A comprehensive study on supermarket indoor navigation for visually impaired using computer vision techniques

Dr. S. Kayalvizhi

Computer science and engineering
Easwari engineering college
Chennai, India
kayalvizhi.s@eec.srmrmp.edu.in

Riya Ponraj

Computer science and engineering
Easwari engineering college
Chennai, India
riyaponraj07@gmail.com

Priya Dharshini S

Computer science and engineering
Easwari engineering college
Chennai, India
priya.srinivas2101@gmail.com

Roshni S

Computer science and engineering
Easwari engineering college
Chennai, India
roshnisrinivas8101@gmail.com

***Abstract*— The ability to navigate is a fundamental skill that every person must have. People who are visually challenged need regular support when travelling from one place to another. It is difficult for them to have an autonomous shopping experience in a supermarket. The project's goal is to make one of the daily life's task easier for visually impaired persons. The project entails aiding visually impaired customers by guiding them to the destination of the corresponding product sections using an indoor navigation technique. The program aids in providing the visually impaired with shopping environment as well as assistance on how to buy their preferred goods. When logging into the Android app, the user or someone who is blind will speak about their shopping list preferences (voice input). The camera is turned on to provide the live-streaming video. The user can then begin navigating indoors to reach the section where the product is found. This incorporates the technology that is used in automated self-driving cars. Modeled high-definition simulated maps of the supermarkets are used to navigate inside employing the shortest distance. The first item from the list is selected, and the user's location is identified by matching the features in the image frame with the map and used to compute the path to the product section from there. It provides speech output for navigation. Someone who assists the people to get the products from shelves will help them to get the exact item. The buying process continues until the last item on the list is bought. As a result, this will make tasks comparatively easier for visually impaired persons and make indoor navigation possible.**

Keywords*— *indoor navigation, machine learning, deep learning, image recognition, CNN, neural networks, SLAM, voice to text

I. INTRODUCTION

Millions of people struggle with visual impairments. A diminished capacity to see to a point where it results in issues that are unfixable by standard methods, such as wearing glasses, is referred to as a visual disability, also known as vision impairment or vision loss. A person's quality of life and capacity to communicate with others might be negatively impacted by a visual impairment, which can also impede their ability to carry out daily duties. The most severe kind of visual impairment, blindness, can affect a person's ability to accomplish daily tasks. Everybody has to be able to navigate, it's a fundamental skill. People with vision impairments require ongoing assistance when moving from one location to another. It is quite challenging for them to shop independently in a grocery. In large buildings, GPS-based location is inconsistent, which contributes to the fact that interior navigation is frequently more difficult than outdoor navigation. Although several systems for providing interior localization have been tested, none has yet reached a widely acknowledged standard equivalent to GPS-based navigation systems. Aside from technical limits, a

978-1-6654-9294-2/$25.00 ©2023 IEEE

consideration that exacerbates the indoor navigation challenge is multiple-levels of 3D constructions with limited nomenclature standards for accessible sections or propping up of specific places, that enhance outdoor transport. Obstacle detection applications for indoor autonomous robots or assistive devices for the blind require centimeter-level accuracy and cannot rely on approximation localization. The accuracy standards for contemporary AR solutions are significantly higher. The user's mobile device's position and orientation must be centimeter- to millimeter-accurate for these apps to provide seamless integration of the multimedia material through video flow on smartphones or across the lenses of smart glasses. In this work, the viability of a system design for interior navigation by visually impaired people using deep learning methods for image recognition by extracting the features from image and comparing it with the HD simulated maps of the shopping complex is investigated.

II. LITERATURE REVIEW

[1] The paper titled “Visually Impaired Indoor Navigation using YOLO Based Object Recognition, Monocular Depth Estimation, and Binaural Sounds” (Xinrui Yu, 2021), by Xinrui Yu, Sukesh Davanthapuram, and Jafar Saniie, advocated the construction of authentic spatial audio-generating software that uses binaural sound and computer vision techniques to assist visually impaired people in indoor navigation.. It includes four components YOLO (a computer vision technique) for object detection, monocular depth estimation methods for creating depth information from a picture. The KITTI dataset is applied to train the model. Several objects were used to check the precision of the algorithm. A 400mm average inaccuracy was estimated for object detection. The sample items chosen included a bench and a bottle. The presented result includes the azimuth angle, object's direction, elevation angle, and distance from the camera. For azimuths, the HRTFs are measured in increments of 15 degrees, and for altitudes, 5.625 degrees. Real-time spatial audio is created utilizing mono audio sources after these indices are computed. Text files are used to store the categories of the discovered objects and their locations. In this work, the viability of a system design for interior navigation by visually impaired people was investigated. (Xinrui Yu, 2021)

[2] The paper titled “An Indoor Navigation app using CV and sign recognition”, by Seyed Ali Cheraghi, Giovanni Fusco, Leo Neat, and James M. Coughlan proposed an easy-to-use iOS navigation software that delivers turn-by-turn directions to the specified location It combines dead reckoning with visual-inertial odometry (VIO) and understanding about the destination in the environment deduced from information - based sign detection systems and map constraints. The model incorporates a 2D plan, a sign recognition algorithm, and a dead reckoning algorithm. With the help of U-Net implementation and their own assumptions which gave satisfying results, they devised an effective algorithm that can detect the Exit signs. Next, The turn-by-turn navigation instructions based on their localization algorithm delivered spoken directions that guided a visually impaired passenger to the intended destination in real time. The future scope of this project focuses on optimizing the user interface, optimizing the sign recognition algorithm, and experimenting with augmented reality capabilities.

[3] Wuttichai Vijitkunsawat and Peerasak Chantngarm's paper titled “Comparison of Machine Learning Algorithms on Self-Driving Car Navigation Using Nvidia Jetson Nano” (Vijitkunsawat, 2020)describes the various machine learning algorithms that can be used in self-driving cars for navigation. It focuses on SVM, CNN-LSTM, and ANN-MLP and discovers that CNN-LSTM produces satisfactory results. Each model was trained using an Nvidia Jetson Nano developer kit board and a PID system for automated driving. SVM can help with classification and regression problems, but it is inefficient. ANN-MLP can detect and classify using feed-forward propagation but fails to detect at 3km/hr. CNN-LSTM has five layers that solve classification problems and provide the highest accuracy. The camera records live data at 5 frames per second. CNN-LSTM performs well in testing by showing high efficiency.

[4] The paper titled “Monocular Camera-based Point-goal Navigation by Learning Depth Channel and Cross-modality Pyramid Fusion”, by Heming Du, Tianqi Tang, Xin Yu , Yi Yang, proposed Geo-Nav, a highly efficient point-to-goal navigation structure. It includes visual perception as well as navigation. In first part, a self-supervised depth estimation network (SDE) is built which can learn RGB images and their depth using cross-modality pyramid fusion. In the latter part, the obtained visual representations are used in the navigation policy network and LSTM is used to generate navigation actions. It can efficiently explore scene geometric cues without relying on ground-truth depth information. First, SDE is trained on the training data sampled from the Gibson environment. After that, their navigation network, including the policy network and visual perception encoder is optimized by employing two branches to extract features from the RGB images and depth maps. Geo-Nav obtains informative visual representations and obtains a 91.1% of success rate. In conclusion, the model achieves superior visual perception ability by using SDE and improves both the effectiveness and efficiency of the navigation system by outperforming the state-of-the-art.

[5] “Nitika Garg, Kanakagiri Sujay Ashrith, Gulab Sana Parveen, Kotha Greshwanth Sai, Anish Chintamaneni, Fatima Hasan”, and states use AI to spot and present the HD mapped route for proper operation of self-driving cars in their paper titled “Self-Driving Car to Drive Autonomously Using Image Processing and Deep Learning” . Convolution2D was trained using the NVIDIA version. Their simulator vehicle captures images from the left, right, and center. The data was then preprocessed, with 5 Convulation2D layers, 4 dropout layers, and 4 dense layers added. Regression, pattern recognition, SVM, cluster algorithms, and decision matrix algorithms were all tested. The decision matrix algorithm performed better than the others. The set of rules is provided by AdaBoosting.

After ten epochs, the loss in training was 0.0343 and the loss in validation was 0.0275.

[6] The neural network-based self-driving car model is described in the paper "Real-Time Self-Driving Car Navigation Using Deep Neural Network (Do, 2018) by Truong-Dong Do, Minh-Thien Duong, Quoc-Vu Dang, and My-Ha Le". CNN, in particular, has demonstrated superior performance in perception tasks. A Raspberry Pi 3 with a front-facing camera was used. The road images with coordinated gearshift angles were used to train the model. The results show improved accuracy in autopilot mode, which can reach speeds of up to 6km/hr. The end result is 92.38% precision on training and 89.04% on testing data.

[7] The research paper "Navigating Self-Driving Vehicles Using Convolutional Neural Network" by
Minh-Thien Duong, uses a method for self-driving automobile navigation. A method has been created using the UDACITY platform for CNN training and simulation of a vehicle model with no man. Three cameras were placed in front of an automobile in the left, right, and centre locations to collect data. The data are in form of images captured by cameras. In all, there are 15504 samples. In this investigation, labels with two parameters would also be created for the steering angle and speed.. After data gathering, CNN will be trained and used to direct the vehicle to achieve these criteria. In the study, we simulate automated driving by estimating steering wheel angles and speed values from unprocessed photos using CNN training. Three cameras were used to gather data, which was then analysed before being put into a CNN to determine the value of steering angle and speed. The weight of the CNN is changed to provide a better outcome after comparing the suggested command to the intended command for that picture. The model accuracy in predicting steering angle and speed, which is 98,23%, is remarkable. This research's modelling software output shows that the car can drive autonomously, follow the lane steadily, and stay in the lane. On the other hand, this work contains a number of flaws. First off, because the replicated atmosphere is so precise, there is hardly any noise coming from the external environment. The second is the issue of the vehicle's mechanical fault, which is generally disregarded.

[8] The paper titled as "A Hassle free Shopping Experience for the Visually Impaired: An Assistive Technology Application" by Kumudha Raimond and Sherin Tresa Paul have proposed system. The goal of this work is to provide an in-depth explanation of some of the challenges that visually impaired people encounter, as well as to suggest an Assistive Technology framework. The goal is to enhance the Visually Impaired's Quality of Life in as many ways as possible. The suggested AT framework use Optical Head-Mounted Display (OHMD) that responds to voice instructions and has shown good results for the visually handicapped. However, the program's reliance on QR codes is inconvenient. For object recognition tasks, a visual-tactile fusion approach is developed.

[9] The paper was written by Satan and was named "Bluetooth-based indoor navigation mobile system". In this research, an unique Bluetooth beacon-based system for Android is proposed. The purpose of this paper is to outline the creation of an Android indoor navigation system that uses Bluetooth beacons. Most people find basic floor maps difficult to interpret, and it takes money to rebuild them when the information on them changes. Databases for indoor positioning and navigation applications are simple to update when necessary, since everyone today carries a smartphone. These systems should leverage sensors, which are already present in contemporary smartphones, for high availability. These sensors include gyro, Bluetooth, and WiFi. Bluetooth is therefore utilised here. Radio frequency signals that are emitted by Bluetooth beacons can be utilised to determine distance. The applicability of the offered Indoor Navigation Application is supported by experimental findings. The application's on-route accuracy was 89.2% tracking and planning.

[10] The paper titled as "Deep Camera: A Fully Convolutional Neural Network for Image Signal Processing" (Ratnasingam, 2019)by Sivalogeswaran Ratnasingam proposes a typical camera that would execute numerous signal processing stages in order to rebuild an image. When these functions are conducted in stages, the residual error from each stage converges in the image, reducing the characteristics of the finished reconstructed image. This study employs a CNN network to undertake defect pixel correction, noise removal, white balancing, visibility correction, demosaicing, color transformation, and gamma encoding (CNN). Extensive testing has shown that the envisioned CNN-based visual signal signal processing processing system outperforms most systems. To generate a picture from a raw Bayer picture for display technology, an ISP pipeline would be required. Picture signal processing is done in a traditional camera. Picture signal processing in a traditional camera is done on specialized hardware with a modular design. To rebuild an image, a typical ISP pipeline goes through several processing phases.

[11] The research article "Vision-Based mobile Indoor Assistive Navigation Aid for Blind People" (Li, 2022) was published in March 2019 by Juan Pablo and Bing Li. The plan for the supermarket's indoor map is created in this approach using a 2D grid. The technology uses a camera specialized with RGB-D to detect objects, enabling users to navigate without obstructions. Map with a time stamp For an Obstacle Avoidance Approach, one uses the Kalman filter (TSM-KF) method. ISANA is a well-known system that can identify obstructions in the way of the next path and, at the same time, updates a new path to get around them. Only obstacle alerts are provided by the ISANA model for obstructions that do not interfere with the intended route.

[12] The paper title "Monocular Depth Estimation Based On Deep Learning" by (Xiaogang, 2020) Qiyu Sun, Chaoqiang Zhao, Chongzhen Zhang, Yang Tang have surveyed the current monocular depth estimation methods and reviewed some representative existing methods. As the geometry-based and sensor-based methods have their own drawbacks like dependency on many image pairs and

sensitivity to sunlight, deep-learning based methods have proved to provide the promising results as they are based on single image. Various neural networks like CNN, RNN, VAE, GANs shows effectiveness in analysing the depth of an image. KITTI dataset, cityscapes, and make3D have been used to test the model. Evaluation metrics like RMSE, RMSE log, Abs Rel are used for checking the accuracy of the model. Different frameworks involving GAN, supervised methods based on GAN and unsupervised and semi-supervised based on GAN are evaluated. Unsupervised methods have shown higher accuracy compared with other methods as it finds a more efficient geometric constraint. Even though loss terms are provided, unsupervised algorithms trained on monocular films exhibit scale uncertainty and lack of consistency, and the issue remains unresolved. When the train and test sets are in diverse spheres or gathered by multiple cameras, interchangeability influences the quality of same system. Finally, there has been little research on the framework of deep learning-based monocular depth estimation techniques.

[13] Federica Barontini, Manuel G. Catalano, and Lucia Pallottino proposed a wearable haptic technology-based indoor navigation module in their article titled "Integrating wearable haptics and obstacle avoidance for the visually impaired in interior navigation." It includes an RGB-D camera, a processor to identify impediments, and a wearable gadget to help with interior systems. Interviews are done with a person who has been blind since a young age, and the findings of the interviews are then presented. It demonstrates how blind people like to explore objects up close and touch them in unfamiliar settings, demonstrating the need for a typical white cane. These findings led to the development of the system's wearable and hands-free components, a dependable cane-free obstacle detection method, and a haptic feedback and communication system to warn the user about the outdoors when travelling autonomously. The navigation algorithm is then utilized to obtain the input image and produce commands to turn right, turn left, turn straight, and turn around. Testing the system's functioning on a variety of visually impaired people allows for an analysis of its effectiveness. Even though there were some variations in the results for the command recognition test, the participants
could readily understand it, and a joystick was employed for convenience. The navigation task experiment is separated into three groups: only using a cane, only using a cuff device, and using a cuff and cane. According to the results, participants can quickly and readily understand the instructions. For those who are skilled at handling a cane, employing a cane alone produces better results than using a cane and a cuff device together. However, those who struggle with cane use exhibit noticeable results when using cane and cuff together. Therefore, using CUFF with cane does not impair the performance. Everyone is capable of avoiding obstacles. Future goals include making the device more feasible, addressing more user needs, reworking the hardware design, and enhancing the object detection algorithm. The environment's lighting conditions are also very important because a camera cannot see in complete darkness.

III. SURVEY OVERVIEW:

TABLE 1

Paper No.	Technique Used	Results	Limitation/Future scope
1	Object detection and depth estimation	The model can predict an average of 400mm	Recent studies on this process have not been very extensive.
2	CV and sign recognition	Added enhancements by improving Exit sign segmentation, determining a sign's range.	Optimizing the user interface, estimating floor changes with the built-in smartphone barometer, improving the sign recognition algorithm, and experimenting with augmented reality features.
3	Machine Learning Algorithms	Shown CNN-LSTM has given high accuracy	When more obstacles and high rates are added, the percentage of all algorithms' accuracy rate gradually decreases.
4	Learning Depth Channel and Cross-modality Pyramid Fusion	Obtains a 91.1% of success rate by using an evaluation metric called Success weighted by Path Length (SPL)	SDE produces inconsistent depth maps and hindrance sometimes
5	Image Processing and Deep Learning	Obtains 0.0343 training loss	Requires much data
6	Deep Neural Network	Obtains 92.38% accuracy	improved prediction accuracy when training the network, identifying and employing low-latency cameras
7	Convolutional Neural Network	Obtains 98.23% accuracy	The issue of the vehicle's mechanical fault is

			generally disregarded.
8	An Assistive Technology Application	Focusses on recent advancement in assistive technology to increase quality of line in visually impaired community	Building a prototype, putting it through testing with VI and sighted users, evaluating the findings, making an algorithm to determine the item relevance factor, and testing it concurrently with prototype development
9	Bluetooth-based mobile system	The software's detecting and prediction efficiency while in the movement was 89.2%.	Administrative duties posed the most difficulty. To make the system installation process easier, a visual interface to aid in planning might be created.
10	Fully Convolutional Neural Network for Image Signal Processing	ISP based on CNN performs sup erior than the rest of the approach	Trained neural network restore images have been damaged by unknown degradation functions
11	Vision-Based mobile Aid	Pose update rate is 100 HZ through VIO, obstacle detection takes place under 5 HZ, obstacle avoidance occurs in 3 – 5 seconds	The entire mechanism uses a lot of energy. And ISANA only lasts for about an hour and a half due to operating on heavy load.
12	Monocular Depth Estimation Based On Deep Learning	Unsupervise d methods have shown higher accuracy compared with other methods	Scale inconsistency and ambiguity hamper unsupervised techniques. Performance is impacted by network transferability when data is gathered by several cameras.
13	Integrating wearable haptics and obstacle avoidance	Participants were able to properly identify the direction as a result discriminatio n task's performance.	Improving the object detection algorithm, addressing more user needs, increasing the device's viability .

IV. ISSUES IN EXISTING SYSTEMS

First off, the majority of research uses wearable technology, which is costly for someone who is visually impaired. The device is not accessible to everyone. Next, some systems use cutting-edge technology, such as augmented reality, which isn't always useful and exists only for object detection. Furthermore, not all supermarkets can afford to set up as many RFID tags as the majority of indoor navigation modules require. Most of the technologies were implemented in real life only for outdoor navigation like automated cars and robots. There is no implementation where the methodologies used for self-driving cars can be used for assisting visually impaired in indoors for navigating.

V. METHODOLOGIES USED:

1) *User Interface* : The user interface allows users to interact with the application and accept voice inputs that are converted into text to obtain the preferred route for indoor navigation. The user interface is in charge of communicating information to the user, such as the pathway model as seen by the camera on a mobile phone, with the voice assisting in navigation. The user interface is also responsible for providing the user with timely alerts.

2) *Deep Learning:* Deep learning is a branch of machine learning that uses artificial neural network (ann) to undertake complex calculations on huge amounts of data. Deep learning algorithms teach computers by using instances to train them.

- CNN: Convolutional neural networks, or ConvNets, are multi-layer neural networks used mainly for image processing and object detection. A pooling layer is one of several layers in CNNs that practice and retrieve features from data. It is specifically designed for image recognition and pixel data processing tasks.

- LSTM: Long-Term RNNs with Short-Term Memory can understand and retain long-term interconnections. For long stretches of time, the default value is to recollect past information. LSTMs store information over time. They can assist you in recalling previous inputs. They have such a chain-like framework with four interconnected layer upon layer that interact in different ways.

- SLAM: SLAM (simultaneous localization and mapping) is a driverless car technique that makes it easy to create a map even while positioning your machine inside it. Utilizing SLAM algorithms, the vehicle could indeed map out unidentified surroundings. Map data is utilized by engineers to accomplish duties such as trajectory tracking and collision avoidance.

- YOLO: You only have one look. (YOLO) is a cutting-edge, real-time object detection system. In terms of speed, YOLO does have an obvious advantage. The YOLO algorithm is much faster than its contenders, reaching rates of up to 45 frames a second.

- Computer vision: Computer vision is a subfield of artificial intelligence which enables computers and structures to derive meaning from digital photos, videos, as well as other graphic inputs and afterwards behave or suggest on that data.

- NLP - Natural Language Processing (NLP) is a branch of science of computer, human language, and artificial intelligence. Machines use this technology to understand, analyze, manipulate, and interpret human languages. It assists developers in organizing knowledge for tasks like translation, fully automated summarization, text translation, and voice recognition

VI. CONCLUSION

An overview of all the cited papers is provided in this survey paper. The feasibility of developing a technology to aid visually challenged people in navigating an interior surrounding has been investigated in this paper. Unlike tags and wearable gadget systems, which are expensive and not accessible to everyone, the system is a user-friendly mobile application. The ability for visually impaired persons to shop independently in stores allows for a hassle-free experience. Though present systems do not have navigation within a supermarket with an average accuracy rate and picture recognition using various ways, our system will be able to design a system providing indoor navigation and image recognition.

VII. REFERENCES

[1] Xinrui Yu, S. D. (2021). Visually Impaired Indoor Navigation using YOLO Based Object Recognition, Monocular Depth Estimation, and Binaural Sounds. *2021 IEEE International Conference on Electro Information Technology (EIT).* Mt. Pleasant, MI, USA: IEEE.

[2] Fusco, G. C. (2020). An indoor navigation app using computer vision and sign recognition. *International Conference on Computers Helping People with Special Needs* (pp. 485-494). springer.

[3] Vijitkunsawat, W. a. (2020). comparison of machine learning algorithm's on self-driving car navigation using Nvidia Jetson Nano. *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (pp. 201-204). IEEE.

[4] Tang, T., Du, H., Yu, X. and Yang, Y., 2022, June. Monocular Camera-Based Point-Goal Navigation by Learning Depth Channel and Cross-Modality Pyramid Fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 5, pp. 5422-5430).

[5] Garg, N., Ashrith, K.S., Parveen, G.S., Sai, K.G., Chintamaneni, A. and Hasan, F., 2022. Self-driving car to drive autonomously using image processing and deep learning. *Int. J. Res. Eng. Sci. Manage.*, *5*(1), pp.125-132.

[6] Do, T. D. (2018). Real-time self-driving car navigation using deep neural network. *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)* (pp. 7-12). IEEE.

[7] Duong, M.T., Do, T.D. and Le, M.H., 2018, November. Navigating self-driving vehicles using convolutional neural network. In *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)* (pp. 607-610). IEEE.

[8] Paul, S.T. and Raimond, K., 2019. A Hassle-Free Shopping Experience for the Visually Impaired: An Assistive Technology Application. In *Computer Aided Intervention and Diagnostics in Clinical and Medical Images* (pp. 207-217). Springer International Publishing.

[9] Satan, A., 2018, May. Bluetooth-based indoor navigation mobile system. In *2018 19th international carpathian control conference (ICCC)* (pp. 332-337). IEEE.

[10] Ratnasingam, S., 2019. Deep camera: A fully convolutional neural network for image signal processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 0-0).

[11] Li, B., Munoz, J.P., Rong, X., Chen, Q., Xiao, J., Tian, Y., Arditi, A. and Yousuf, M., 2018. Vision-based mobile indoor assistive navigation aid for blind people. *IEEE transactions on mobile computing*, *18*(3), pp.702-714.

[12] Xiaogang, R. W. (2020). Monocular depth estimation based on deep learning: A survey. *2020 Chinese Automation Congress (CAC)* (pp. 2436-2440). IEEE.

[13] Barontini, F., Catalano, M.G., Pallottino, L., Leporini, B. and Bianchi, M., 2020. Integrating wearable

haptics and obstacle avoidance for the visually impaired in indoor navigation: A user-centered approach. *IEEE transactions on haptics*, *14*(1), pp.109-122.