Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2023)
DOI: 10.54254/2753-8818/5/20230369

Research Advanced in Blind Navigation based on YOLO-SLAM

Hongyu Chen[1,†], Xiang Li[2,*,†], Zongyang Zhang[3,†], Ran Zhao[4,†]

[1] University of Electronic Science and Technology of China, Sichuan Province, China
[2] University of Jinan, Jinan City, Shandong Province, China
[3] China University of Petroleum, Beijing, China
[4] Pingfeng Campus, Zhejiang University of Technology, No.288, Liuhe Road, Xihu District, Hangzhou City, Zhejiang Province, China

* 202130312009@stumail.ujn.edu.cn
†These authors contributed equally.

Abstract: Automated blind guidance has been a hot research topic, which aims to develop efficient and inexpensive technologies to help blind people meet their daily needs. Benefiting from the rapid development of deep learning and machine vision, artificial intelligence-based blind guidance technology, especially blind guidance technology based on simultaneous localization and mapping (SLAM), has become a promising alternative. In this paper, we introduce the relevant research results of YOLO-SLAM technology in the guidance of blindness. We began by highlighting the power of YOLO, SLAM technology, and the promising prospects for current research in this field. In order to ensure that the information has a higher reference value, we focus on the practical application and improvement optimization of related papers in the past four years. We analyzed existing surveys and looked at current work, using several dimensions such as the data obtained, the sensors used, the models learned, and the human-machine interface. We compared the different methods, evaluated their testing sessions, summarized their similarities and differences, and drew conclusions by analyzing future trends in the field.

Keywords: Blind Navigation, YOLO, SLAM, Artificial Intelligence.

1. Introduction

According to the World Health Organization, there are approximately 285 million people with severe vision loss worldwide, and the number will continue to grow as the population ages. Blind guidance has always been a problem in real life that has long been expected to be improved and solved. To help these visually impaired individuals in meeting their daily needs, most common travel assistance methods for the blind are based on traditional cognition, which exists great drawbacks and deficiencies in practical applications. For example, most of the blind roads paved on the road are illegally and maliciously occupied; training guide dogs requires a lot of time, money and energy. To this end, it is imminent to develop an efficient and inexpensive general guide-blind technology.

Benefiting from the rapid development of deep learning and machine vision, artificial intelligence-based blind guidance technology, especially blind guidance technology based on simultaneous

DOI: 10.54254/2753-8818/5/20230369

localization and mapping (SLAM), has become a promising alternative. The purpose of SLAM technology is to achieve simultaneous positioning and mapping Construct. First, it moves in an unknown environment, reads information such as images of the surrounding environment in real time through multiple sensors such as cameras and inertial sensors, and preprocesses it to achieve self-positioning in places where repeated observations are made, and then constructs the surrounding environment based on the current position. map. The SLAM task perfectly fits the needs of blind navigation. Therefore, the application research prospect is very broad, attracting the research interest of a large number of scholars. Visual-SLAM is a special type of SLAM system that only utilizes cameras as sensors to perceive surrounding environment information. Compared with radar SLAM, its data collection is fast, convenient, and cheap. Visual-SLAM has become the mainstream framework in current SLAM-based blinding guidance. As shown in Figure 1, the research work related to Google Academic guide-blindness shows a clear upward trend. The basic assumption of SLAM is that the objects in the scene are stationary. However, when there are a large number of dynamic objects in the practical application of blinding guidance, such as walking people, cars, etc., they will bring wrong observation data to the system and reduce the accuracy and robustness of the system. At present, the design idea of SLAM system in dynamic environment is to treat dynamic objects in the environment as outliers, remove them from the environment, and then use conventional SLAM algorithms to process them. However, this often splits the contextual information of moving objects in different image frames, inhibiting the registration accuracy of the front end.

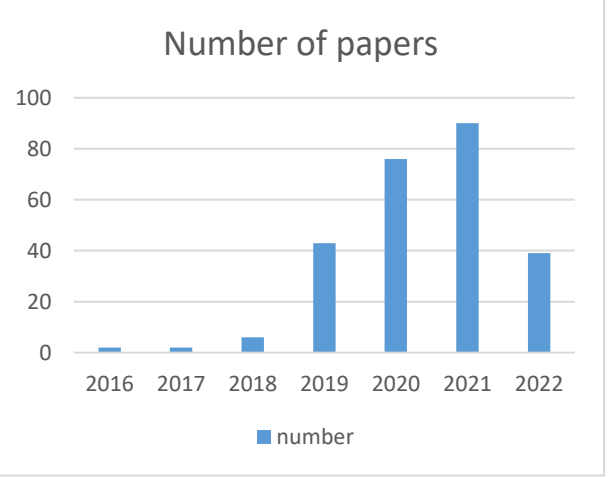

Figure 1. Number of papers referred to blind navigation.

In recent years, deep learning-based dynamic region detection methods, such as object detection, have begun to be extended to dynamic SLAM systems. YOLO is a representative object detection technology, which can realize object detection and recognition at a very fast speed. Many scholars have made many attempts in the field of blind guidance and put forward different methods combining the SLAM and YOLO. Xie et al. [1] used a depth camera and audio input from the user to obtain RGB-D video frames. The optimized path is calculated based on the dense map and the information in the scene is extracted based on YOLO. They also developed a path planning algorithm (PPA) based on A-Star algorithm and a direction correction algorithm (OCA) based on Bresenham algorithm to optimize the shortest path. Andres et al. [2] developed a ground real-time segmentation algorithm based on inertial sensors, normal vectors, gravity vectors and random sample consensus (RANSAC), which generated a global occupancy grid to perform the function of the guide. Gunethilake et al. [3] developed a prototype of obstacle detection and distance estimation using deep neural network (DNN) using SSD MobileNet architecture. A deep learning model is trained using the tens or stream object detection API and deployed on a smartphone. Promsurin et al. [4] designed compensation methods to allow a three-ring or three-wheeled scooter to travel along a planned path through a 2D navigation stack. They used a 3D Kinect camera

DOI: 10.54254/2753-8818/5/20230369

and liDAR sensor to start the mapping by using Hector Slam mapping and uses dark web YOLO to achieve obstacle avoidance function. Nafisa et al. [5] describe four different obstacle detection methods based on different sensing technologies, including ultrasound, RGB cameras, and RGB-D depth cameras. They also develop new interaction techniques. Ferdelman et al. [6] developed a headband or smart glasses, including an RGB camera and a stereo depth camera, implementing constrained RGBDslam, ASIF-NET algorithm and a convex neural network to help realize navigation for blind people.

2. Blind Navigation based on YOLO-SLAM

At present, there are a large number of experimental methods and research results in the field of blind guidance, this paper will analyze the articles in related fields in recent years, analyze and discuss how to improve the convenience and stability of blind guidance from different angles according to the comparison of different methods, and finally draw the future development trend of this field.

Depending on the sensor and subject technical framework, the existing representative blind navigation algorithms are shown in Figure 1. Common sensors include RGB-D camera, Inertial sensors, ZED Mini camera, Phone camera, 3D Kinect camera, lidar sensor, collision avoidance sensor, high-quality RGB camera and stereo depth camera. While involved key technologies include point cloud filtering, A-Star algorithm, Bressenham algorithm, RANSAC algorithm, CNNs, SSD, YOLO, AMCL and ASIF-Net, et al. We will detail the different solutions for guiding the blind in the following.

Table 1. Representative blind navigation algorithms based on various sensors and technologies.

Methods	Sensor	Key technology	Feedback
[1]	RGB-D camera	Point cloud filtering, A-Star algorithm, Bressenham algorithm	voice, vibration
[2]	Inertial sensors, ZED Mini camera	RANSAC algorithm	vibration
[3]	Phone camera	Deep neural network, SSD MobileNet architecture	audio
[4]	3D Kinect camera, lidar sensor, collision avoidance sensor	YOLO, SLAM, AMCL, dark web deep learning algorithm	assisted driving
[5]	High-quality RGB camera	CNNs, SFIT, analytical image processing, YOLO	audio
[6]	RGB camera, stereo depth camera	SLAM, ASIF-Net, YOLO V4, convective neural networks	audio

2.1. Visually Impaired Using YOLO and ORB-SLAM

They are RGB-D video frames that were acquired with the user's voice input and the depth camera. All inputs are first transmitted to the controller process, which has the ability to activate and deactivate other processes based on acoustic inputs [7]. SLAM mapping, local navigation, and object detection procedures receive video frames. The navigation process generates forward direction information and optimizes routes based on detailed maps. The fusion process receives orientation information. Based on the input video frames, the local navigation process creates a local map, determines the orientation of the barrier-free objects, and then sends the orientation data to the fusion process. By combining direction information, the fusion process creates an optimal positive direction, which is then transformed into a vibrational signal [1]. In addition, a YOLO-based object detection model is used to extract information

DOI: 10.54254/2753-8818/5/20230369

about the items in the scene from video frames that are sent to the object detector process. For the user's convenience, audio cues can be generated [1].

The goal of SLAM mapping is to create a detailed point cloud map for navigation. They created a sparse point cloud map of the user's surroundings using the ORB-SLAM framework [1]. The three essential parts of the ORB-SLAM framework's structure are loop shutdown, local mapping, and tracking. The RGB-D camera's location and monitoring, as well as the estimate and optimization of gestures by comparing feature points, are all handled through tracking [7]. Keyframes and feature points can be added or removed using local mapping. Additionally, it offers local keyframe and mapping point improvements. Once the existence of the loopback is established, the loopback is aligned and fused using a similarity transformation. The loop-off section searches for each keyframe that is detected [1]. They use a statistical filter to process a dense point cloud map in order to increase navigational efficiency, assuming there are some discrete error points in the dense point cloud map. Based on the A-Star algorithm, they created the path planning algorithm (PPA). The PPA algorithm they provide is based on greedy approaches that can discover the shortest path more quickly than certain common algorithms [1]. After establishing the optimal path, the system is designed to provide directional recommendations. They developed a Direction Correction Algorithm (OPA) based on Breesenham's algorithm to optimize the shortest path [1].

2.2. Vision-Based Blinder Assistance

Take into account camera tracking by sensing the environment, dividing the floor into three dimensions, combining local and global 2D grids, responding to closing obstructions, and producing a tactile band vibration pattern. They assessed the camera's orientation and normal vector based on the depth and inertial data, respectively, to partition the floor into three dimensions [2]. The equations for the support plane are then efficiently calculated using RANSAC (floor). In order to create a free area that contains the full trajectory and a global map that fills the area data, the local mesh is fused. They took advantage of the Jetson TX2's capabilities for high performance, low power consumption, and portability when processing dense data in parallel. These are their significant contributions [2]:

An inertial sensor, a normal vector, a gravity vector, and RANSAC are used in a real-time ground (surface) segmentation technique (random sample consistency). an algorithm that detects open space by creating a two-dimensional mesh in real time. In order to create a global footprint mesh, local data from different locations are combined. an algorithm that creates motion commands through the tactile band in response to the approach of an obstruction. An efficient obstacle detection and avoidance method based on the time-stamped map Kalman filter (TSM-KF) algorithm was developed using RGBD. designed a multimodal human-machine interface (HMI) with a voice-to-audio interface and robust electronic intelligent channel haptic interaction and designed a new map building algorithm to connect visual area description files (ADFs) and semantic navigation maps, this locates users in a semantic navigation map. This process is called semantic localization.

2.3. Blind Navigation Based Obstacle Detection

A prototype was created that uses deep neural networks (DNNs) for obstacle recognition and distance estimation, taking advantage of their real-time and high accuracy [3]. To find obstacles that could endanger navigators, combine binary threshold methods with deep learning-based monocular depth estimation techniques. A simulated environment is used to gather data that will be used to train the DNN to recognize obstacles. Estimate the distance to the obstruction using the output of the obstacle detection model. The final outcome is communicated to the user via an auditory queue after being combined with data from obstacle detection and distance estimation. The prototype system is installed on mobile devices, and an obstacle detection algorithm is applied to a live video stream taken with the camera of the mobile device. They employed the SSD MobileNet architecture and produced the training data in a simulation environment to train the DNN for obstacle detection [3]. The detection of barriers is done using a DNN-based single-depth algorithm, which estimates their distance.

DOI: 10.54254/2753-8818/5/20230369

All classes of DNN obstacle detection have an average accuracy (mAP) of greater than 70%. The system prototype can detect obstacles more quickly and with more accuracy, however there is a delay in distance estimation. Feedback from usability tests indicated that the prototype system had a greater than 65% availability and effectiveness rate [3].

The system is deployed on a pair of wearable optical glasses. Obstacle detection uses fisheye cameras and depth cameras, and distance calculations use ultrasonic sensors [8]. All algorithms, such as visual SLAM, obstacle detection, pathfinding, route tracing, and speech synthesis, are executed on one CPU. Feedback from the CPU is provided to the user via headphones and AR glasses.

AirSim is a simulation platform introduced by Microsoft to support machine learning and deep learning experiments. To build this prototype, they used an AirSim generated dataset to train a deep learning model for obstacle detection. Using the SSD MobileNet architecture, a deep learning model is trained using the tensor stream object detection API, and distances are estimated, using a single-depth PyTorch implementation. To provide audio feedback to the user, they use a pre-trained audio queue. Using the tensor stream Lite library, the prototype is converted to a mobile-compatible version and deployed on smartphones [3].

2.4. Artificial Intelligence Autonomous Vehicle for the Blind

A probabilistic positioning system for motion robots in two dimensions is called AMCL. It uses an adaptive Monte Carlo positioning method to monitor the robot's angle with respect to a predetermined map. AMCL needs a lot of parameters [9]. Laser-based maps, laser scans, conversion data, and attitude estimation are all acquired by AMCL. There are already two drive configurations for the common AMCL algorithm: differential driver and omnidirectional drive. One of their major contributions to this study was the creation of a few compensatory techniques that would enable a three-ring wheel or three-wheeled scooter to go through a 2D navigation stack along the intended path [4].

They begin mapping utilizing Hector Slam mapping with a 3D Kinect camera and lidar sensor [4]. Additionally, they try to use rtabmap to create maps (mapping based on real-time appearance) [9]. The velocity sensor's data is then required in order to determine the actual driving path. Distance, speed, and direction are all part of this information. The front wheels have encoders attached to track speed and distance. To determine the direction in which the data is supplied to the ROS, an image processing angle detection device is mounted on the steering column. To transmit signals to braking automobiles, the dark web YOLO algorithm is predicated on a 70% confidence level [4]. When the automobile starts to lose control, it will send a digital signal to the controller, which will then communicate with the relay switch underneath the car to stop the car.

2.5. Smart Walker for People with Both Visual and Mobility Impairment

On the basis of several sensing technologies, they describe four different obstacle detection approaches (ultrasonic, RGB camera, and RGB-D depth camera). In light of this, they discuss suitable processing methods that can turn the sensor output into a signal to recognize observed obstructions [5]. A voice-based user interface that communicates information through spoken messages and haptic feedback provided by a coin vibration motor attached to the walker handle were the two UI techniques they decided to employ and compare.

2.6. Navigation based on Computer Vision

They suggest creating a headband or pair of smart spectacles, each with an integrated sensor [6]. The computer will execute a Python application that applies the constraint RGB-D SLAM, the ASIF-NET algorithm, and a convex neural network to handle all the data gathered from these sensors, which will include an RGB camera and a stereo depth camera [10]. The orientation of barriers and objects of interest, their distance from the user, and the kind of information that is provided must all be revealed by the data gathered by these three components of the software. The SLAM arithmetic will be used for this. The ASIF-Net method serves as the foundation for the software's second component, which recognizes

DOI: 10.54254/2753-8818/5/20230369

objects. A straightforward concurrent neural network is then used to classify the identified objects. This allows the system to create a map while preserving hazards and interesting objects [6].

3. Experiments and performance analysis

As shown in Table 2, we have summarized the products for different years, from the types of products they use, sensors, algorithm models, the type of feedback given to the user, as well as the test objects, environments, and test results. This includes products from 2018 to the near future. First of all, from the results, as the years change, the average accuracy of object detection is also improving year by year, and the average detection time required is gradually decreasing. At the same time, comparing the test subjects of different years with the test environment, we can find that the more complex the detection accuracy in more complex environments, such as lack of light, fast movement, or rainy weather, the accuracy is also increasing, from an average accuracy of 70% to more than 90% of the object detection accuracy later. After comparison, these advances are mainly caused by the following factors. First of all, the equipment is more advanced. In different years, the camera or camera is an indispensable presence for detecting objects, and as the camera is updated, its higher and higher pixels make the input of the sample that needs to be processed have higher clarity so that it can be more accurately compared with the corresponding items in the database. Secondly, the algorithm used to process the detected image is also updated generation by generation, from the initial simple use of the SLAM algorithm to the later combination of machine learning, and the use of neural networks to supplement and optimize the algorithm. In the process, the neural network used is also gradually adjusted, which improves the processing efficiency and reduces the overall error. The application of convolutional neural networks (CNNs) to SLAM has greatly improved the accuracy of detection. However, in contrast, combining YOLO with SLAM shows a higher fit between them, which greatly improves efficiency while also reducing the error. Besides, the upgrade of hardware equipment also enables the system to carry the optimized algorithm, and accelerates the information transmission between different plates, reducing the error generated in the transmission process. From these product types, we can also note that the product types are gradually becoming more portable, from crutches to mobile apps to headwear and belts, and the way test results are fed back is gradually becoming more diverse, which provides more convenience for blind people.

In addition, for the above papers, there are still some general deficiencies, which are mainly reflected in the testing process. Firstly, most of the relevant research is based on laboratory environment testing, which is out of the actual application of the study, and the blind guide device is to assist blind people in the streets of reality And these studies generally do not do the test on the actual road; Secondly, these studies did not take into account the effects of sufficient weather conditions on product performance, such as rain and snow, smog, etc., and most of them were tested in sunny weather or undisturbed indoor environments, which is not conducive to judging the robustness of the system; Thirdly, most of the studies did not take into account the actual use of blind people. not enough blind volunteers were recruited to participate in the test, and most of the testers were visually normal researchers. In addition, for the physical condition of blind people, whether it is too eye-catching in daily use, most of these scenes have not been taken into account, and it is hoped that it can be paid attention to in future research.

DOI: 10.54254/2753-8818/5/20230369

Table 2. Test scenes and performance of representative blind navigation systems.

Methods	Test	Product type	Outcome
[1]	Indoor, no dynamic obstacles, no volunteers, no rain and fog testing	sash	The accuracy of tactile perception is 90.75%.
[2]	Indoor, outdoor, static obstacles, 10 volunteers of different ages and genders, rain and fog-free test	crutch	none
[3]	none	Mobile apps	mAP is more than 70%. The availability of the prototype system exceeded 65%.
[4]	Indoor obstacle avoidance test	Scooters	The highest result of stability and accuracy is at 35%
[5]	Item detection, daily activities, and surroundings	Camera app	Object detection is 90%, mean square error is 4.8%
[6]	Outdoor, dynamic obstacles	headband	The accuracy of close-range detection averages 94%.

4. Discussion

For a long time, with the progress and development of science and technology, helping people with disabilities has been an important research topic, and the explosion of technology in daily life has greatly accelerated the development of this field. Although it lacks certain features, there are already some sensors and technology products that can really work in real life. For example: RGB-D camera; Inertial sensor ZED Mini camera; 3D Kinect camera, lidar sensor, collision avoidance sensor; The combination of technology products: blind belts, crutches, glasses, mobile apps, headbands, and belts.

Thanks to the development of various sensors and artificial intelligence, under the basic model of YOLO-slam, the technology of blind guidance has been developed and maintained with good accuracy and efficiency (the chart has shown detailed data), in simple terms, Yolo first detects moving objects, such as car people. Then the feature points of the SLA detection are eliminated if they fall in the box of this object, and do not participate in pose estimation, reducing the dynamic object to the SLAM The impact of positioning. Shortly, this technology product will provide the blind with a safe and reliable wayfinding system, as well as detailed scene descriptions and custom object recognition. As the productivity of the above many technological products continues to increase, devices are becoming cheaper, more comfortable to wear, and more conducive to daily life.

Despite a thorough analysis of the YOLO-SLAM system and expectations for its future potential, there are still several points that require special attention and further processing: (1) The judgment of obstacles in the YOLO algorithm, which ones need to be identified and which ones do not need to be recognized, need to be summarized in more detail. (2) Since the current research-related literature has less image filtering and electromagnetic filtering, further research is needed, and there are many interferences in the real environment, which also need to be further divided and summarized. (3) For the aspect of blind navigation, because it is related to the safety of the user's life, the accuracy needs to be further improved, and based on the current research products, the emergency response to accidents

DOI: 10.54254/2753-8818/5/20230369

should also be considered comprehensively. (4) Potential in terms of governance, unified monitoring management in the cloud can be attempted to improve efficiency and security.

In addition, in most research papers, the proposed prototype has not yet been formed. The difficulty of achieving development is caused by several factors that must be addressed, including the lack of complete and available solutions, and researchers still face serious problems when developing, many situations that can be challenging for guide blindness, navigation systems: (1) There are too few test obstacles and a single scene, that is, a more comprehensive test scene is needed. (2) Performance is not strong enough, image refresh rate is low. (3) The equipment used is easily damaged, and it is difficult to use it. Moreover, some products due to product strength and other factors, such as: the cost is higher and depends on computing power; No personnel testing was conducted, and too much reliance was placed on experimental data. Here we put together these suggestions for improvement into the following mind map (See Fig.2). Finally, in the near future, SLAM and YOLO technologies interact, and more and more researchers will be working on this aspect of research, there is great potential in connected areas, smart cities, and other aspects of the integration of people with disabilities and their helpers into daily life.

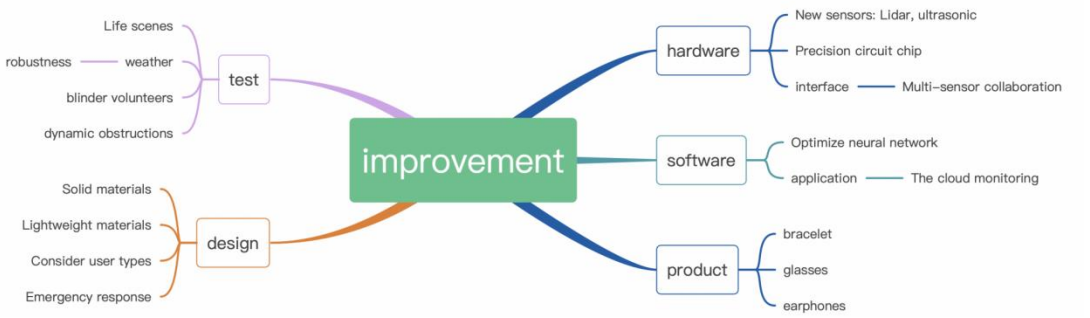

Figure 2. Possible improvement directions for the blind navigation.

5. Conclusion

In this article, through the reading summary of a variety of literature, we first classified and summarized the characteristics of SLAM and YOLO technology in recent years, respectively, introduced the advantages of SLAM and YOLO technology. Based on this, in this article, combined with the real environment, put forward the basis. The feasibility of YOLO-SLAMs in areas such as guide blindness and a powerful slam system called Yolo-Slam to reduce the influence of moving objects in dynamic environments utilize the intrinsic geometric depth information of feature points and collect, summarize specific methods.

References

[1] Xie, Zaipeng, Zhaobin Li, Yida Zhang, Jianan Zhang, Fangming Liu, and Wei Chen. 2022. "A Multi-Sensory Guidance System for the Visually Impaired Using YOLO and ORB-SLAM" Information 13, no. 7: 343. https://doi.org/10.3390/info1307034

[2] Andrés A. Díaz-Toro, Sixto E. Campaña-Bastidas, Eduardo F. Caicedo-Bravo, "Vision-Based System for Assisting Blind People to Wander Unknown Environments in a Safe Way", Journal of Sensors, vol. 2021, Article ID 6685686, 18 pages, 2021. https://doi.org/10.1155/2021/6685686

[3] Gunethilake, W.A.D.P.M.,2021."Blind Navigation using Deep Learning-Based Obstacle Detection". https://dl.ucsc.cmb.ac.lk/jspui/handle/123456789/4530

[4] Promsurin Phutthammawong.2020."Artificial Intelligence Autonomous Vehicle for the Blind". http://www.dulyachot.me.engr.tu.ac.th/EECON-43_CP-5.pdf

[5] Mostofa, Nafisa, Christopher Feltner, Kelly Fullin, Jonathan Guilbe, Sharare Zehtabian, Salih Safa Bacanlı, Ladislau Bölöni, and Damla Turgut. 2021. "A Smart Walker for People with Both

DOI: 10.54254/2753-8818/5/20230369

Visual and Mobility Impairment" Sensors 21, no. 10: 3488. https://doi.org/10.3390/s21103488

[6] Ferdelman, Kai (2021) Using computer vision to aid navigation for people with visual impairements. https://purl.utwente.nl/essays/87769

[7] Aladrén, G. López-Nicolás, L. Puig, and J. Guerrero, "Navigation assistance for the visually impaired using rgb-d sensor with range expansion," IEEE Systems Journal, vol. 10, no. 3, pp.922–932, 2016.

[8] D. Tudor, L. Dobrescu, and D. Dobrescu, "Ultrasonic electronic system for blind people navigation," in 2015 E-Health and Bioengineering Conference (EHB), pp. 1–4, Iasi, Romania, 2015.

[9] Xu, Q.; Lin, R.; Yue, H.; Huang, H.; Yang, Y.; Yao, Z. Research on small target detection in driving scenarios based on improved YOLO network. *IEEE Access* **2020**, *8*, 27574–27583.

[10] Tapu, Ruxandra & Zaharia, Titus. (2017). Seeing Without Sight — An Automatic Cognition System Dedicated to Blind and Visually Impaired People. 1452-1459.10.1109/ICCVW.2017.172.