such as lower latency and face detection. With 5G technology a major availability in the future, the application could access advanced algorithms and image processing services in the cloud and retrieve the results almost instantaneously. Moreover, the application could be updated to de-

tect and describe human faces, which will be a great support in the communication of blind people with others.

References

Apple (2018). Accessibility on iOS. Retrieved from https://developer.apple.com/accessibility/ios/

Aipoly (2018). Aipoly Vision: Sight for Blind & Visually Impaired on the App Store. Retrieved from https://itunes.apple.com/us/app/aipoly-vision-sightfor-blind-visually-impaired/id1069166437?mt=8

Brady, E., Morris, M. R., Zhong, Y., White, S., & Bigham, J. P. (2013). *Visual challenges in the everyday lives of blind people*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France.

Cohen, J. P. (2015). BlindTool – A mobile app that gives a "sense of vision" to the blind with deep learning, https://github.com/ieee8023/blindtool

Dharmale, R. D., & Ingole, D. P. V. (2015). *Text Detection and Recognition with Speech Output in Mobile Application for Assistance to Visually Challenged Person*.

Google (2017). Android accessibility overview - Android Accessibility Help. Retrieved from https://support.google.com/accessibility/android/answer/6006564

Hulstaert, L. (2018). A Beginner's Guide to Object Detection. Retrieved from https://www.datacamp.com/community/tutorials/object-detection-guide

Kawazoe, Y., Shimamoto, K., Yamaguchi, R., Shintani-Domoto, Y., Uozaki, H., Fukayama, M., & Ohe, K. (2018). Faster R-CNN-based glomerular detection in multistained human whole slide images. *Journal of Imaging, 4*(7).

Khurana, K., & Awasthi, R. (2013). *Techniques for Object Recognition in Images and Multi-Object Detection*.

Microsoft Docs (2020) What is Computer Vision? - Computer Vision - Azure Cognitive Services. (2020). Retrieved from https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/home

Pascolini, D., & Mariotti, S. P. (2012). Global estimates of visual impairment: 2010. *British Journal of Ophthalmology, 96*(5), 614.

Punchoojit, L., & Hongwarittorrn, N. (2017). Usability Studies on Mobile User Interface Design Patterns: A Systematic Literature Review. *Advances in Human-Computer Interaction, 2017*, 6787504.

Ramkishor, V., & Rajesh, L. (2017). Artificial Vision for Blind Peoples using OCR Technology. *International Journal of Emerging Trends & Technology in Computer Science, 6*(3), 30-33.

Riddering, A. T. (2016). *Visual Impairment and Factors Associated with Difficulties with Daily Tasks* (Doctoral dissertation, Western Michigan University). Retrieved from https://scholarworks.wmich.edu/dissertations/2465

Zainal, M., Ismail, S. M., Ropilah, A. R., Elias, H., Arumugam, G., Alias, D., . . . Goh, P. P. (2002). Prevalence of blindness and low vision in Malaysian population: results from the National Eye Survey 1996. *The British journal of ophthalmology, 86*(9), 951-956.