Indoor Navigation System for Visually Impaired People using Computer Vision

Adrian Leyte Fraga[1,2], Xinrui Yu[1], Won-Jae Yi[1] and Jafar Saniie[1]
[1]*Embedded Computing and Signal Processing Research Laboratory (http://ecasp.ece.iit.edu)*
Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago IL, U.S.A.
[2]*E.T.S.I. Industriales, Universidad Politécnica de Madrid, Madrid, Spain*

***Abstract*—This paper studies an indoor navigation guidance system for visually impaired people using Artificial Intelligence (AI) and computer vision techniques to guide users via optimal path based on quick response (QR) code markers and collision avoidance system supported by the monocular depth estimation algorithm. The proposed system utilizes a set of QR code markers, as location beacons, to generate an optimal path for the user to reach the destination. The identified QR code markers are used to correlate and compare information from the online database containing valuable information such as current location information, and the next possible locations from the current location. In addition, our system embraces a safety feature, a collision avoidance system, by utilizing the monocular depth estimation algorithm to identify obstacles in the path to the desired destination. All obtained information is provided to the user through a text-to-speech engine, where the system can direct the user to the optimal path via audio output.**

I. Introduction

According to the American Foundation for the Blind and the 2018 National Health Interview Survey, it is estimated that 32.2 million Americans older than the age of 18 are visually impaired [1][2]. The World Health Organization (WHO) estimates that globally over 2.2 billion people have a vision impairment and at least 1 billion of these cases could have been prevented or can still be corrected [3]. Visual impairment is a serious and common global issue that is commonly tied to socioeconomic factors as existing treatments can be expensive or difficult to obtain for most people.

In this study, we investigate methods and applications to improve the everyday life of visually impaired people. One of the main issues, when we are depraved of sight, is guidance from one point of location to the next point of location independent from others' assistance. It is possible to guide visually impaired people using computer vision techniques coupled with quick response (QR) codes, acting as beacon locators, and a database containing floor or building point of location information. Such a system would require a camera to capture images including QR codes and their surroundings, and a processing unit to analyze QR codes and captured images from different frames to correlate information from the database. Along with the monocular depth estimation algorithm implemented on the proposed system, it can provide the optimal path to the visually impaired user as well as provide collision avoidance capability.

II. System Design

The proposed system in this paper is divided into two main components: an indoor navigation guidance system for optimal path generation and a collision avoidance system using monocular depth estimation algorithms. The indoor navigation guidance system includes the QR code marker recognition application, the optimal path calculation application, and the guidance application. The collision avoidance system uses the monocular depth estimation algorithm to identify the average distance from the camera to the identified QR code markers to warn the user to prevent collision with the obstacle. Prior to the realization of this system, a database must be created that contains knowledge of the surrounding environment of the visually impaired user such as building floorplans. For every point of interest within the building, QR code markers will be placed on the wall to provide guidance information to the user including turning left or right, moving forward or backward, and the number of steps to proceed to the next point of location. The user with a camera on the front will walk inside the building and each time the camera sees a QR code, it will automatically scan it and read out the information to the user via audio output. Once the user provides the destination information to the proposed system, it will guide the visually impaired person automatically to where the next node (QR code marker) in the path is located.

A. Indoor Navigation Guidance System

This guidance system utilizes a database that contains the ID of each QR code marker location inside a building where the system is used for guidance. Each ID contains information about its location including the coordinates of the QR code marker (location beacon) inside the building and the possible other locations that the system can direct the user to proceed. Along with this information, inspired by Petri net diagrams [4], the system creates the shortest possible path to the destination. Once the user enters the building, the system downloads the full database of the building from the QR codes placed on the entrance doors of the building. This allows the system to work inside a building without an Internet connection.

The database is created based on the information collected from the QR code markers, which will include the marker's ID, next possible movement locations (nearby QR code markers), name of the location, and its current location in coordinates as shown in Fig. 1. Based on the created database, it is possible to determine each QR code marker's position and its neighboring nodes and to create a path to the desired destination.

ID	Possible_movements	fname	Coordinat
1	2	'Class 1'	[0,0]
2	[1,3,4]	'Class 2'	[2,0]
3	2	'Restroom'	[4,0]
4	[2,5,6,8]	'Hall'	[2,2]
5	4	'Elevator'	[0,2]
6	4	'Class 3'	[4,2]
7	8	'Class 4'	[0,4]
8	[4,7,9]	'Restaurant 1'	[2,4]
9	8	'Restaurant 2'	[4,4]

Fig. 1. Database Sample

The indoor navigation guidance system is designed with two different modes. First, the *normal mode* of the system identifies the QR code markers and retrieves necessary files from the online database to enable the system to operate offline. Each time a QR code marker is identified, the system will provide feedback to the user via audio output including the label of the node and the contextual information about the current location. Also, when different QR code markers are identified, the system can learn how the user proceeds from one marker to the next marker, providing path experience to the guidance system. Lastly, the monocular depth estimation function is called to calculate the distance from the camera to the QR code marker. Fig. 2 describes the overall system flow of the indoor navigation guidance system.

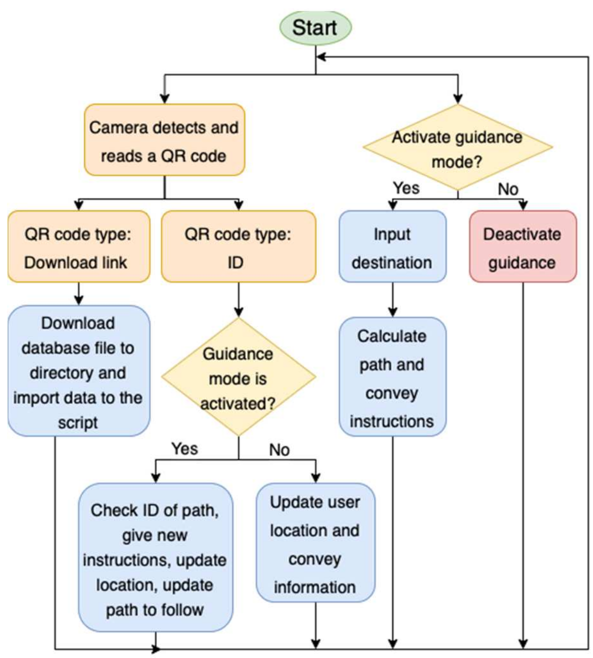


Fig. 2. Flowchart of Indoor Navigation Guidance System

Second, the *automatic guidance mode* is implemented on our system where it will provide the optimal (shortest) path possible to the user's choice of destination. When the user avoids the planned path, which can be recognized by the system if an unexpected QR code marker is identified, the system proceeds to recalculate the path and directs the user to the new path (see Fig. 3). If a correct QR code marker is identified along the planned path, the system will guide the user for the next movement directions (e.g., turn left/right, move forward/backward) including the distance information. Once the user arrives at the destination, the system exits the guidance mode automatically and turns back to the *normal mode*. All direction commands and information feedback are achieved by a text-to-speech module [5].

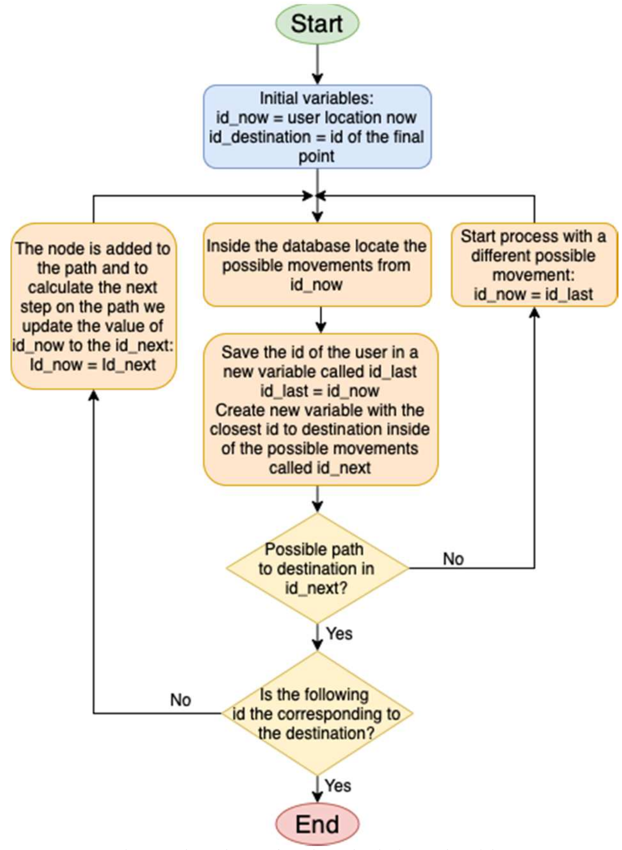
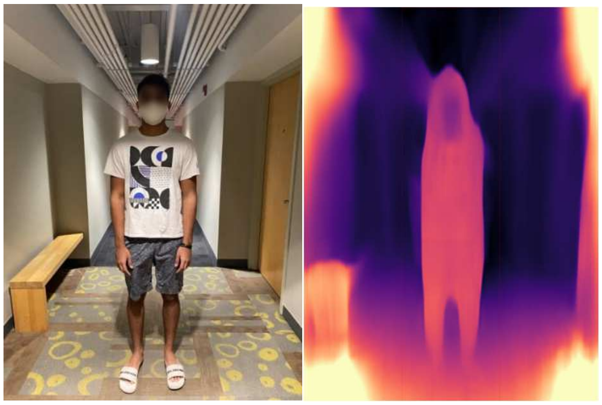


Fig. 3. Flowchart of Path Calculation Algorithm

B. *Monocular Depth Estimation System*

A monocular depth estimation algorithm is used to infer the scene geometry from the captured 2D images using computer vision techniques. The goal of the monocular depth estimation system in our design is to predict the depth of each pixel of the captured scene from a single input RGB image from the camera to determine the distance from the camera to the QR code marker. The output of the monocular depth estimation algorithm can be either a disparity map or a depth map depending on the technique used. In this study, we use a pre-trained mono and stereo models from Monodepth2 [6]. Monodepth2 is an implementation of a depth estimation model trained with KITTI dataset, consisting of 12,919 stereo images, whose output is a disparity map from a single RGB image [7].

In order to obtain the distance from the camera to the QR code marker, it is required to obtain the depth of an object shown in the captured image. Based on the output disparity map result, it is possible to convert it to a depth image map, using a given formula: $depth = \frac{baseline*focal}{disparity}$, where the baseline is the distance between the two cameras used for taking the KITTI dataset images (approximately 54 cm), and the focal length is the measured distance between the lens and the image sensor when focused to the object of interest. As our system utilizes captured image from the system's camera, the disparity value needs to be adjusted by normalizing it to the image width (in pixels of the image). The disparity value is calculated by the model for each pixel of an image and the formula can be applied to each of these pixels for obtaining a final depth map. Fig. 4a shows the original captured RGB image from the camera whereas Fig. 4b shows the converted image in a form of a disparity map.



(a) (b)

Fig. 4. (a) Captured Original RGB Image and (b) Converted Disparity Map

C. *Collision Avoidance System*

The collision detection system is designed to detect if there is a person or an obstacle on the path of the visually impaired user. In this study, not utilizing any conventional object detection algorithms, we solely use the captured image from the camera for the collision avoidance system. Our system divides the image into three vertical sections equally, assuming that the person walking towards the user, or the obstacle is visible in the center of the image as shown in Fig. 5. The system calculates the disparity maps from the captured image, and also calculates the average sum of every pixel disparity value in the region of interest created on the image (see Fig. 5). After performing experimental trial and error procedures, where a person would move towards the camera in different places, we obtained a threshold value to distinguish if the person was close or not. After calibrating the threshold of the average value of the disparity pixel values, the system was able to detect if a collision was about to happen, or if the path was free of obstacles or was far away.

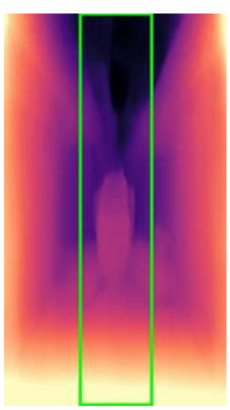


Fig. 5. Region of Interest (green box) for Detecting Obstacle

III. Implementation and Results

For demonstration purposes, the system was tested in a controlled environment with a typical USB camera and a laptop (Intel Core i9-8950HK, 32 GB RAM, Intel UHD Graphics 630). The database was created based on one of the floors of the building for demonstration (see Fig. 1) with multiple hallways and rooms. Each QR code marker was specifically located in corners and critical locations to assist guidance as well as doors and other important points of interest. Fig. 6 shows the sample path of the building where each QR code marker location is presented as an ID number, and the passage to each location is presented as a line between two markers. In our experiment, QR code marker ID 7 was the starting location for the user to reach the destination, QR code marker ID 3. Following the algorithm shown in Fig. 3, our system was able to find the optimal path to the destination as shown in Fig. 7 along with its calculated path to follow. Once the user approaches the QR code markers, the system will provide feedback via audio output including direction to the next node with the number of steps required, and information of arrival at the destination.

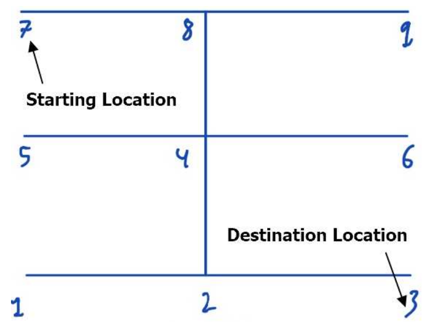


Fig. 6. Sample Building Path with QR Code Marker Locations

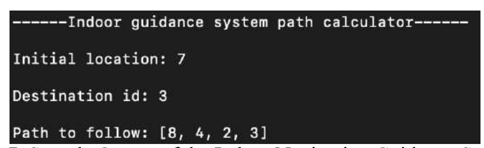


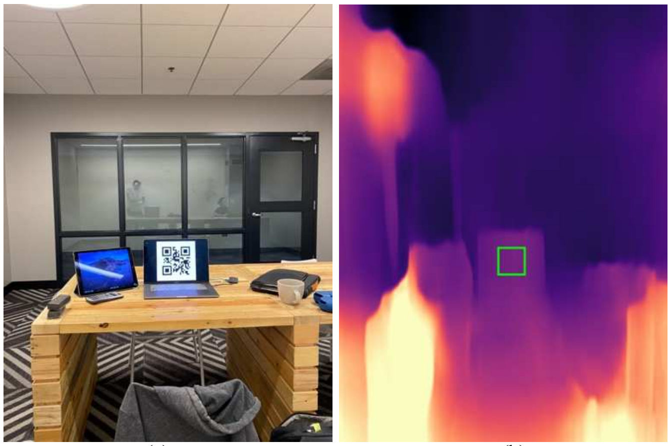
Fig. 7. Sample Output of the Indoor Navigation Guidance System

The collision avoidance system experiment was conducted by simulating a person walking through the same corridor and observing if the system correctly identified the risk and provided feedback to the user. Prior to implementing the system, the monocular depth estimation algorithm was tested by detecting QR code marker distance from the camera and comparing the measured distance and the estimated distance. As shown in Table I, the results obtained from the monocular depth estimation algorithm were within a margin of error compared to the measured distance. As the second experiment case (1.5 meters) has resulted in approximately 25 centimeters difference, it is still acceptable for the operation of the system.

TABLE I. Depth Estimation Test Results

Measured Distance (meter)	Estimated Distance (meter)
1	1.0174
1.5	1.2544
2	1.9289

Fig. 8 is a sample image taken using the attached USB camera to the system where Fig. 8a is a captured image containing a QR code marker in a complex background setting for image analysis, and Fig. 8b is the extracted depth map image with detected QR code marker. From Fig. 8b, the QR code marker was successfully identified and the distance estimation from the camera to the QR code marker was approximately 1.93 meters where the measured distance was 2 meters.



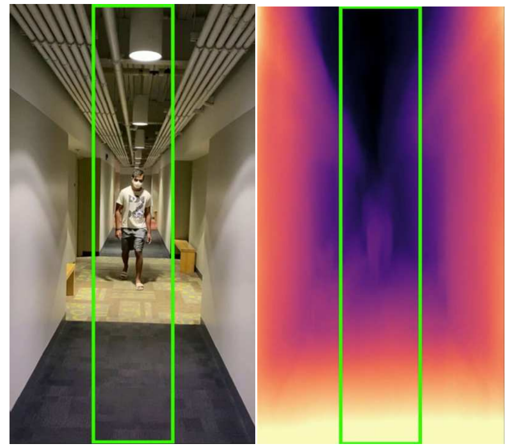
(a) (b)

Fig. 8. Sample Image for QR Code Marker Detection (a) RGB Sample Image, (b) Extracted Depth Map with Detected QR Code Marker

The collision avoidance system was also tested for the feasibility of the system when confronting any obstacle along the path to the destination. Fig. 9 and Fig. 10 illustrate recognition of the obstacle (person) identified inside the region of interest for detecting any obstacle in the captured image. Based on the threshold setting in our application, the system can identify if it is safe to proceed to the path or needs attention to avoid collisions. Fig. 9 is an example of when it is safe to proceed to the path as the estimated distance to the obstacle is not close enough to trigger a warning. However, as shown in Fig. 10, if the obstacle is within the threshold of the estimated distance, the system will generate an alert to the user, and it is shown as a red box in the extracted depth map.

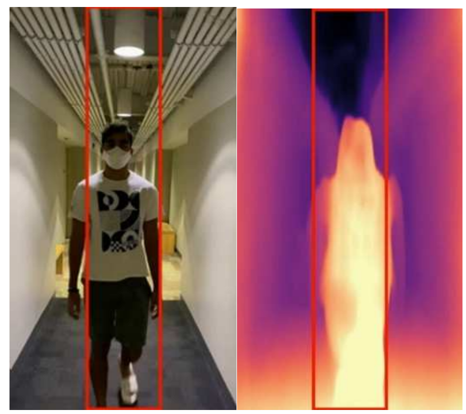
IV. Conclusion

In this paper, we explored a solution for an indoor navigation guidance system for visually impaired people by using QR code markers as point of location information and a collision avoidance system using monocular depth estimation algorithms. Our demonstrated results show that using a typical camera is feasible to detect obstacles in the path to the destination via artificial intelligence and computer vision techniques. Based on the collected information from the QR code markers, a database was created to contain all valuable information to guide the user and to create an optimal path to the desired point of location. By using a typical USB camera, our system identified the QR code markers placed on the walls and critical points of locations within the path. With the obtained image, our system utilizes pre-trained AI model to perform monocular depth estimation to detect the distance from the camera to the object on the path. Our system was able to accurately detect the distance to the obstacle and to warn the user to avoid collision with the obstacle or the person walking towards the user. Our system has the potential to be adopted to a smartphone or to a low-power AI accelerator where it can operate offline with the predetermined and up-to-date AI model stored in the device.



(a) (b)

Fig. 9. Sample Image for Collision Avoidance System (no warning) (a) RGB Sample Image of Obstacle Detected, (b) Extracted Depth Map



(a) (b)

Fig. 10. Sample Image for Collision Avoidance System (warning) (a) RGB Sample Image of Obstacle Detected, (b) Extracted Depth Map

References

[1] L. A., Blewett, J. A. Rivera Drew, M. L, King, and K. C. Williams, "IPUMS health surveys: National Health Interview Survey (Version 6.4)". 2019. https://doi.org/10.18128/D070.V6.4

[2] American Foundation for the Blind, "Blindness Statistics". [Online] Available: https://www.afb.org/research-and-initiatives/statistics/adults [Last Accessed: Mar. 6 2022]

[3] World Health Organization (WHO), "Vision impairment and blindness". [Online] Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment [Last Accessed: Mar 6. 2022]

[4] P. J. Fortier, H. E. Michel, "Petri Nets", *Computer Systems Performance Evaluation and Prediction*, pp. 279-303, 2003

[5] H. Jiang, T. Gonnot, W.-J. Yi and J. Saniie, "Computer vision and text recognition for assisting visually impaired people using Android smartphone," *2017 IEEE International Conference on Electro Information Technology (EIT)*, 2017, pp. 350-353

[6] GitHub - wkentaro/gdown. 2021. [Online] Available: https://github.com/wkentaro/gdown

[7] C. Godard, O. M. Aodha, M. Firman, G. Brostow, "Digging into Self-Supervised Monocular Depth Estimation", 2019. arXiv:1806.0