2019 IEEE 4th International Conference on Computer and Communication Systems

Smart Vision: Assistive Device for the Visually Impaired Community Using Online Computer Vision Service

Hasventhran Baskaran[1,a], Rachel Lum Mei Leng[2,a], Fiza Abdul Rahim[3,a,b], Mohd Ezanee Rusli[4,a,b]
[a]College of Computing and Informatics,
Universiti Tenaga Nasional, Kajang, Malaysia.
[b]Institute of Informatics and Computing in Energy (IICE),
Universiti Tenaga Nasional, Kajang, Malaysia
e-mail: hasventhran@gmail.com[1], rachel.lum14@yahoo.com[2], fiza@uniten.edu.my[3], ezanee@uniten.edu.my[4]

***Abstract*—The visually impaired community face several difficulties in their daily life. With the increase of commercial assistive devices, it can greatly improve their lives, especially for navigation and orientation, which are their main obstacles. Recently, it has been introduced many powerful online image processing services, based on machine learning and deep learning. Microsoft is one of the main players in online image processing services. In this paper, we discussed a prototype development by utilizing the online image processing service, Microsoft Cognitive Service. We expect to further evaluate the accuracy and the usage analytics in order to understand long term system behavior.**

Keywords-vision cognitive services; visually impaired; blind assistive

I. INTRODUCTION

Great vision is a valuable blessing however lamentably loss of vision is getting to be noticeably regular nowadays. Blindness may come about because of a sickness, damage or different conditions that limits vision and as a result of which blind individuals stand up to various challenges in their daily life.

Blindness may result from a disease, injury or other conditions that limit vision. World Health Organization reported that there are about 1.3 billion people live with some form of visual impairment and 36 million of them are completely blind. [1]. They are not able to experience the world the way we do. People with complete blindness or low vision often have a difficult time self-navigating outside well-known environments. Indeed, physical development is one of the greatest difficulties for visually impaired individuals, clarified World Access for the Blind [2].

Visually impaired is regularly characterized as a best adjusted visual sharpness of more regrettable than either 20/40 or 20/60. The term visual impairment is utilized for finish or almost total vision misfortune. Visual debilitation may cause individuals challenges with typical every day exercises, for example, driving, perusing, mingling, and strolling.

Usually, blindness influences a man’s capacity to self-explore outside well known environment. It influences a man's capacity to perform many job duties and furthermore exercises outside of a work environment, for example, academics and sports. A large portion of these social difficulties restrain a visually impaired individual's capacity to meet people, and this exclusive adds to low self- esteem.

Voyaging or just strolling down a swarmed road may posture extraordinary trouble for visually impaired and blind people. Along these lines, many individuals with low vision will bring a sighted companion to help explore obscure conditions. They also cannot recognize object without touching and cannot enjoy the beauty of the nature visually.

A number of assistive devices for visually impaired and blind community are commercially available to interpret objects in their environment [3], enhancing the existing cane used by them, and improving their experience in reading using tablet [4].

Moreover, many approaches and systems have been proposed by previous researchers in navigation systems. Researchers developed a prototype to assist visually impaired people to navigate on a sidewalk [5], improving their dining experience [6], and collision avoidance device [7].

Another example, the navigation system using small lightweight drones was proposed to assist user in navigating to a defined address or room [8]. In addition, the drones can utilize computer vision to identify misplaced objects [9]. The system was tested and revealed that visually impaired participants navigate significantly faster compared to the audio navigation [10].

Karkar et. al summarized enormous prospect of assistive technology for mobile devices which can be presented in different forms, including general assistive mobile applications and advanced sensory based applications [11]. Based on their survey, there is a need to develop a robust system that is capable to manage the diversified services and provides support for visually impaired people.

In this paper, we investigate the potential of computer vision services to allow visually impaired and blind users aware what is happening around them. A prototype called Smart Vision is developed to tackle this problem. The prototype aims to bring the beautiful world as a narrative to the visually impaired and blind users.

The narrative is generated by converting the scenes in front of them to text which describes the important objects in the scene. Examples of text include “a group of kids playing”, “a man posing for a picture”, “a bird sitting on tree branch”. One line along with some keywords are narrated as an audio to the users.

978-1-7281-1322-7/19/$31.00 ©2019 IEEE

II. Background

Technology for the disabled has made various advances over the current years. The things they have been doing to help disabled individuals to have the capacity to remain with the advances that have been going ahead on the planet is continuing pleasantly. This helps the disabled to do essential things that we have thought to be an effortlessly routine chore.

However, most of these solutions come with a high price that the visually impaired people cannot afford. In Malaysia, we can see their difficulties in earning money by singing on the streets, selling tissues or making some handcrafts. By imposing hefty prices on new assistive device for visually impaired, it is not going to reach the targeted user nor fulfilling the functions of the device.

Utilizing online computer vision service is one of our initiatives to reduce the development cost for this prototype. There are several online computer vision service that has been developed to analyse images, providing image interpretation, and text descriptions of the images. These services are based on machine learning and deep learning systems.

From a comparison test conducted, Microsoft Cognitive Service perform better for object recognition and human face recognition [12]. It also provides a full description of a complete scene on an image. Therefore, we adopted Computer Vision API, which is one of the cognitive services API provided by Microsoft Cognitive Service in the prototype.

III. Methodology and Prototype Components

We conducted preliminary investigation through interview to understand the limitations faced by visually impaired people and to identify the relevant components for prototype development.

A. *Interview Interpretation*

We interviewed three visually impaired respondents, to understand their needs to overcome visual limitations. Our interviews gave us insights of their needs and allowed us to develop the prototype.

One of the respondents stated that his biggest concern that interacting with others when they do not start the conversation. In getting their insights to the needs of wearable assistive device, it should not draw attention due to its appearance and should not add affliction and burden to their current restrictions.

B. *Prototype Components*

Smart Vision is specially invented for the visually impaired. This device will take the picture of the scenery in front of the user and narrate it back to the user through a headset. To perform this function, this device requires a Raspberry Pi, Pi Camera, Lithium Powerpack, and a headset. Raspberry Pi will process the image that is captured by the PI Camera. The processed description of the image will be played as a voice narration to user.

1) Python: To build web app and to automate small tasks on systems.

2) Microsoft Cognitive Service: The API used for this project is the cloud-based Computer Vision API, which gives developers access to cutting edge algorithms for handling pictures and returning information. By uploading an image, the system returns the information by reading out the visual contents found in that image to the user.

3) Google Text to Speech (gTTS): It is a Python interface to Google's Text to Speech API. Create an mp3 file with the gTTS module or gtts-cli command line utility. It allows for unlimited lengths of spoken text by tokenizing long sentences where the speech would naturally pause.

4) Raspberry PI 3 Model B: This small single-board computers is appended with Pi Camera, which have 5 Megapixels and can record video with a high definition quality of 1080P. A button embedded on a breadboard is connected to the Raspberry Pi.

IV. Prototype Development

A. *Algorithm Development*

As shown in Figure 1, when the system is running, the user has to press the button to take a picture. The captured picture is sent to Microsoft Cognitive Service to be analysed. If the picture is clear, the Microsoft Cognitive Service will process the description of the picture and send back to Raspberry Pi. Then, the description of the picture is narrated to the user through a headset. If the picture is blur, the system will notify "A blurry image was taken, please take a picture again".

```
while (Power On)
   {
   start program;
   snap picture;
   picture analysis;

   if(clear picture?) {
     narrate description;

    if(!user repeats)
     {
       end program;
       GPIO cleanup;
     }
   }
   else
   {
     narrate blurry image;
   }
   restart program;
}
```

Figure 1. Algorithm development.

B. *Circuit Diagram*

Figure 2 shows the circuit diagram of the system. The camera module is connected to the camera serial interface (CSI) port on the Raspberry Pi 3 board, and takes in the digital output form the camera module. The button switch is connected to the GPIO pins, GND and 15. When the button switch is pressed, a Python script is executed which activates the camera to take picture and send it to Microsoft Cognitive Service.

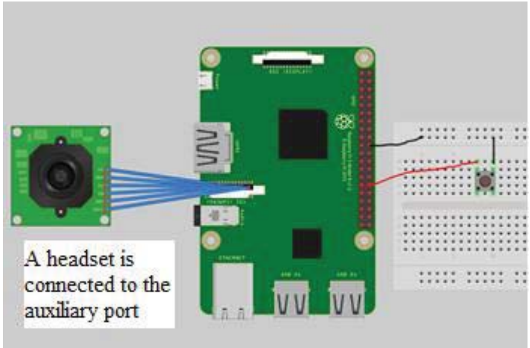

Figure 2. Circuit design of the smart vision.

To enable the Raspberry Pi to "talk", gTTS will be embedded in the Raspbian system. The module is then imported to Python and commanded to read the image classifications result.

V. IMPLEMENTATION AND TESTING

A. Picture Analysis

In order to run Computer Vision API's service, the API key and the Computer Vision API's URL is embedded in the code. The parameters to be descripted in JSON file is also defined. Once the coding is debugged, the Raspberry Pi sends the captured image from camera to Microsoft Cognitive's knowledge base. The online API run its' service and classifies the picture. The description is generated in JSON format and sent back to the Raspberry Pi.

During the testing for image classification, the first trial was unsuccessful due to the failure of extraction of image from the local device. However, the second attempt was successful after the image's path is defined and the image was translated as a binary information to the Cognitive Service.

B. Setting GPIO Pins and Pi Camera

To detect the camera and to enable it to capture picture, Python codes are used to connect the button on pin numbers and set the image name, URL, height and width as shown in Figure 3.

```
GPIO.setmode(GPIO.BCM)
BUTTON = 15
SYSTEM_READY = 26
SYSTEM_RUNNING = 13

GPIO.setup(SYSTEM_READY, GPIO.OUT)
GPIO.setup(SYSTEM_RUNNING, GPIO.OUT)
GPIO.setup(BUTTON, GPIO.IN, pull_up_down=GPIO.PUD_UP)

IMG_WIDTH = "1920"
IMG_HEIGHT = "1080"
IMG_NAME = "/home/pi/image.jpg"
```

Figure 3. Lines of Code to Connect GPIO Pins and Button

The Pi camera module is imported into Python codes to activate the Pi camera to take picture after being triggered when the button is pressed. Figure 4 shows lines of code that triggers the camera to capture image.

```
print "System Ready - push button to take picture.\n"
try:
    GPIO.wait_for_edge(BUTTON, GPIO.RISING)
    print "Program running...\n"
    GPIO.output(SYSTEM_READY, False)
    GPIO.output(SYSTEM_RUNNING, True)

    print "Capturing photo...\n"
    ret = subprocess.call(snapCommand, shell=True)
    photo = open(IMG_NAME, 'rb')
finally:
    GPIO.cleanup()
Print 'Picture Captured"
```

Figure 4. Lines of code for camera activation.

C. Voice Narration

To enable the Raspberry PI to "talk", Google Text to Speech (gTTS) is embedded in the Raspbian system. The module is then imported in Python and commanded to read the image classifications result as shown in Figure 5. During the testing for voice narration, the first trial was unsuccessful due to the version of Python (Python 3) used. However, the second attempt was successful after reverting the Python to an earlier version (Python 2.7).

```
data = response.json()['description']['captions'][0]
  ['text']
tts = gTTS(data, lang = 'en')
tts.save("result.mp3")
os.system("mp321 result.mp3")

mixer.init()
mixer.music.load('result.mp3')
mixer.music.play()
```

Figure 5. Line of codes for voice narration.

D. Looping the Program

It is expected that the user will not need someone to run the program each time it finishes executing. The prototype is tested to enable the program to loop on its own.

During the testing of system looping, the first attempt failed due to the use of 'While' loop. The system was not able to loop with the While statement. However, the second trial was successful after using 'For' loop with a number of count (10000) as shown in Figure 6.

```
import os
import voice

for i in range(10000):
  execfile("new.py")
```

Figure 6. Line of codes for looping.

Figure 7 shows our prototype is embedded into cap for testing purpose. It is in our plan to improve the prototype by attaching aesthetic value for user's suitability. More considerate designs are expected to be designed to interact with others in a safe environment

This is in line with one of the requirements that we gathered from the visually impaired people, although they cannot see like an ordinary people, but still they want to look good with all the accessories attached with them.

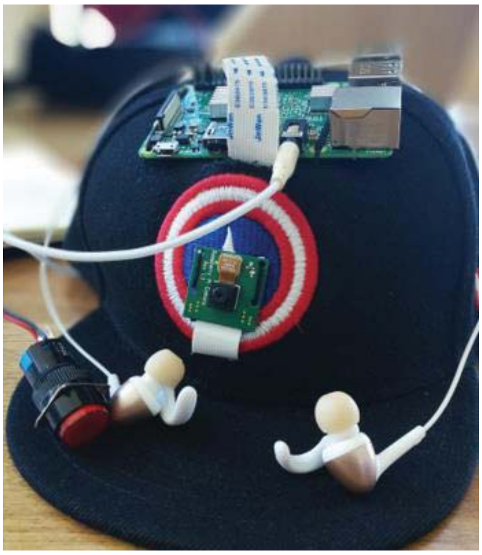

Figure 7. Final prototype of smart vision.

VI. Future Works

It is our intention to further develop the proposed prototype by focusing on the accuracy and the usage analytics in order to understand long term system behavior. Later, we will further evaluate the usability and functionality with visually impaired people in a laboratory environment. The prototype is expected to be redesigned by using a pair of sunglasses with integrated cameras that capture images. Analyzed information will be narrated and translated to users.

In our effort to reach more potential users in our local community, the narration with the aid of Google Translate API will be considered to accommodate with Malay, Mandarin and Tamil languages. It is also in our planning to include facial expression feature that will allow users to improved their social interaction by knowing the expression of other peoples they are interacting with.

To produce a more rounded device that can replicate the eyes' functions, we are developing navigational ability for this device. This feature will enable the target user to avoid obstacles at head level. These people come across a lot of head level obstacles like street signs when they are strolling through city streets.

In our survey, visually impaired people admit that they would need a device that helps them to avoid these obstacles. This problem can be rectified by embedding ultrasonic sensors on the device. These sensors will be programmed to detect objects from a preset distance. Whenever it detects an object or obstacle, it will alert the user through a buzzer or wearable vibration modules. Along with this feature, target user also will be able to send their current location to their family during emergency situations, with a push of a button.

The main constraint in this prototype is the need for Internet connectivity. Since it is a project that dwells with Internet of Things (IoT), the Internet connection should be strong. Without Internet connection, the device will not be able to function. Currently, the device is connected to a mobile hotspot to be able to function. It is planned to embed a GSM module with the Raspberry Pi to have a sim card on board to always provide Internet connectivity. This will remove the dependence of Smart Vision on an external device to connect to the Internet.

VII. Conclusion

We intend to provide an affordable and convenient blind aid device by breaking the trend of making commercially-driven prototypes and make devices for those who are in need. By packing more features in this prototype in future, it will provide a complete blind aid device which will enable them to do many things on their own.

In future, Smart Vision is expected to reach visually impaired people from multiple layers of society. Whether they are rich or poor, they can afford to have a smart companion that helps them to be more independent. This will put technology into good use and aid in the betterment of the blind society.

It is our hope that the outcomes of this initiative will provide a reliable solution for visually impaired people to ease and improve their daily life. In the future, it will contribute to the betterment of the visually impaired society and encourage efficient use of technology in helping disabled people.

Acknowledgment

We would like to express profound gratitude to Research & Development Unit, Malaysian Association for the Blind for their useful suggestion for this project. This project is funded under Yayasan Canselor Uniten (YCU) Community Grant.

References

[1] World Health Organization, "Global Data on Vidual Impairments 2010," 2012.

[2] World Access for the Blind, "Blindness: Challenge and Achievement," 2015. [Online]. Available: https://waftb.net/blindness-challenge-and-achievement. [Accessed: 20-Sep-2018].

[3] BrainPort Technologies, "BrainPort Vision Pro," 2018. [Online]. Available: https://www.wicab.com/brainport-vision-pro. [Accessed: 20-Sep-2018].

[4] Zach Pontz, "New devices help the blind 'see' differently," *From The Grapevine*, 2016. [Online]. Available: https://www.fromthegrapevine.com/innovation/eyecane-eyemusic-help-blind. [Accessed: 10-Sep-2018].

[5] F. Ahmed, S. Mahmud, R. Al-fahad, S. Alam, and M. Yeasin, "Image Captioning for Ambient Awareness on a Sidewalk," *2018 1st Int. Conf. Data Intell. Secur.*, pp. 85–91, 2018.

[6] C. Lee, Y. Chu, L. Cheng, Y. Lin, and K. Lan, "Blind Assistive Device - Smart Lazy Susan," in *Proceedings of the 2017 International Conference on Machince Learning and Cybernetics*, 2017.

[7] A. Ali and M. A. Ali, "Blind Navigation System for Visually Impaired Using Windowing-Based Mean on Microsoft Kinect

Camera," in *2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME)*, 2017.

[8] M. Avila, M. Funk, and N. Henze, "DroneNavigator: Using Drones for Navigating Visually Impaired Persons," *Proc. 17th Int. ACM SIGACCESS Conf. Comput. & Access.*, pp. 327–328, 2015.

[9] C. Yi, R. W. Flores, R. Chincha, and Y. Tian, "Finding objects for assisting blind people," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, 2013.

[10] M. Avila Soto, M. Funk, M. Hoppe, R. Boldt, K. Wolf, and N. Henze, "DroneNavigator: Using Leashed and Free-Floating Quadcopters to Navigate Visually Impaired Travelers," *Proc. 19th Int. ACM SIGACCESS Conf. Comput. Access. - ASSETS '17*, pp. 300–304, 2017.

[11] A. Karkar and S. Al-maadeed, "Mobile Assistive Technologies For Visual Impaired Users : A Survey," *2018 Int. Conf. Comput. Appl.*, pp. 427–433.

[12] A. Reis, D. Paulino, V. Filipe, and J. Barroso, "Using Online Artificial Vision Services to Assist the Blind - an Assessment of Microsoft Cognitive Services and Google Cloud Vision," in *WorldCIST'18 2018: Trends and Advances in Information Systems and Technologies*, 2018, pp. 174–184.