



Computer Vision and Deep Learning-Based support system for Visually Impaired People in an outdoor environment

El-Mehdi Baahmed, Ahmed Rafik, Sarah Ouarab, Meriem Smati, Randa Cheima

► To cite this version:

El-Mehdi Baahmed, Ahmed Rafik, Sarah Ouarab, Meriem Smati, Randa Cheima. Computer Vision and Deep Learning-Based support system for Visually Impaired People in an outdoor environment. Colloque sur les Objets et systèmes Connectés 2022, May 2022, Dakar, France. <hal-04404843>

HAL Id: hal-04404843

<https://hal.science/hal-04404843v1>

Submitted on 19 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Computer Vision and Deep Learning-Based support system for Visually Impaired People in an outdoor environment

Ahmed Rafik El-Mehdi BAAHMED, Said Abdesslem MESSADI, Sarah OUARAB,
Meriem SMATI, Randa Cheima BENDJEDDOU.

[a.baahmed](#), [s.messadi](#), [s.ouarab](#), [m.smati](#), [rc.bendjeddou](#)

Higher School of Computer Science, Sidi Bel Abbes, Algeria

ABSTRACT: Visual assistance systems are still severely constrained by sensor capabilities, form factor, battery power consumption, computational resources and the use of traditional computer vision algorithms.

The challenge is to find innovative solutions that encompass such limitations and solve mobility problems of visual impaired persons (VIP), thus improving their quality of life through using new technologies and AI-Based algorithms. Smartphones and wearable devices, together with an ever-growing uptake of deep learning, computer vision, the Internet of Things (IoT), and virtual and augmented reality are key technologies in this context.

In this paper, we propose to design and implement a novel visual assistance system that relies on deep learning and edge processing to perform advanced perception tasks on a Raspberry pi 4, that-is a cost-effective, low power mobile computing platform.

The system accomplishes several tasks. It detects objects in the nearby environment of the VIP, and transforms text to sound. It achieves also obstacle avoidance through an ultra-sensor using sound alerts. It can also communicate with family members through a web server. A GPS module helps track the VIP anytime anywhere through this web server. The proposed system design incorporates an AI-based voice interface that allows for user-friendly interaction and control with an acceptable latency. It is also shown that the proposed prototype is a simple low cost, portable and unobtrusive visual assistance device.

Keywords: Visually impaired persons, Live Object Detection, Real-time, Assistive systems, Computer vision, Deep learning, YOLO, Independence for the visually impaired, Raspberry Pi, Internet of things, Kafka, NodeJS, Twilio.

1 INTRODUCTION

Vision impairment is one of the major health problems in the world and it cannot be cured only by wearing glasses. Recognizing and navigating becomes more difficult around places other than the visually impaired person's own home or places that are familiar. That's why the main contribution of the proposed work is to design a fully automated assistive technique for those people to help them perceive the objects in the surrounding, provide obstacle-aware navigation, as well as an emergency and location modules, where auditory inputs are given to users in real-time.

This system follows a Kafka architecture. A Kafka cluster consists of servers. Its management is done by Zookeeper. Producers are processes that push records into Kafka and a consumer pulls the records off it. Topics are divided into partitions, which contain records. Kafka will be illustrated in the next section.

The rest of the paper is organized as follows: First, the methodology is explored in [Section 2](#), followed by the hardware description presented in [Section 3](#), Then the prototype in detail in [Section 4](#). Finally, our conclusions and ideas for future work are discussed in [Section 5](#).

2 METHODOLOGY

2.1 General Architecture

In this paper, we present our system which is based on Edge-Fog cloud architecture that distributes task pro-

cessing on each layer. It's based on classifying compute devices into Edge and Fog layers, depending on their capabilities and ownership. The Edge and Fog layers envision bringing the cloud and its performance closer to users. As shown in *figure 1*, the Edge layer is responsible for data acquisition and local processing, and communicates heavy data to the Fog layer using a Kafka cluster for more complex processing.

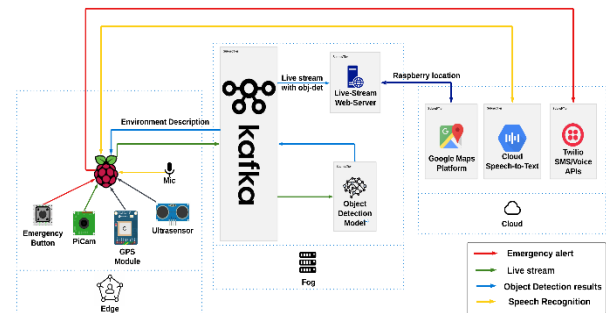


Figure 1 : General Architecture

2.1.1 Live Object Detection

When looking at images or videos, we can easily locate and identify the objects of our interest within moments. Passing on this intelligence to computers is nothing but object detection — locating the object and identifying it. Object detection is a computer technology related to computer vision and image processing that deals with identifying and tracking objects present in images and videos. The YOLOv3 algorithm is used. It predicts bounding boxes and class probability of an image that

is divided into $N \times N$ grids, and each grid is given a label depending on the objects that are marked in it.

To define which object is more dangerous than the other, a priorities system has been set by editing the used dataset. Additionally, our YOLO network has been customized to detect the five more dangerous objects.

The Raspberry Pi camera module takes the image and accomplishes the object detection algorithm on it. The detected image is converted to speech and the audio result is provided to the user through a headset. The resulting video-stream is then transferred to the web server (Node application). Node.js allows the creation of Web servers and networking tools using JavaScript and a collection of "modules" that handle various core functionalities.

2.1.2 Obstacle Avoidance

The ultrasonic sensor is designed to perform an obstacle avoidance system with voice alerts that are sent to the VIP through a Bluetooth headphones. Ultrasonic sensors work by sending out a sound wave at a frequency above the range of human hearing. The transducer of the sensor acts as a microphone to receive and send the ultrasonic sound. It uses a single transducer to send a pulse and to receive the echo. The sensor determines the distance to a target (the obstacle) by measuring time lapses between the sending and receiving of the ultrasonic pulse. Depending on the safety distance between person and obstacle, the ultrasonic will activate the voice sound to alert the VIP.

2.1.3 Self-navigation with GPS & Google Maps

This functionality consists of sharing the user's location using a GPS module (Neo6M). It is connected to the Raspberry Pi board to get the position, altitude, as well as speed, date and time on UTC. To visualize the GPS data, the Google Maps API is deployed on the web application for a geo-localization of the VIP at anytime, anywhere by VIP family members.

2.1.4 Text to Speech

Bluetooth earphones are attached to the Raspberry Pi so we can transmit voice/sound outputs. For conversion of text to speech, Pyttsx3 library is used for local processing on the Raspberry Pi. Consequently, types of objects detected in the nearby environment are transformed into sounds, as well as sound notifications in case of obstacles, and navigation directives. For example, if the VIP chooses object detection mode, then the output obtained will be the sound description of the objects in the environment. Similarly, if the VIP chooses navigation assistance mode, then the output will describe the exact path to the selected destination. The Obstacle avoidance module uses Text to Speech permanently to be able to warn the user at any time of obstacles close ahead.

2.1.5 Emergency Help

Personal safety is hard to conserve especially for blind people, interacting with an environment one cannot see

creates potential health and safety hazards. With emergency help, the user can call for help quickly and easily only by pressing a button for ten seconds. An alert message will be sent to the emergency contacts as well with the user's current location.

The activation of this trigger is done by using Twilio. Twilio is a customer engagement platform to build personalized experiences for users. It is known for democratizing channels like voice, chat, video and email through APIs.

3 HARDWARE DESCRIPTION

The Raspberry Pi 4 is a powerful low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. It is a capable little device that enables explore computing capabilities. The operating system (OS) used with Raspberry Pi in our project is Raspbian OS which is a Debian-based free operating system.

Its GPIO pins are useful to control sensors/actuators. In our case, we use as inputs an emergency button, a GPS module and an ultra-sensor. The camera is connected to the camera slot and earphones to the jack slot.

4 PROTOTYPE

As shown in *figure 2*, the system consists of a camera fixed on the glasses and attached to the Raspberry Pi located in the box with our GPS Module, Ultrasonic sensor, and the Raspberry Pi Lithium Battery.

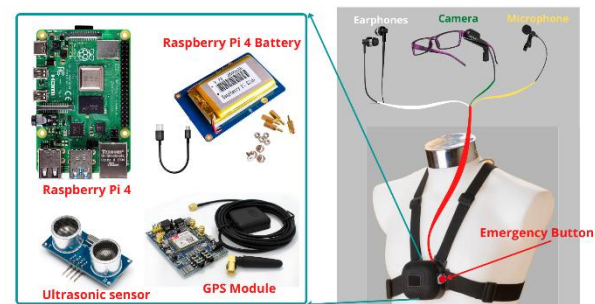


Figure 2 : Prototype

5 CONCLUSION

An assistive, deep-learning based system is proposed for visually impaired people through which they can perceive their surroundings and objects in real-time and navigate in an outdoor environment. It is a camera based visual assistant with a Raspberry Pi which can detect objects, label them and convert those labels to speech since it has an extra voice feature, ultra-sensor that can tell the distance between the VIP and the object for obstacle avoidance. In addition to that, this system has also a web application where the streaming video and the user's location will be displayed.

Future work will emphasize on face recognition in the nearby environment of the VIP.

Further improvement of this system may focus also on ways to overcome the problem of power-intensive

Graphical Processing Unit (GPU)-based hardware required by most deep learning algorithms for real-time inference. This can be done by employing instead edge Artificial Intelligence (AI) accelerators such as the Neural Compute Stick-2 (NCS2), model optimization techniques such as OpenVINO, and TensorFlow Lite, and smart depth sensors such as OpenCV AI Kit-Depth (OAK-D).

Bibliographie

- [1] B R, Sandhya; Sahana, S; Sneha, S "Third Eye for Blind", *Perspectives in Communication, Embedded-Systems and Signal-Processing (PiCES) – An International Journal* ISSN: 2566-932X, Vol. 4, Issue 11.
- [2] Rakesh Chandra Joshi, Saumya Yadav, Malay Kishore Dutta, Carlos M. Travieso-Gonzalez "Efficient Multi-Object Detection and Smart Navigation Using Artificial Intelligence for Visually Impaired People".
- [3] Lovisa Johansson "Apache Kafka for beginners - What is Apache Kafka?"