Blind's Eye : IoT based real-time surrounding identification and object detection

ABSTRACT

By

HIMANSHU SHEKHAR
ENROLLMENT NO. : 12017002002067
SUJOY SEAL
ENROLLMENT NO. : 12017002002010

Under the supervision of
PROF. TUFAN SAHA

Department of Computer Science and Engineering
Institute of Engineering and Management

West Bengal, India
September, 2020

Abstract

This project aims to provide software e-solution for visually impaired people. We want to address the general problem that little has been done for providing real time visual aid to the blind people in terms of low cost and high usability facilities.

According to World Health Organisation (WHO), more than a quarter of the world's population approximately 2.2 billion people, are visually impaired.[1] Around 1.8 billion people are suffering from Presbyopia (a condition where it is difficult to see nearby objects) and 2.6 billion people from Myopia, (a condition where it is difficult to see distant objects). 65.2 million from Cataract and rest by many other diseases. In 2010, WHO reported that approximately 39 million people are completely blind globally.[2]

Some common causes for blindness are cataract, glaucoma, age-related macular degeneration, corneal opacification, childhood blindness, trachoma, etc.

Difficulties faced by blinds in their day-to-day life:

A. Navigating around places: The biggest challenge for any visually impaired person, especially with those with complete loss of vision is to navigate around places.

B. Reading materials: Blind people have a tough time finding good reading materials in accessible formats. Internet, the treasure trove of information and reading materials, is mostly inaccessible for the blind people. Even a blind person can use screen reading softwares but it does not make the surfing experience very smooth if the sites are not designed accordingly. Blind person depends on the image description for understanding whatever is represented through pictures. But most of the time, websites do not provide clear image description.

C. Overly helpful individuals: It is good to be kind and help others. But overly helpful individuals often create problems for the blind person. There are lots of individuals who get so excited to help a disabled person that they forget even to ask the person whether he needs help or not.

D. Recognizing their own family members: Blind people sometimes aren't able to recognize their own family members and think of them as a stranger. This also creates a problem for a visually impaired person in this life. Situation may arise like a blind person can not take help from a person sitting beside (for an etiquette) thinking of him as a stranger while he is his own family member.

And there are many more difficulties which visually impaired people have to face but here we are trying to solve major problems as mentioned above. Later in this chapter we will discuss briefly how we tried to solve the major problems visually impaired people are facing.

A desire visually impaired people having no vision, have, is to navigate and recognize the environment. They want to know what's going on in their surroundings, therefore, we came up with this solution.

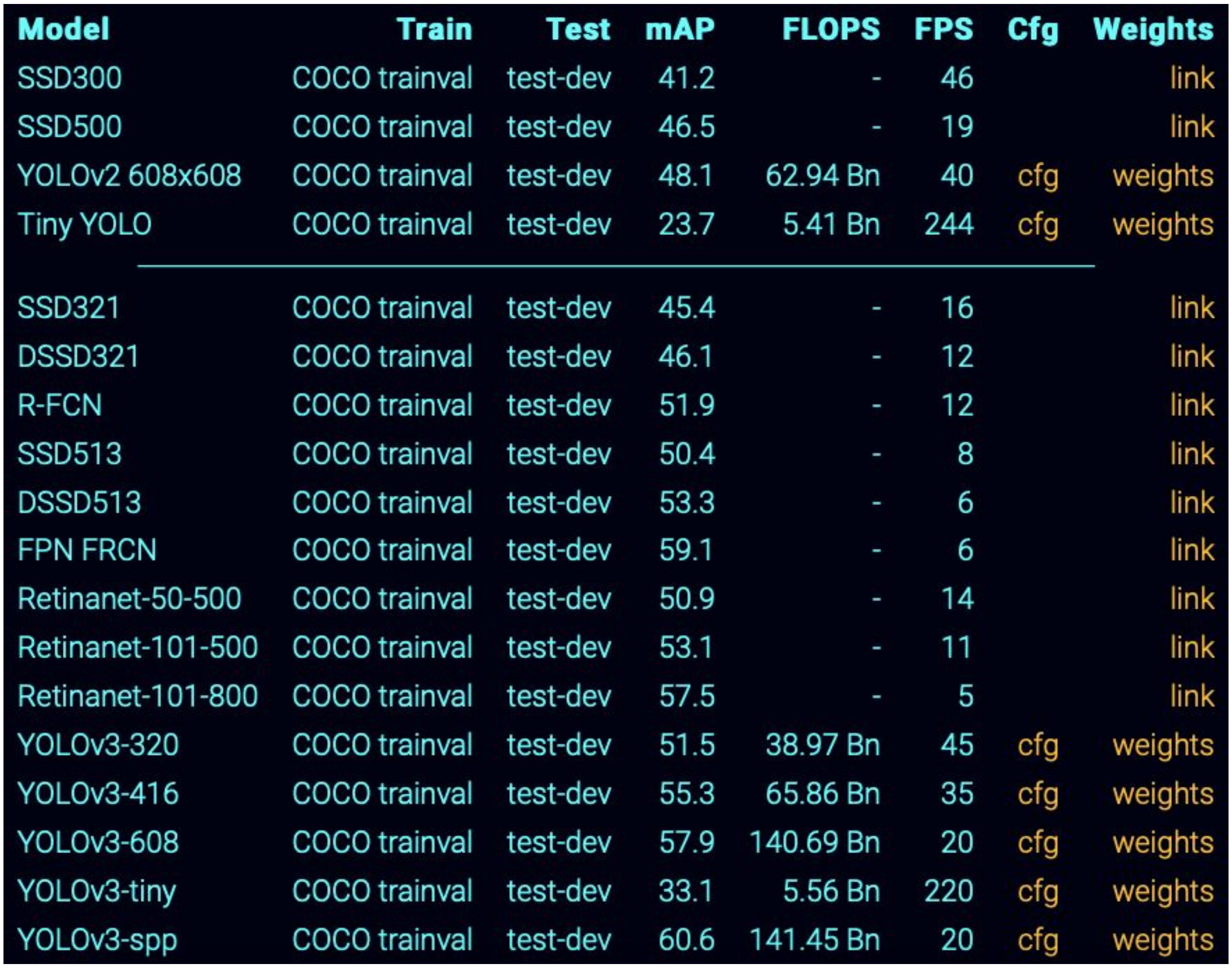
The first phase of our model is object-detection for which we'll use a camera to capture the environment. We're using pycam here with optical size ¼". And we are using machine learning for identifying objects along with accuracy of detection. With the help of a deep-learning model (YOLOv3-tiny) we are trying to detect objects and the output will be in speech format.

Furthermore, we are using Google's Text-to-speech API for converting the results into sound vibration. That sound vibration will come out through a speaker or an earpiece. We're using GPS to track the current location and a network card (eg. SIM card) which connects it to the internet.
All these are running on a Raspberry Pi 4 module.

Literature Survey

Since the beginning of the 19th century enthusiasts have been trying to help visually impaired people in exploring this universe as normal humans do (not exactly). In 1805[3], Grade-2 Braille was released for educating visually impaired people. And today is the 21st century, undoubtedly we make lots of success in this field but still many fields are untouched or not giving appropriate results. In 2005, Apple Inc. introduced VoiceOver in Mac OS X 10.4. In 2009, Google Inc. introduced TalkBack in Android version 1.6 (Donut) by Google’s Eyes-Free Project. And now many projects like GoogleGlass, IrisVision, AceSight, OrCam, NuEyes, eSight, etc. are helping visually impaired people in exploring the world using many advanced modern techniques like Augmented Reality (AR), Natural Language Processing (NLP), etc.

For many years there had been a lot of research for giving virtual vision to blinds and still going on, where few organisations came up with some solutions. Organisations like Aira, Google glasses[4], etc. developed an augmented reality (AR) based platform that offers visual interpreter service for visually impaired persons. Aira uses a technique of providing real-time solutions by implying some customer care agents who speak to the visually impaired person via telephone. On the other hand the visually impaired person have to wear the spectacles provided by Aira which connects to the internet and shares the real-time video of surrounding via camera embedded in the spectacles. The flaw in this solution is that there should be constant connectivity of the internet. Therefore, we tried to solve the problem by introducing this manual work into automated form using deep learning.



Model	Train	Test	mAP	FLOPS	FPS	Cfg	Weights
SSD300	COCO trainval	test-dev	41.2	-	46		link
SSD500	COCO trainval	test-dev	46.5	-	19		link
YOLOv2 608x608	COCO trainval	test-dev	48.1	62.94 Bn	40	cfg	weights
Tiny YOLO	COCO trainval	test-dev	23.7	5.41 Bn	244	cfg	weights
SSD321	COCO trainval	test-dev	45.4	-	16		link
DSSD321	COCO trainval	test-dev	46.1	-	12		link
R-FCN	COCO trainval	test-dev	51.9	-	12		link
SSD513	COCO trainval	test-dev	50.4	-	8		link
DSSD513	COCO trainval	test-dev	53.3	-	6		link
FPN FRCN	COCO trainval	test-dev	59.1	-	6		link
Retinanet-50-500	COCO trainval	test-dev	50.9	-	14		link
Retinanet-101-500	COCO trainval	test-dev	53.1	-	11		link
Retinanet-101-800	COCO trainval	test-dev	57.5	-	5		link
YOLOv3-320	COCO trainval	test-dev	51.5	38.97 Bn	45	cfg	weights
YOLOv3-416	COCO trainval	test-dev	55.3	65.86 Bn	35	cfg	weights
YOLOv3-608	COCO trainval	test-dev	57.9	140.69 Bn	20	cfg	weights
YOLOv3-tiny	COCO trainval	test-dev	33.1	5.56 Bn	220	cfg	weights
YOLOv3-spp	COCO trainval	test-dev	60.6	141.45 Bn	20	cfg	weights

Fig. 1. Comparison of YOLO with other detectors on the COCO dataset [5]

We're using YOLOv3-tiny for object-detection because it is quite fast in detecting objects with 220 fps. We chose this because our device will work even if the user travels in a vehicle (like car, train, etc.).

Obviously, a visually impaired person can't see but in our case he should not be deaf too because we're using a speaker or an earpiece to deliver him the message. We're directing him/her by converting text results into speech format therefore we need to convert from text to speech. Now for converting text format into speech format, we are using Google cloud text-to-speech (gTTS) because it is faster than all other TTS APIs available in the market, it is easy to use as well as has a high word accuracy.

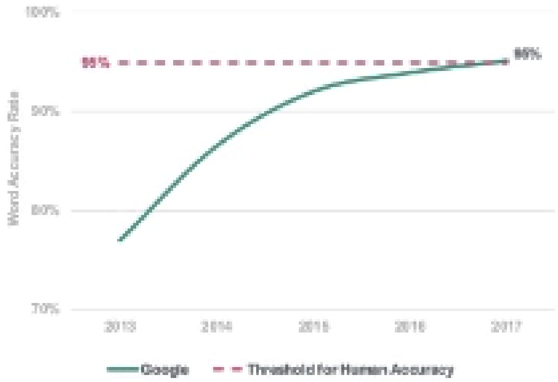


Fig. 2. Google machine learning word accuracy graph [6]

Although according to some research works it is found that Amazon Web Service Polly TTS (Polly TTS) is better than gTTS[7] but, Polly TTS is not affordable to us therefore we use gTTS.[9]

In cloud storage service we are going to use Mega cloud service because it helps to keep privacy, features end-to-end encryption by using AES-128 encryption technique and also provides huge free cloud storage space of upto 50 GB.[10]

Our project basically deals with basic requirements of a visually impaired person in exploration or movement of their bodies. It is faster, reliable and cheaper than what exists till now.

Project Proposal Plan

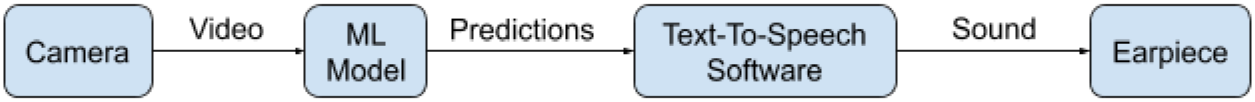


Fig. 3. Block diagram of the process of detecting environment and informing to the user

Firstly, we'll capture the scenario of the environment through a camera embedded on the device in a video format which is an input for our machine learning model. Within the model the video will pass through an object-detection algorithm which will give us the set of objects within the video frame. This set of objects will pass through a classifier whose job is to find whether the device is in a home environment, or in a forest, or any outdoor places like amusement parks, roads, etc. Finally, output of the previous will be an input for Text-To-Speech (TTS) software whose output will be an input for the earpiece or the speaker connected to the device and hence, the user will get the information about his/her surroundings.

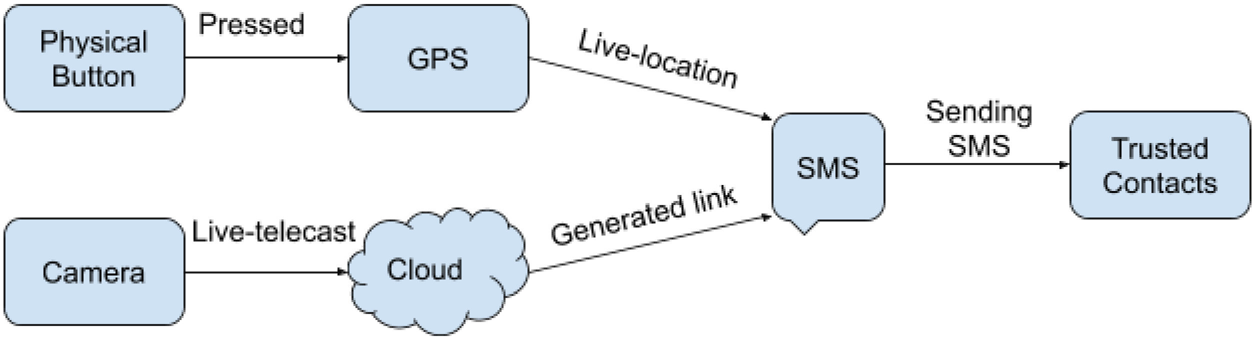


Fig. 4. Block diagram of the process execution in security feature

Security plays a vital role in the field of IoT. This is the first thing which usually comes to our mind. Here we are providing security to the user person by tracking him using a GPS module and broadcasting live-telecast (through camera) to trusted contacts saved.

This security feature will be triggered when the user presses a physical button. Along with the current live location of the user, the live-telecast through camera will be sent to all trusted contacts in the form of SMS. The video output from the camera will be first uploaded to an online server then a link will be generated for the same and that link will be added to the SMS for people to view. This is how a user will be secured when he'll feel any type of threat.

References

[1] DTE Staff. (2019. October 08). More than a quarter of the world's population has vision impairment: WHO [Blog Post]. Retrieved from https://www.downtoearth.org.in/news/health/more-than-a-quarter-of-the-world-s-population-has-vision-impairment-who-67147 [Last accessed on 14/09/2020].

[2] World Health Organization. (n.d.). GLOBAL DATA ON VISUAL IMPAIRMENTS 2010. WHO. Retrieved from https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf?ua [Last accessed on 05/09/2020].

[3] Wikipedia contributors. (2020, September 1). Braille. In *Wikipedia, The Free Encyclopedia.* Retrieved 08:46, September 8, 2020, from https://en.wikipedia.org/w/index.php?title=Braille&oldid=976232006 [Last Accessed on 08/09/2020].

[4] Thomas Macaulay. (2020, March 9). Google's AI-powered smart glasses help the blind to see [Blog Post]. Retrieved from https://thenextweb.com/plugged/2020/03/09/googles-ai-powered-smart-glasses-help-the-blind-to-see/ [Last accessed on 14/09/2020].

[5] Joseph Redmon. (n.d.). YOLO: Real-Time Object Detection. Pjreddie. Retrieved from https://pjreddie.com/darknet/yolo/ [Last accessed on 11/09/2020].

[6] Aguirre, C. C., Kloos, C. D., Alario-Hoyos, C., & Muñoz-Merino, P. J. (2018, September). Supporting a MOOC through a conversational agent. Design of a first prototype. In 2018 International Symposium on Computers in Education (SIIE)(pp. 1-6). IEEE.

[7] Jofish Kaye. (2020, May 7). Mozilla research shows some machine voices score higher than humans [Blog Post]. Retrieved from https://blog.mozilla.org/blog/2020/05/07/mozilla-research-shows-some-machine-voices-score-higher-than-humans/ [Last accessed on 14/09/2020].

[8] Cambre, J., Colnago, J., Maddock, J., Tsai, J., & Kaye, J. (2020, April). Choice of Voices: A Large-Scale Evaluation of Text-to-Speech Voice Quality for Long-Form Content. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

[9] Francesco Malatesta. (2018, April 11). It's Show Time: AWS Polly Vs Google Cloud Text-To-Speech [Blog Post]. Retrieved from https://medium.com/francesco-codes/its-show-time-aws-polly-vs-google-cloud-text-to-speech-498bcb44b3e5 [Last Accessed on 14/09/2020].

[10] Good Cloud Storage. (2020, August 31). MEGA Review – Get Free Mega Cloud Storage up to 50GB? [Blog Post]. Retrieved from

https://www.goodcloudstorage.net/mega-review/ [Last Accessed on 14/09/2020].

[11] Matthijs Hollemans. (2017, May 20). Real-time object detection with YOLO [Blog Post]. Machine Think. https://machinethink.net/blog/object-detection-with-yolo/ [Last Accessed on 08/09/2020].

[12] Karol Majek. (2017, May 1). Convolutional Neural Networks. Github. https://github.com/karolmajek/darknet-pjreddie [Last accessed on 01/09/2020].

[13] Pengyi Zhang. (2019, October 25). Github. https://github.com/PengyiZhang/SlimYOLOv3 [Last accessed on 01/09/2020].

[14] Adarsh, P., Rathi, P., & Kumar, M. (2020, March). YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)(pp. 687-694). IEEE.

[15] Zhao, H., Zhou, Y., Zhang, L., Peng, Y., Hu, X., Peng, H., & Cai, X. (2020). Mixed YOLOv3-LITE: A lightweight real-time object detection method. Sensors, 20(7), 1861.

SIGNATURE OF STUDENTS :

1. Himanshu Shekhar
2.

SIGNATURE OF MENTOR :

1.

DATE : 12/09/2020