

# Assistant for visually impaired person

Zacharia Azzouzi

*School of CESI*

*Computer Engineering*

*Lyon, France*

*Email: zacharia.azzouzi@viacesi.fr*

Fabien Richard

*School of CESI*

*Computer Engineering*

*Lyon, France*

*Email: fabien.richard1@viacesi.fr*

Rémi Papin

*School of CESI*

*Computer Engineering*

*Lyon, France*

*Email: remi.papin@viacesi.fr*

Guillaume Woreth

*School of CESI*

*Computer Engineering*

*Lyon, France*

*Email: guillaume.woreth@viacesi.fr*

**Abstract**—In the context of helping the visually impaired this document will talk about technologies capable of reading documents, counting peoples, detecting pedestrians vehicle and signage. We compare our project with existing implementation like glasses produced by the company OrCam or the EVA project. The objective of our research team is to train on this technology and to add the detection of dynamic objects. Calculation of the speed and estimation of the possibility of interaction with other persons or not.

We are going to offer two versions of our final product, the first more affordable in order to reach a very large audience and a second a little more advanced for people with means of obtaining these glasses. Thanks to the multilingual management, our goal is to reach not only France, but a global audience.

**Keywords**— Computer vision, SLAM, DATMO, MOT, Visual Assistant

## 1. Introduction

The contemporary issue that we decide to address in this document is:

How to reduce the risk and help blind people to better perceive the world around them?

SLAM (Simultaneous Localization and Mapping) [1] methods make it possible to estimate both the trajectory of a blind person and the structure of the world from data acquired from an on-board sensor. The goal is to produce from visual data of a dynamic scene a description usable by an information system: modeling of the static world, location of the human in this world, and how other mobile objects move there. An approach combining Segmentation by Clustering and Classification makes it possible to detect on the one hand static points taken into account in visual SLAM, and on the other hand groups of mobile points used to detect and follow the dynamic components of the scene. The overall approach is assessed on the basis of images acquired in an urban environment [2]. First, we will see the existing solutions as well as their limitations. Then we will explain why our project is an

innovation with the list of functionalities from a hardware and software point of view. Finally, we will conclude with the technical validation of our project.

## 2. Project

Today people who are blind or with reduced vision use tools that are proven to be effective (white cane, brailles, guide dog). But these technologies make it possible to perceive the world in a limited way. There are technologies capable of reading documents [3], language support, technology miniaturization, counts the number of people, helps to see situations on screens, currency and face recognition, able to detect pedestrians vehicle and signage or again to driving assistance. This implementation already exists in the form of glasses produced by the company OrCam [4] or in the EVA project [5]. The objective of our research team is to train on this technology and to add the detection of dynamic objects. Calculation of the speed and estimation of the possibility of contact with the person or not. To meet our needs, we will use existing SLAM and DATMO (for Detection and Tracking of Moving Object) [6] technologies. In addition, we will implement Artificial Intelligence to perform model recognition from a collaborative database, allowing us to rapidly increase our datasets and offer a more efficient solution and implement more 'tools to increase the autonomy of blind and partially sighted people. In 2016, JAMA Ophthalmology [7] conducted a study involving 12 legally blind participants to assess the usefulness of a portable artificial vision device (OrCam) for visually impaired patients. The results showed that the OrCam device would improve the patient's ability to perform tasks simulating those of daily living, such as reading a message on an electronic device, a newspaper article or a menu. The OrCam project still contains many limitations such as reading non-handwritten text as well as battery autonomy.

The EU-funded EVA project has developed voice-activated glasses for the visually impaired. EVA, the acronym for "Extended Visual Assistant" or "Extended Visual Assistant", is based on a machine vision system that recognizes objects, text and signs and orally describes what it sees. These glasses have the advantages of being light and discreet.

The EVA project still contains many limitations such as a mandatory connection to a smartphone, the use of the smartphone

speaker, constraint in a street with a lot of noise, it is also a very expensive technology.

### 3. Application to our project

In this article, we consider that there may be moving objects in the human perceived world; despite this, we want to solve the problem of SLAM, and at the same time, estimate the state of each moving object. This also requires addressing the detection and tracking of these moving objects, while retaining the ability to detect a sufficient number of fixed landmarks in order to locate the human.

### 4. Our solution

To implement this project our team will have to design hardware but also software. The main goal of our project is to implement more features than our concurrent.

We have therefore defined the main technical evolution's that we are going to implement in our beta product.

The goal is to release a first product with features that our competitors already have (Proof of Concept).

This first product (Beta), will have the following features:

- An on-board system which attaches to glasses and does not require a smartphone
- Bone conduction headphones.
- Marker in space and vocal guide.
- Voice and gestural interaction with the product.

Secondly, after having produced our beta product, we will focus on producing a range of products (three in number):

- Product 1: this product contains all the basic functionalities (SLAM / Document reading (multilingual))
- Product 2: this product contains all the basic functionalities as well as more advanced functionalities:
  - Integrates the functions of version 1;
  - Languages and text translation;
  - Face emotion;
  - Brand identification and Identify popular places.

The hardware part will allow us to work on the physical parts, the choice of components (the majority open-source) as well as the design.

The software part will take care of embedding the SLAM and AI functionalities

#### 4.1. Hardware

We are going to use open source and accessible components, these are technologies that we already master. In addition, components such as batteries, headphones as well as cameras are easily implemented on a raspberry. An evolution is possible at the component level. The goal is to bring together the smallest and most ergonomic possible in one piece.

- Hearing system: Bone conduction
- Environmental sensor system: camera
- Materials used: Carbon (lightness)
- Power supply (24 hours of autonomy min)
- Micro computer : Raspberry pi

As we can see in the diagram above, each part of the cycle can be represented with a physic part. First, we got the user movement that is obviously made by the user. Then the identification is possible thanks to the cameras and the microprocessor. This part is also capable of making all calculations we need. Finally, the warning is sent with the bone conduction system.

RaspberryPi	Processor Cores	Processor Speed	RAM	Price
Pi 4	4	1.5GHz	2GB, 4GB, 8GB SDRAM	83,99€
Pi Zero	1	1GHz	512MB RAM	5,50€
Pi Pico	2	133 MHz	264KB SRAM	4,20€

TABLE 1. LIST OF MICRO COMPUTER

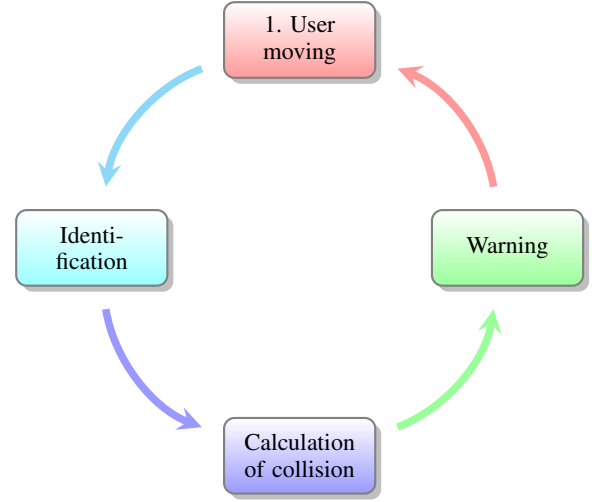


Figure 1. Visual assistant operating diagram.

#### 4.2. Software

For the software part we will focus on containerization technologies. These technology will allow us to re-use existing algorithms like SLAM and DATMO. They are based on series of image analyse. First of all, SLAM detect the interest points, then thanks to a collaborative database, we identify all the objects present in the picture. Finally, DATMO calculate the speed of all incoming objects.

With the localization of our glasses, we have the speed and the direction of the user. Our last homemade algorithm calculate the percentage of possibility of collision. If one or multiple objects are threatening the user, a precise warning is send to the user with a description of the threat.

Moreover, the use of containers allows us to deploy or modify them quickly. We will therefore be able to add containers for reading text, translating, etc.

On the IA part, we will be able to rely on containers prepared by google, allowing us to save time and make it more feasible.

#### 4.3. Technical architecture

For the realization of the technical part we will orient ourselves towards a python code of our own manufacture using the OpenCV library which encapsulates a large library for image processing.

To achieve the real-time SLAM of our camera, we will first process the image before using it. First of all, we will therefore use our camera to retrieve a video stream in real time. in the following we will spread the histogram so as to use the whole range of values of the pixels. Carry out a median blur for the treatment of the noise, work on the net, then use the harris detector in order to carry out edge detection.

Environment informations and distance detections are based on SLAM technology. The project use the world reconstructed in 3D by the SLAM to compute distance between the camera and obstacle in front of the user.

Our approach is based on LINEMOD, an efficient method that exploits both depth and color images to capture the appearance and 3D shape of the object in a set of templates covering different views of an object. Because the viewpoint of each template is known, it provides a coarse estimate of the pose of the object when it is detected.

Then we can build a 3D map environment with all the objects surrounding the user. With this map, we are now able to determine the distance of each object from the user. Moreover, from the precedents images, we can determine the speed of moving objects by using their coordinate from each 3D map.

For the sentiment detection, face detection, brand and popular place part we will rely on the Google AI image API, allowing us to retrieve this information.

## 5. Validation

### 5.1. Feasibility

This project use cheap and lightweight hardware so it's easy to make multiple embed system without having a high starting capital. For the software part we use existing technologies that we can easily bind together to match our needs. Considering the previous statements our project is simple to build and setup.

### 5.2. Usefulness

The embed system is designed to be discreet and to give information about the environment of the user. Blind people can use this object to simplify their interactions with the world. They use it to read text in front of us and even to detect logos. They can also use it to detect car or people moving around us to avoid accident. This equipment saves in autonomy and time for the user.

## 6. Conclusion

In this article we have presented a system that effectively addresses the problem of navigation in dynamic environments for blind people.

We have also achieved the state of the art at the level of our competitors in order to clear their limits.

We found lightweight equipment capable of running our different algorithms in order to produce the most suitable and efficient solution possible. For people with sight problems we have found fast algorithms capable of providing information in real time as well as competing with our various rivals on the subject. We are going to offer two versions of our final product, the first more affordable in order to reach a very large audience and a second a little more advanced for people with means of obtaining these glasses. Thanks to the multilingual management, our goal is to reach not only France, but a global audience.

## References

- [1] Ó. M. Mozos, A. Gil, M. Ballesta, and O. Reinoso, "Interest point detectors for visual slam," in *Conference of the Spanish Association for Artificial Intelligence*. Springer, 2007, pp. 170–179.
- [2] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual slam," *Machine Vision and Applications*, vol. 21, no. 6, pp. 905–920, 2010.
- [3] R. Neto and N. Fonseca, "Camera reading for blind people," *Procedia Technology*, vol. 16, pp. 1200–1209, 12 2014.
- [4] "About orcam: Meet the purpose behind the tech of amnon shashua & ziv aviram," Mar 2021. [Online]. Available: <https://www.orcam.com/en/about/>
- [5] "Extended visual assistant," Mar 2021. [Online]. Available: <https://www.eva.vision/>
- [6] A. Pancham, N. Tlale, and G. Bright, "Literature review of slam and datmo," in *Detection And Tracking Of Moving Objects*, Nov 2011. [Online]. Available: <http://hdl.handle.net/10204/5457>
- [7] M. Adrienne W. Scott, "Survey of americans on eye and vision health," Oct 2016. [Online]. Available: <https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2540516>