This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Ghosh, Amit; Al Mahmud, Shamsul Arefeen; Uday, Thajid Ibna Rouf; Farid, Dewan Md.

Assistive Technology for Visually Impaired using Tensor Flow Object Detection in Raspberry Pi and Coral USB Accelerator

Published in:

Proceedings of the 2020 IEEE Region 10 Symposium, TENSYP 2020

DOI:

[10.1109/TENSYP50017.2020.9230630](https://doi.org/10.1109/TENSYP50017.2020.9230630)

Published: 05/06/2020

Document Version

Peer reviewed version

Please cite the original version:

Ghosh, A., Al Mahmud, S. A., Uday, T. I. R., & Farid, D. M. (2020). Assistive Technology for Visually Impaired using Tensor Flow Object Detection in Raspberry Pi and Coral USB Accelerator. In *Proceedings of the 2020 IEEE Region 10 Symposium, TENSYP 2020* (pp. 186-189). [9230630] IEEE.
<https://doi.org/10.1109/TENSYP50017.2020.9230630>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

© 2020 IEEE. This is the author's version of an article that has been published by IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Assistive Technology for Visually Impaired using Tensor Flow Object Detection in Raspberry Pi and Coral USB Accelerator

Amit Ghosh[§], Shamsul Arefeen Al Mahmud*, Thajid Ibna Rouf Uday[†], Dewan Md. Farid[‡]

^{§†}Department of Computer Science & Engineering, [†]Department of Electrical & Electronic Engineering, United International University, United City, Madani Avenue, Badda, Dhaka 1212, Bangladesh

*Department of Electrical Engineering & Automation, Aalto University, Finland.

Emails: [§]amitbd1508@gmail.com, *shamsul.almahmud@aalto.fi, [†]tuday121102@bsee.uui.ac.bd, [‡]dewanfarid@cse.uui.ac.bd

Abstract—Assistive Technology (AT) becomes an interesting field of research in the present era. According to the World Health Organisation (WHO - <https://www.who.int>), there are approximately 285 million visually impaired people around the world. To address this issue, many researchers are employing new technologies, e.g. Machine Learning (ML), Computer Vision (CV), Image Processing, etc. This paper aims to develop an assistive technology based on Computer Vision, Machine Learning and Tensor Flow to support visually impaired people. The proposed system will allow the users to navigate independently using real-time object detection and identification. Hardware implementation is done to test the performance of the system, and the performance is tracked using a monitoring server. The system is developed on Raspberry pi 4 and a dedicated server with NVIDIA Titan X graphics where Google coral USB accelerator is used to boost processing power.

Keywords—Assistive Technology; Computer Vision; Object Detection; Tensor Flow; Visually Impaired;

I. INTRODUCTION

Assistive Technology (AT) is used for the betterment of persons with some disability, e.g., blindness and deafness. Visually impaired people suffer from numerous difficulties, such as working on their own, not understanding the changes in their surroundings. In recent years, researches are amalgamating emerging technologies like Artificial Intelligence (AI), Deep Learning to assist the visually impaired people. Physical and social barriers, along with accessibility, make it burdensome for a visually impaired person. Blindness and visual impairment are one of the highest rates in Asia that round 63% of people have a visual impairment [1]. Several assistive devices e.g. earphones, smart cane, goggles, etc have been introduced to aid blind people in last decades. Assistive technologies can enhance the quality of life of a blind person by introducing specific devices, services, systems, processes, and environmental modifications [2]. Recent studies observe the impact of state-of-art assistive technologies such as cane, goggles, wearables, etc. for the visually impaired person [3]. Auditory vision can be a solution for the visually impaired person as many studies illustrate that blind persons do have better sound sense than a sighted person.

Recently, researchers are looking for a way to integrate different technologies with voice commands to assist a blind person in navigating and understanding his/her environment.

Smartphones are the most common device in assistive technology development. Voice assistant using a smartphone is the most common option for the visually impaired person. Such technologies have limitations as smartphones need to be focused on a specific thing to know about it. Voice assistant applications are not perfect and cannot answer all the necessary queries. There are many applications of assistive technology for visually impaired people, such as navigation in the road, navigation assistance in a vehicle, etc. In intelligent transport system (ITS), assistive technologies can be merged for visually impaired people to help them navigate through the road [4], [5]. Ghosh et. al. [6] presented a video based system to count the number of vehicles, measure speed where the assistive system for a blind person can be easily implemented. Babir et. al. [7] discussed the advantages of radio over fiber-based vehicle communication systems where real-time data can be easily transferred through the internet, which can be a useful factor in assistive technology to collect real-time data from the environment.

Computer Vision (CV) is gaining popularity for processing digital images nowadays, which is a branch of Artificial Intelligence (AI) and Machine Learning (ML). In this paper, a wearable Computer Vision system based on Internet of Things (IoT) embedded platform is proposed where image classification is implemented by integrating with the voice to assist the visually impaired person. The system is developed and tested in Dhaka, Bangladesh. Visual data are collected through a camera and then processed through Tensor Flow, an open-source framework developed by Google. Image classification, object detection, speech recognition, and voice command for navigation are the main strength of this work.

The remainder of this paper is organised as follows. Section II discusses about the related works to understand the recent development in this area. Section III presents the system overview along with data processing technique. Section IV presents the experimental test and results. Finally, Section V presents the conclusion with future works.

II. RELATED WORK

Computer Vision (CV) is creating an innovative step towards assisting visually impaired people. Leo et. al. [8] discussed the use of Computer vision (CV) that can help

to develop multiple assistive technologies for blind or visually disabled people. Software solutions are becoming more common in assistive technology. Tapu et. al. [9] presented the difficulties in pointing smartphones towards an object to describe the environment. Digital image processing is stepping forward in the development of assistive systems for visually disabled people. Numerous works have been done using image processing or computer vision technology. Roberto and Dan developed a laser-based virtual cane on top of computer vision, which can assist in navigation [10]. Rajalakshmi et. al. [11] discussed about an assistive system using object detection by the tensor flow and ultrasonic sensors. This work is similar to [11], however, in the proposed method, computer vision-based goggles are implemented without any sensor, which is reducing the cost. Moreover, we have considered the hearing problem of the person and integrated a bone conduction earphone with the system. The whole system is based on raspberry pi. Raspberry pi is a well known embedded platform for development projects. Goel et. al. [12] used raspberry pi to develop readers for blind people. The voice vision technology developed a visual experience system that can provide an image to sound description to the users [13]. It uses a camera on top of goggles to capture images and map them according to the objects in the image. Navigation is a hurdle for the visually impaired person. Ross [14] developed a system for blind people navigation, and the system was tested to cross a road using an assistive wearable. This work taken to imply CV by integrating voice command, speech recognition and object detection. The uniqueness of this work is including Tensor Flow for image processing, object detection, expression recognition, real-time ID, text to speech conversion and navigation system. The high processing unit is used to process the images with proper efficiency.

III. METHODOLOGY

A. System Overview

The proposed system is a wearable goggle with earphones based on IoT embedded platform using raspberry Pi 4. It has multiple segments includes image collection, image processing, object detection, expression recognition, real-time ID, text to speech conversion, and navigation system. The camera is an essential device in this system, which is responsible for collecting images of the real-time environment. The entire system is based on goggles, which include a camera and earphones. The whole system works depending on the image collected by the camera. Then the collected images are processed through Tensor Flow to identify different objects in the images. In Fig. 2, there are four modes of operation after getting the image data. In describe mode, the system uses local or cloud processing to explain the user about the image. In announcement mode, the system will narrate the object in the images in real-time. In search operation mode, object search and find direction are done by using voice recognition. In the proposed system, to identify the object's exact position and distance from the user, the system calculates the angle from the user's current location. In the last mode, real-time identification and expression recognition are made by the system. It has an algorithm to detect new faces and saved it in the database for future reference. After all of these steps, the system generates text and process it for text to speech conversion. Then the speech output is sent to the user through a bone conduction

earphone. A prototype is built to test the system and evaluate the performance. Table I shows the prototype system in details.

Fig. 1: Image of the prototype system, which contains a camera, goggles, USB coral, raspberry pi 4 and headphone.

TABLE I: Prototype system details.

Specifications	Description
Camera	Fixed focus, 720p/30fps
Goggles	Ordinary glasses
Processing Unit	Google Coral
Embedded Platform	Raspberry Pi 4

Fig. 2: Flow chart of the proposed system with different modes.

B. Data Collection and Processing

Images of the local environment are used as data for the system in this work. As mentioned earlier, the visual impairment rate in Asia is one of the highest in the world. Based on this, for testing the system, images were taken in Dhaka, Bangladesh. Camera video was taken with a speed of 30 frames per second. After taking the video, images are extracted from the video and then converted into a matrix shape to analyse each pixel. After that, images are levelled to identify objects in the picture. The training pipeline is created by developing a workspace where all the images are annotated. From these annotated images, label mapping is done to detect objects in the images. Label mappings are converted in a tensor flow file format to process them in the system. After developing the training model, the interface graph is exported. Then these graphs are converted into tensor flow lite and fed them into google coral accelerator.

TABLE II: Data set details.

Specifications	Description
Number of Images	5,000 samples
Number of labeled objects	16,000 objects
Train data	80 percent
Test data	20 percent

MobileNet is a network designed specifically for the deployment in mobile and embedded vision applications. The idea behind MobileNet was to build a small network with low latency for platforms with limited computational resources. To train the data set, MobileNet open-source data is used along with the locally processed images to increase the data range. The model is built and trained using NVIDIA Titan due to the availability of a GPU for faster training. The model is created using Python 3, and Tensor Flow for building the classifier and OpenCV for image processing. The evaluation of each model, especially with respect to the suitability for deployment on an embedded platform, was performed on a Raspberry Pi 4. There is one special reason for using Raspberry pi 4 as it has a dual-band WiFi facility, which helped the system to store real-time images from time to time. These real images are assisting the system in adopting real-life situations more accurately. From ethical perspective, user consent is taken before performing any data collection. European union data policy was used to take consent from users and people in the images.

C. Hardware

The system is developed on raspberry Pi 4 platform. Table I shows the specifications of Raspberry Pi 4. A googles is used with an integrated earphone. A camera is placed on top of goggles to capture video and images will be process on the IoT embedded platform based on Raspberry Pi.

TABLE III: Specifications of Raspberry Pi 4.

Specifications	Description
Processor	64-bit ARM Cortex-A72 CPU
Memory	Up to 4GB
Connectivity	Gigabit Ethernet; Dual-band Wi-Fi

As image processing needs a high processing unit, for that purpose, Google coral USB accelerator is used to accelerate

Fig. 3: Sample of the dataset where in (a) a normal image of the bottle is shown; where in (b) the system detected the object as bottle with 65.64 percent accuracy.

the processing operation. Tensor Flow model is converted to a tensor flow light model as raspberry pi alone cannot process a large amount of data. To increase the processing power tensor processing unit (TPU) as it can accelerate the system efficiency. NVIDIA Titan X is used as a GPU to train the data set faster. A dedicated server with 32Gb ram along with 4Gb NVIDIA Titan X graphics is used for the faster processing of the system. This server is used to accelerate image processing and object detection. A bone conduction headphone is used to convey the voice command to the user. According to medical practitioner visually impaired person usually have some issue with their hearing as well. A bone conduction headphone is designed for this purpose.

TABLE IV: Specifications of the dedicated server.

Specifications	Description
Processor	Intel core-i7
Memory	32GB
Graphics	NVIDIA Titan X 4Gb

D. Software

The scripts for this system are written in Python. The tensor processing unit is controlled by a Python script. When the data set is large to process, then TPU will be enabled using a script. The data analysis is done through the Python script. As the system will contain multiple modes including, describe mode, identification mode, expression recognition mode, real-time location announcement, and object search. These modes are controlled and enabled by different scripts. CUDA programming is used to run the training set in the GPU. Google voice default api is used for the voice command in Bengali.

IV. EXPERIMENTAL STUDY, RESULT AND ANALYSIS

An experimental study has been done to validate the proposed model. We have consulted with two visually impaired persons who agreed to perform the experiment. We have asked them to use our system for a week. The experiment was tracked to understand the performance improvement by

the users. The experiment was conducted in different places including, university classroom, university cafeteria, university lobby, office room and an office meeting room. The experiment environments are selected based on room size. The location was United International University, Bangladesh, and office location is ANTT Robotics Limited, Bangladesh. From the experiment, we have collected accuracy data to understand how conveniently our users can identify and roam around the places independently. As mentioned above, as experiment was conducted in different places by the help of two visually impaired persons. From the experiments we have collected image frames from the camera, and tried to calculate the accuracy of the overall system. Table V is containing the data from our experiment. The accuracy is representing the precision of identifying objects in the environment. This accuracy is the determinant about how precisely a visually impaired person can recognise objects within the coverage area of the system. Data column is the number of frame collected during the experiment. The system was tested in different room environment where classroom and office room were quite similar in size. University cafeteria and office meeting room were bit larger than classroom size, where university lobby was the largest environment. From Table V, the system performance is better when the room size is small and it degrades when the size increases. System accuracy is 89% and 87% for classroom and office room respectively, where in university lobby the accuracy is lowest (70%).

TABLE V: Result on system accuracy.

Location	Data	Minute	Accuracy
University Classroom	460	80	89 %
University Cafeteria	1090	10	85 %
University Lobby	300	12	70 %
Office Room	633	15	87 %
Office Meeting Room	522	7	81 %

V. CONCLUSIONS AND FUTURE WORK

The proposed system has achieved acceptable accuracy level for standard size room environment. The accuracy is near 90% in classroom and office room. A bone conduction earphone is considered in the system, as visually impaired people may also have some problem in their hearing. This system can be developed more efficiently by selecting proper high functioning camera, more processing power for labelling and image processing. As future work, we would like to consider wide angle camera to cover more objects in single frame and connect the system with internet more efficiently to collect images in real time with powerful processing unit.

ACKNOWLEDGMENT

We appreciate the support received from the a2i Innovation Fund of Innov-A-Thon 2018 (Ideabank ID No.: 12502) from

a2i-Access to Information Program – II, Information and Communication Technology (ICT) Division (<https://ictd.gov.bd>), Government of the People’s Republic of Bangladesh. We would like to thank “a2i (Access to Information) Innovation Lab” (<https://a2i.gov.bd/innovation-lab/>).

REFERENCES

- [1] Y.-C. Tham, S.-H. Lim, Y. Shi, M.-L. Chee, Y. F. Zheng, J. Chua, S.-M. Saw, P. Foster, T. Aung, T. Y. Wong *et al.*, “Trends of visual impairment and blindness in the singapore chinese population over a decade,” *Scientific reports*, vol. 8, no. 1, pp. 1–7, 2018.
- [2] M. A. Hersch and M. A. Johnson, “On modelling assistive technology systems—part i: Modelling framework,” *Technology and disability*, vol. 20, no. 3, pp. 193–215, 2008.
- [3] A. Bhowmick and S. M. Hazarika, “An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends,” *Journal on Multimodal User Interfaces*, vol. 11, no. 2, pp. 149–172, 2017.
- [4] I. Alam, D. M. Farid, and R. J. F. Rossetti, “The prediction of traffic flow with regression analysis,” in *International Conference on Emerging Technology in Data Mining and Information Security (IEMIS)*, Kolkata, India, February 2018, pp. 1–10.
- [5] I. Alam, M. F. Ahmed, M. Alam, J. Ulisses, D. M. Farid, S. Shatabda, and R. J. F. Rossetti, “Pattern mining from historical traffic big data,” in *IEEE Technologies for Smart Cities (TENSYP)*, and *IEEE Xplore Digital Archive*, Cochin, Kerala, India, July 2017, pp. 1–5.
- [6] A. Ghosh, M. S. Sabuj, H. H. Sonet, S. Shatabda, and D. M. Farid, “An adaptive video-based vehicle detection, classification, counting, and speed-measurement system for real-time traffic data collection,” in *The IEEE Region 10 Symposium (TENSYP) Symposium Theme: Technological Innovation for Humanity*, Kolkata, India, June 2019, pp. 1–6.
- [7] M. R. N. Babir, S. A. Al Mahmud, and T. Mostary, “Efficient m-qam digital radio over fiber system for vehicular ad-hoc network,” in *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2019, pp. 34–38.
- [8] M. Leo, G. Medioni, M. Trivedi, T. Kanade, and G. M. Farinella, “Computer vision for assistive technologies,” *Computer Vision and Image Understanding*, vol. 154, pp. 1–15, 2017.
- [9] R. Tapu, B. Mocanu, A. Bursuc, and T. Zaharia, “A smartphone-based obstacle detection and classification system for assisting visually impaired people,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 444–451.
- [10] S. Sivan and G. Darsan, “Computer vision based assistive technology for blind and visually impaired people,” in *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies*, 2016, pp. 1–8.
- [11] M. R. Rajalakshmi, M. K. Vishnupriya, M. M. Sathyapriya, and M. G. Vishvaardhini, “Smart navigation system for the visually impaired using tensorflow,”
- [12] A. Goel, A. Sehrawat, A. Patil, P. Chougule, and S. Khatavkar, “Raspberry pi based reader for blind people,” *International Research Journal of Engineering and Technology*, vol. 5, no. 6, pp. 1639–1642, 2018.
- [13] M. Auvray, S. Hanneton, and J. K. O’Regan, “Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with ‘the voice’,” *Perception*, vol. 36, no. 3, pp. 416–430, 2007.
- [14] D. A. Ross, “Implementing assistive technology on wearable computers,” *IEEE Intelligent systems*, vol. 16, no. 3, pp. 47–53, 2001.