Public Transportation Assistive Device for People With Severe Visual Impairment

Paula Galindo Alvear, Juan José García and Fredy E. Segura-Quijano *Member, IEEE*

***Abstract*—Despite improvements in public transportation, visually impaired people still face difficulties navigating the system, especially when they want to take a bus alone. We proposed a device that allows the user to identify the bus destination sign and detect nearby people, providing greater autonomy in using public transport in Bogotá, Colombia. The methodology and development of the device are described, beginning with the prototype design that captures real-time information through cameras and processes it using an algorithm that communicates with the user through an application. The algorithm's customized image analysis model detects bus destination signs and people using the AWS Amazon Rekognition CustomLabels API and the OCR API. The final prototype detects 66.7% of the bus destination signs correctly identifying their route. It can detect people within two meters, allowing the mobility of visually impaired people in Bogotá's public transportation system.**

***Index Terms*—Assistive device, AWS, cloud computing, computer vision, public transport, Raspberry Pi 4.**

I. Introduction

In Colombia, the proportion of blind people is estimated to be 7,000 per million inhabitants, according to Vasquez [1]. Of this group, 87.8% are in the lowest levels of Colombia's social programs system, SISBEN [2], [3]. According to Bogotá's mobility report [4], visually impaired people prefer walking and using public transportation, around 70%. However, mobility on public transportation in Bogotá continues to be a challenge for this population, as they depend on other people or assistive devices, such as canes or guideposts, to perform daily activities such as finding their way around moving from one place to another. Although the public transportation system has undergone improvements and tools that benefit all users, such as audio stop announcements and the inclusion of Braille information at bus stops [5], visually impaired people still face considerable difficulties.

With this in mind, the main objective of this project is to develop an assistive device for the mobility of visually impaired people that focuses on public transport in the city of Bogotá (SITP). This device would allow the user to identify the bus destination sign and detect nearby people, to avoid possible collisions through computer vision services offered by AWS. Also, it is expected that all detections will be reported to the user through an application designed for the unique needs of our target user.

This work was presented as an undergraduate project at the University de los Andes in December 2022. *(Corresponding author: Paula Galindo)*
This work did not involve human subjects in its research.
The authors are with the Department of Electric and Electronics Engineering, University de los Andes, 111711 Bogotá, Colombia (e-mail: p.galindo@uniandes.edu.co).

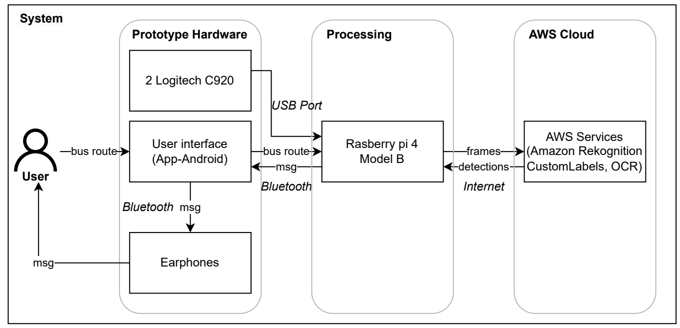

Fig. 1. Block diagram illustrating the inputs and outputs of each module, including the communication protocol denoted in italics.

This paper presents the process of the proposed solution, which begins with a detailed description of the methodology and the development of the physical prototype, as well as the algorithms for detecting people and bus signs. Next, the results of the experimental tests in the Bogotá public transportation system (SITP) are presented, along with a detailed analysis.

II. Methodology and Development

The final system comprises three main modules, as depicted in Fig. 1. The first module, the prototype hardware, encompasses the physical components that directly interact with the user. This module is responsible for collecting data to be processed and facilitating communication with the user. The second module is responsible for processing functions as the receiver and transmitter of the system's inputs and outputs, employing a Raspberry Pi 4. Finally, the AWS cloud module is introduced, where the development of the customized detection model specific to this project is elaborated. The characteristics of each component are presented below:

A. Prototype Hardware

The prototype hardware module comprises two Logitech C920 cameras. The decision to utilize this number of cameras was based on the requirement to cover a broader range of vision while avoiding blind spots. Optimal camera positioning was investigated to ensure real-time image capture and to enable destination signal detection, accounting for various user positions while taking the bus. To achieve this goal, we capture videos from different camera positions at bus stops. Analysis of the video data revealed that when the camera was oriented toward the left, it was more likely to detect the front signal of the bus. In contrast, the side signal was detected more frequently when oriented toward the front. Based on this analysis, we determined that it is necessary to use two cameras to detect the bus signal, and the optimal camera arrangement is shown in Fig. 2.

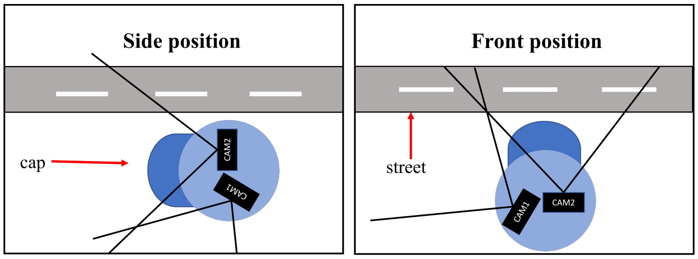

Fig. 2. Final camera arrangement. It shows the angle of visibility of the cameras when the user is in the side position (left image) and when the user is in the front position (right image).

The second component of this module is the user interface. It was determined that an application would be the best way for users to interact with the prototype. The application allows users to input the desired bus route, check their location, and receive auditory messages through headphones. The application connects to the processing module (Raspberry Pi 4) through Bluetooth to receive the messages.

The application was developed using AppInventor and was designed with the specific accessibility and navigation needs of visually impaired users. For this, we used the Android speech-to-text system to facilitate the input of the destination route. Additionally, a minimalist design was implemented to enhance navigation for the user. As we can see in Fig. 3, the interface includes three buttons that provide audible indications of their respective functions when pressed, and the haptic feedback system is activated when the user presses a button, providing a tactile response. Moreover, the user receives audible notifications for specific events, as listed in Table I.

Fig. 3. The application interface displays the process by which the user can enter their desired route. The figure on the right shows the initial interface of the application, where the user can start the process. In the second image, the speech-to-text system is displayed, which is used to enter the desired route. Finally, an interface is opened to verify if the entered route was correctly interpreted by the speech-to-text system, as shown in the third image.

TABLE I
EVENTS AND MESSAGES

Origin	Event	Message
Raspberry	Bluetooth connection realized	Connected system
Raspberry	When receiving the route entered by the user	Route received
Raspberry	When detecting the route entered	Route on the way
Raspberry	When detecting a person person nearby	Person nearby
Cell	By pressing the my location button	Location
Cell	By pressing the Enter route or the No button	Please enter the route
Cell	By pressing the Yes button	Ready for the trip
Cell	When entering the route	The route entered is: + route

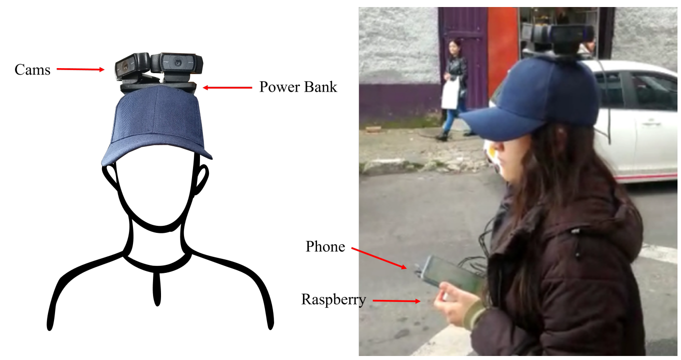

Fig. 4. The final prototype comprises two Logitech C920 cameras, a Raspberry Pi 4, a Power Bank with a capacity of 10 Mha, a Smartphone, and a MicroSD card with 16GB of storage capacity.

B. AWS Cloud

In this module, a custom detection model was designed to detect destination signs of buses and people at a distance of two meters from the user. For this purpose, the AWS Amazon Rekognition CustomLabels API service was used, which allows the creation of object detection models with few images in a short time, with the help of cloud processing. To train the model, a dataset is built with videos recorded by the prototype at different bus stops. About five recordings were made in different positions with the prototype at bus stops in Bogota to create the dataset by extracting the frames from the obtained videos. The dataset is built with images of buses at different angles, distances, and amounts of light. The final dataset consists of 489 images; 389 correspond to training and 91 to testing.

The most common metrics for evaluating the model in object detection problems are accuracy, recall, and F1 score. Accuracy refers to the percentage of correct predictions of signals or people out of the total predictions of the model. Recall refers to the percentage of labels that the model predicts correctly. This model obtained an accuracy of 86.1%, a recall of 84%, and an F1 score of 86.4%. Looking at the metrics obtained, it is concluded that the model performs well in

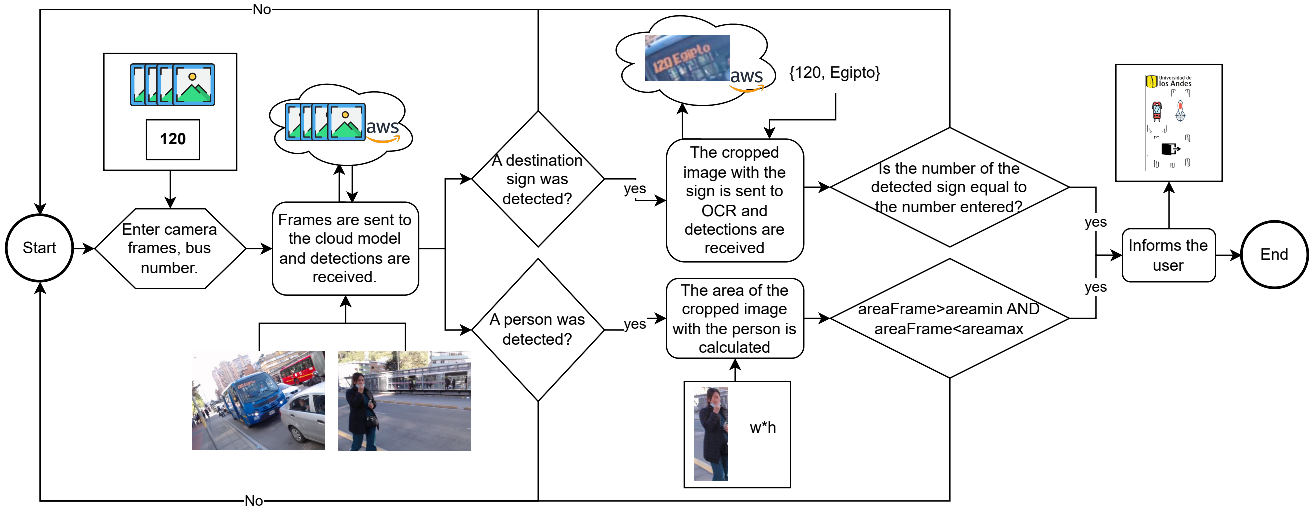

Fig. 5. Flowchart of the processing model.

detecting the specified labels. Another important metric to evaluate the model's performance is the Intersection over Union or IoU, which measures the percentage of overlap between two bounding boxes. The AWS algorithm considers a bounding box to be correct when this value is at least 50%.

On the other hand, this module has as inputs the frames obtained from the cameras which, after processing them, send the obtained detections to the processing module, as shown in Fig.1.

C. Processing

This module is in charge of individually processing the sign and person detections obtained from the AWS cloud module. The raspberry pi is used as a command center, which receives all system inputs (frames, bus number, cloud detections) and generates event messages, and processes the detections made with the algorithm in Fig. 5.

The algorithm receives the frames from the cameras and sends them to the AWS cloud module, which returns the detections as a result, and from these, the decisions are made. In case a signal has been detected, the image is cropped to obtain only the sign and to detect the route number. Then, the image is sent back to the AWS cloud module, where it is processed with the AWS text Rekognition API (DetectText), and the image characters are obtained. These are then compared with the characters entered by the user through the application. Only if the path matches the one entered is a message sent to the user by voice. In the second case, when a person has been detected, see crops the image and then pass through a size filter. This filter evaluates whether the area of the detected image is in a range set to detect people at a 2 m distance and then notifies the user. The detection range values were obtained experimentally after a series of tests with videos of people at approximately the desired distance.

III. Experimental Results

We proposed three test protocols to validate the operation of the final prototype with the integration of the application. The first one is to verify if the response time of the prototype allows informing the user in time to stop the bus. The second and third one consists of testing the algorithms that process the detected images to know their efficiency in detecting signs or people.

A. Response time test

In this test, we want to verify the time it takes for the system to notify the user when an event occurs (detects a sign or a person). An essential factor to consider in this test is that it is necessary to have a stable Internet connection to access the cloud model, so we want to evaluate how the response time changes depending on the speed of the Internet. First, a pre-recorded video is processed with the model to detect routes and people. Internally, the response times to an event are calculated with the function time(); in each case, the total time is saved. This process is performed three times with different Internet speeds, as shown in Fig. 6. As the graph shows, the relationship between the Internet and the waiting time is inversely proportional, i.e., when the Internet speed is higher, less waiting time is obtained when detecting signs or people. It is observed that the maximum waiting times in the sign detector take a value of 1.75s and in the people detector 443.1ms. Further, tests are performed to determine if these response times are enough to notify the user to act.

B. Signal detection test

In this test, we want to check the number of real-time routes the prototype can detect. Some tests are performed at a bus stop (SITP) in different positions, and the routes that pass manually are counted. Internally, the routes the prototype manages to detect are counted to compare results later. This test is performed with nine manually counted routes, 5 in the front and 4 in the side positions. The purpose is to identify the best position to detect the routes.

As shown in Fig. 7, of the nine routes that passed, about 66.7% were detected, most of them in the front position. This

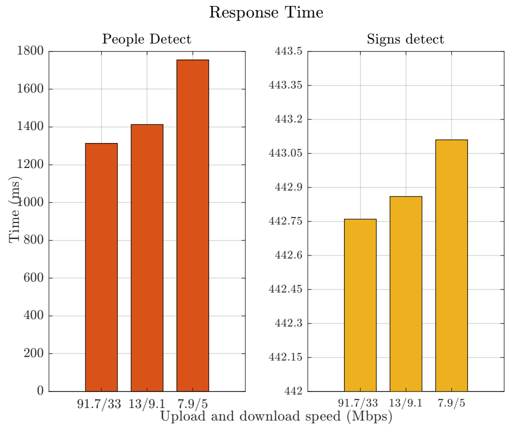

Fig. 6. Response time in terms of internet speed. a) People detect. b) Sign detect

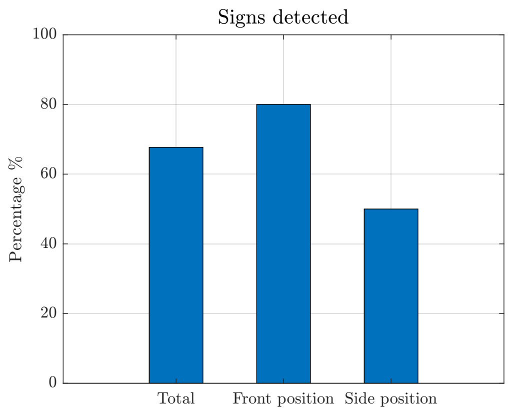

Fig. 7. Detections obtained by position. It is observed that the frontal position is the best for detecting destination signs in 80%.

is because, in this position, both cameras are observing the bus passing by, which increases the probability of detecting a sign; in the lateral position, only one camera is observing the bus. Additionally, in this test, it was observed that of the six detected routes, only three were notified early enough for the user to take action. This is mainly because the buses sometimes pass too fast to detect them and notify the user with enough time.

C. Person detection test

The purpose of this test is to evaluate the detection ability of a person at a distance of two meters from the user. In order to achieve this objective, an experimental study was carried out in a given space, where the distance of two meters from the user was measured. During the experiment, a total of six tests were carried out. A person was positioned within the detection range, i.e., at a distance of two meters, and on four other occasions, was positioned outside the detection range.

The test results showed that, on the six occasions in which the person was positioned at a distance of two meters from the prototype, the prototype correctly detected his presence on four occasions, representing a 66.7% success rate. On the other hand, when the person was located at a distance more significant or less than two meters, no detection occurred, as expected. In percentage terms, it can be concluded that the detection system has an efficiency rate close to 100% when trying to detect a person located within a range of two meters. However, this capability decreases significantly outside this range.

IV. Conclusion

This paper describes the design and development of an assistive device for visually impaired people who wish to take public transportation (SITP) autonomously. The prototype incorporates a cloud-based artificial vision algorithm that allows the detection of bus signs and people in the environment. In addition, it has an application that functions as an interface between the user and the device, allowing him to enter the desired route, know his location, and receive notifications about important events. Regarding the capacity of the route detection algorithm, it was observed that its performance is limited by the speed of the Internet and the speed at which the buses pass, resulting in a partial detection of the signs. Additionally, the prototype has an algorithm for detecting people at a distance of 2 meters that notifies the user in less than 1 second. It was identified that the DetectText API increases the response time of the sign detector, and exploring other options to detect route digits is recommended.

References

[1] R. D. Vasquez Salazar and A. A. Cardona Mesa, ”Dispositivos de asistencia para la movilidad en personas con discapacidad visual: una revisión bibliográfica”, *Rev. polite.*, vol. 15, no. 28, pp. 107–116, jun. 2019.

[2] Ministerio de Salud y Protección Social de Colombia, ”Registro para la Localización y Caracterización de personas con discapacidad - RLCPD”, *RLCPD*. [Online]. Available: https://acortar.link/oopZ1r

[3] J. C. Cubillos Alzate and S. A. Perea Caro, “Boletines Poblacionales: Personas con Discapacidad-PCD,” *Oficina de Promoción Social*, MinSalud, Dec-2020.

[4] Movilidad Bogotá, “Encuestas de movilidad 2019 Indicadores preliminares.” *Alcaldía de Bogotá*, Bogotá, 2019.

[5] J. E. Gómez Ramírez, “Transporte e inclusión, un Camino Borrascoso: Instituto Nacional Para Ciegos,” *INCI*, 15-Sep-2017. [Online]. Available: https://www.inci.gov.co/blog/transporte-e-inclusion-un-camino-borrascoso. [Accessed: 01-Mar-2023].

[6] ChatGPT was used in this paper as a writing tool. Available: https://chat.openai.com/