Integrated Machine Vision and Communication System for Blind Navigation and Guidance

Thomas Gonnot and Jafar Saniie
Department of Electrical and Computer Engineering
Illinois Institute of Technology, Chicago, Illinois, USA

Abstract — **This paper investigates methods and procedures to construct an efficient system to assist blinds in their everyday life. In particular, various technologies that can be utilized to build a wearable system are examined. The machine vision and the communication component of the blind navigation and guidance is designed not only to map the surroundings environment but also to determine a safe path to a desired destination. This work highlights the importance and also provides the instructions to blinds for efficient navigation and safe guidance by incorporating local traffic/pedestrian and alarming signs in real-time.**

I. INTRODUCTION

The issue of helping the blind people is not new. Several infrastructure adaptations show that significant consideration is made to facilitate the lives of the millions of people with visual disabilities in the United States of America (USA) alone [1]. This number grows up to about 285 million people of visually impaired around the world, from which 39 million are actually blind [2]. It is not feasible to equip the entire world with the infrastructures to help them. This limits their autonomy to a certain area, mostly to cities. For the areas that can be improved, the modification comes with a cost. In the USA, about 4 billion dollars per year is spent on infrastructure ameliorations, and on health care [3]. This includes, for example, the hiring of companions to help them in tasks which appear to be simple. The design of a device that helps them doing the simple tasks would greatly improve their independence. The system can help in their day to day lives by reading what they cannot see, prevent them from colliding with any objects, finding relevant information about their environment, and basically to navigate them safely to their destination. In the best case scenario, it would contribute for removing the need for an aid, and therefore reducing the expenditure in health care.

The system investigated in this system is to be worn by the user. It consists in a set of sensors coupled with a smartphone, and if necessary, one or more dedicated processing units. The objective is to minimize the blind person's dependency on any infrastructure modifications designed for them. Figure 1 shows an example of a blind navigation system, featuring one or two cameras integrated on the frame of standard eyeglasses. It also integrates earphones for directions and a vibration array to provide feedback. A dedicated computing unit implements all the algorithms used by the system and can be linked to the smartphone of the user [4].

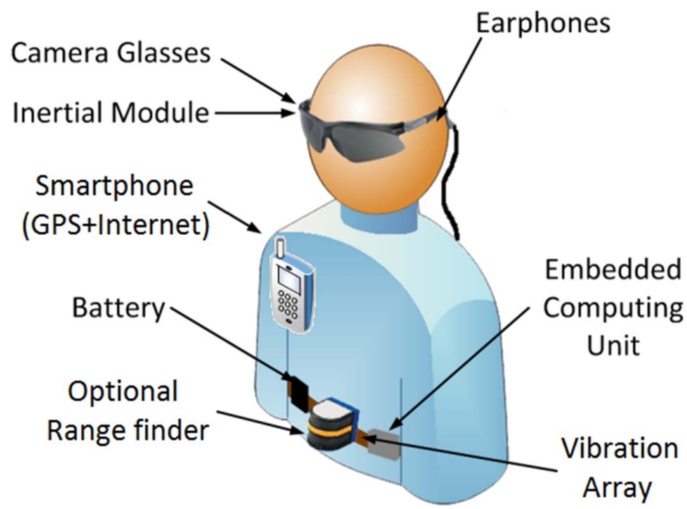

Figure 1. Example of a Blind Navigation System [5]

The system needs to address three different issues to assist the blind user. First, it needs to locate the user in the environment, using different methods involving the mapping of the areas. Furthermore, it needs to track the different elements of the scene to detect any danger, such as collisions. This will be discussed in section II. Secondly, the system needs to derive the trajectory to follow from the current position, the set destination and the map of the environment. It will be described under section III. It can also analyze the images from the camera to find clues about important information, such as traffic signs, as describe in section IV. It then needs to compile this information and provide guidance to the user. This leads to the final issue, the feedback to the user. The system needs to have a way not only to transmit the collected information to the user, but also to get user inputs, such as the destination to reach. This will be detailed in the section 0. The process is showed in Figure 2.

978-1-4673-9985-2/16/$31.00 ©2016 IEEE

II. Localization, Mapping and Tracking

One critical aspect of navigation is the ability to precisely compute the current position of the system. For a car, for example, the GPS provides an indication with a precision that can vary from one meter to several meters [6]. Even though this is enough for a car on a road, the precision is too low to efficiently guide a pedestrian securely. A difference of one meter could lead to the person walking on the road instead of the sidewalk. Of course, the use of a differential GPS would help increasing the resolution, however it requires a ground reference, and it does not help to locate the different obstacles in the environment. Instead, the GPS can be used to do a coarse location of the user to help other methods to refine this.

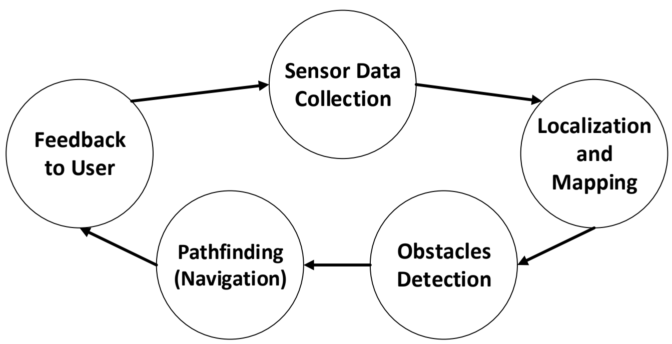

Figure 2. Process of navigating the blind

One type of algorithm that has proven quite reliable is the Simultaneous Localization and Mapping (SLAM) method. It consists in deducing the map of the environment from the previously collected information, and then correlates the present position using the generated map [7]. The mapping itself can be achieved using different types of sensors. One type of sensor is the laser ranging devices. This type of device projects a laser on the scene and analyses the projection to deduce its distance. Two-dimensional data can be obtained by repeating the procedure while rotating the sensor, and even three-dimensional data can be obtained by adding a second motion [8]. This can give depth information of the surrounding environment at a fast rate and with high precision, but is usually very expensive. Another way to get a depth map is to use cameras. This can be done in various different ways, and one of the most recognized methods is stereovision. Based on human vision, two cameras are placed side by side, with a known distance, and the depth map is deduced from the disparity between the views [9]. The drawbacks of this method are the use of very complex algorithm and the requirement for two cameras. Another method to extract the depth map is used by the Microsoft Kinect. In this method, instead of using two cameras to get the depth, an infrared structured light is projected in the scene, and an infrared camera measures the projection of the pattern to deduce the depth map [10]. Despite being very efficient and robust, it is limited in range, and due to the surrounding light intensity, it retrains to an indoor environment. The last method consists of using only one camera, and uses its motion to recover the necessary 3D information [11]. The advantage of such technique is that it just requires a single camera, without modification. However, it is highly computationally intensive and difficult to implement in a real time application.

Another method to refine the position of the system is the use of an Inertial Measurement Unit (IMU). It combines three axis accelerometer and three axis gyroscope to measure the motion of the system. Often used in embedded systems, its precision can vary depending on the technology used for the accelerometers and gyroscopes. The inexpensive devices that are commonly used in smartphones, small drones or other publicly available systems, are the MEMS sensors [12]. They are very small in size, but are less precise. Some systems also include a magnetometer to compensate the deviations of the gyroscopes using the earth's magnetic field as a reference. Even with this improvement, the precision of MEMS based IMU is very limited over time. Some SLAM research projects combine both the depth map generation and the IMU measurements to reduce the complexity of the algorithm [13]. The IMU provides a coarse matching of the current depth map with the memorized map, and then the system can perform a local brute force matching to locate itself more precisely, instead of doing a brute force matching on the entire map.

The advantage of using SLAM to deduce the location of the system is that it makes it possible to generate a map of the environment. This map can be compared to a previously generated map to determine new elements, such as new obstacles, or more importantly moving obstacles. A tracking algorithm can be added to the SLAM to detect potential collisions [14]. Figure 3 shows the organization of the localization and mapping algorithm.

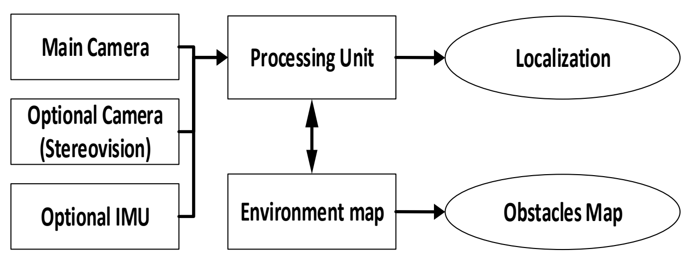

Figure 3. Organization of the localization and mapping algorithm

More exotic means of detection can also be implemented. For example, the recent development in thermal imaging has made it possible to use a thermal camera on embedded systems. This particular type of sensor can alert the person, for example, if the person tries to touch a dangerously hot surface. In the case of recognizing people, an algorithm tracking faces can be confused by a picture of a person, whereas adding the thermal imaging in the process would help making the distinction between a picture and an actual person. Yet another technology is the use of ultrasonic range finders. Generally used as unidirectional sensors, they could help avoiding collision, and can be used as an economic alternative to the laser range-finding devices.

III. Navigation and Assistance

Navigation devices are fairly common nowadays. Starting with dedicated GPS devices for cars, the smartphone took over and now proposes GPS navigation for cars, as well as for pedestrians. Many years of development and use of these systems have helped getting several reliable pathfinding algorithms. The current research is orienting on indoor navigation, searching for alternative localization techniques for using in the phones [15]. This could be very valuable to the blind, but the quantity of information available to the blind is not enough. When a non-blind person wants direction in a building, it requires only simple directions such as which corridor to take turn. A blind person requires more directions, including distances, indications to avoid obstacles on the way or something as simple as the number of steps on a staircase.

When navigating outside, the maps should also include much more information. A non-blind person will probably find it easy managing at a cross-walk, or a sidewalk, but this task is very difficult for the blind. Consequently, a map designed for the blind should include very precise position of roads, sidewalks, crosswalks, and also the ground level changes and light posts, or any permanent obstacles.

One solution for the map to stay updated is to constantly map the areas used, and also to share these maps between the users. The smartphones provide this possibility due to their data connection feature. This way, the shared maps can also be complemented by external data sources. For example, a city would be able to provide the information about all the traffic lights at any given time, to allow the user to know when to cross, and also give the positions of any point of interest. If for some reason, the mapping partially fails, the system can rely on this information to provide reliable indications for the navigation. Moreover, in a situation where the automation completely fails, the smartphone can be used to contact an assistant, accessing the current information about the environment and providing directions. Figure 4 shows the organization of the navigation system.

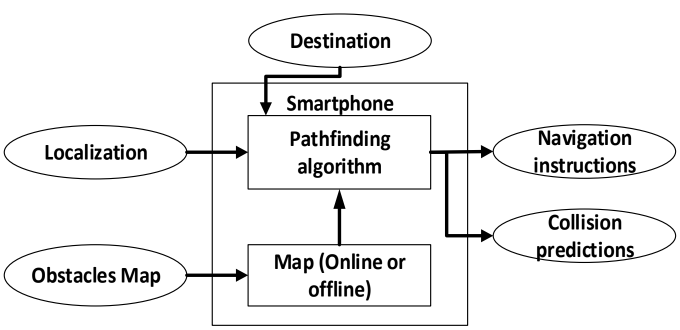

Figure 4. Organization of the navigation system

IV. Environment Analysis

Despite navigation, the system can also help the user to detect relevant information, such as the indications on pedestrian signs.

Figure 5 shows an example of situation where recognizing the pedestrian signs can help the navigation and avoid a potential dangerous situation.

Figure 5. Example of complex situation and relevant information

This doesn't have to be limited to pedestrian signs, it can also be used for anything else, such as reading the text on an object help in front of the camera, or the room of apartment numbers in a building. The diversity of signs that could be recognized requires the use of a database, storing the different signs. However, an offline system can't store all the different signs by itself, and should be connected to an online database capable of matching the signs with greater efficiency and with a larger signs collection. Figure 6 shows the design flow of the sign recognition system, including both offline and online databases and using a smartphone for the Internet connectivity. The system also includes an Optical Character Recognition (OCR) to interpret the content of the different signs, which can also be made online if necessary.

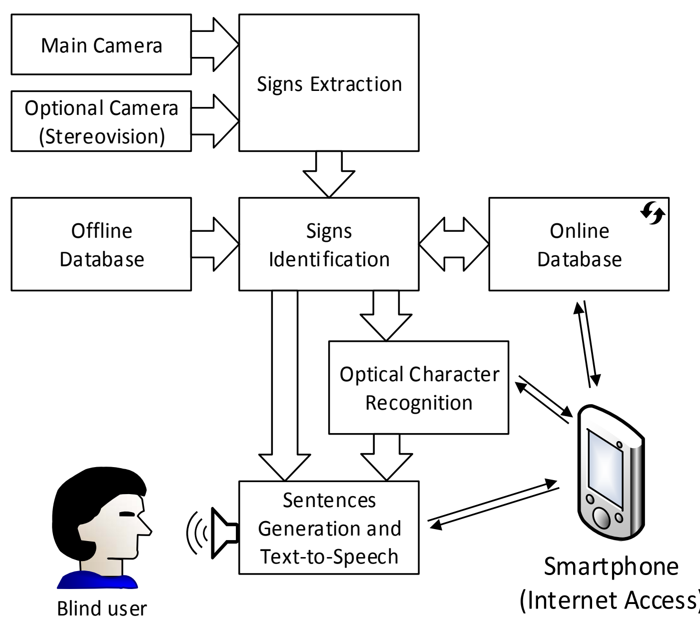

Figure 6. Sign recognition system design flow for the blinds

V. INTERFACE AND INTERACTIONS

The system is useless without a comprehensible way to transmit the information to the user. Therefore, it needs to integrate one or multiple systems to communicate with the user. One very common way to do this is the Text-to-Speech. The system generates sentences or words describing the situation, and an algorithm converts it to voice. It can also be used in conjunction with a speech recognition engine to process commands or answers from the user. Most of the currently available smartphones implement an online speech recognition engine, and Android even proposes one that remains available offline. This is the most natural way people can communicate with the system. The technology is so advanced nowadays that someone with a smartphone can capture a text with a smartphone, and gets a translation directly on the image. This system requires the capabilities to detect and read the text. In the situation where the person present a text in front of the camera, the system could interpret it as the need to read it, and so it would be able to do so and use text-to-speech to dictate it. This not only removes the need for braille, but also allows the person to get any text read to him.

Despite the intuitive nature of the verbal communication, it can be quite slow when describing the surrounding. Therefore, another method can be used in conjunction to allow more information to be available. For example, when using the sound, the system can generate a set of tones representing different directions, and their intensity can represent the distance of the closest object in that direction. The blind people usually compensate their lack of vision by increasing the sensitivity of their other senses, which can be used as guidance. Another solution is to take into account the geometry of the ears in order to make the person feel like the sound comes from a certain direction. This principle is the base of 3D surround systems applied to stereo headphones of earphones. This make it more natural since the brain already interprets the distance and the orientation depending on the information it receives on both ears.

Another possible feedback method takes advantage of the haptic sensitivity of the blind people. Vibration arrays can be used on the body, for example on a belt, increasing the vibration as an obstacle comes by on a specific direction. The array can also be a two-dimensional, representing the obstacles as they appear in front of the user. Electrodes can also be used, on more sensitive parts of the body, such as the tongue. Figure 7 shows the organization of the feedback mechanism.

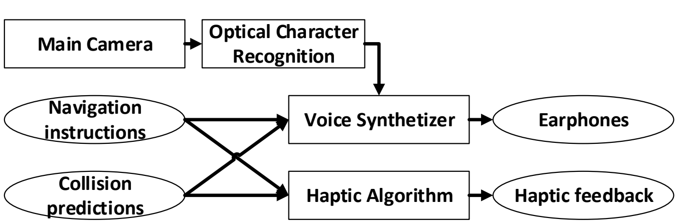

Figure 7. Organization of the feedback mechanism

In the case of special objects, such as doors, stairs, signs or tools, the system can be programmed to send specific patterns to signal the user of special actions. In the case of the voice, it could indicate the presence of directions when it recognizes signs on a wall in a corridor.

One last way to exchange information between the system and the user is by selecting the data that will be communicated. A specific action from the user can trigger a certain type of information to be highlighted. The system can be following the movement of the hand of the user, and when pointing at an object, the system would try to interpret whatever is in the vicinity of the pointed area. If we take the previous example in front of signs giving directions, pointing at it would trigger the system to enumerate the different elements of the sign.

VI. CONCLUSION

In this paper, we presented an assembly of methods and technologies for the design of a navigation system for the blinds. It features the use of different sensors to recover the map of the environment as well as the precise localization of the system in order to properly guide the user. It also proposes the use of a smartphone to manage the navigation and the interaction with the user using either audio or haptic feedback, or both.

REFERENCES

[1] "Blindness Statistics," National Federation of the Blind, [Online]. Available: https://nfb.org/blindness-statistics. [Accessed February 2015].

[2] "Prevalence of Vision Impairment," Lighthouse International, [Online]. Available: http://www.lighthouse.org/research/statistics-on-vision-impairment/prevalence-of-vision-impairment/. [Accessed February 2015].

[3] "Blindness and Vision Impairment," Center for Disease Control and Prevention, [Online]. Available: http://www.cdc.gov/healthcommunication/toolstemplates/entertainmented/tips/blindness.html. [Accessed February 2015].

[4] T. Gonnot and J. Saniie, "Embedded and modular video processing design platform," *2014 IEEE International Conference on Electro/Information Technology (EIT),* pp. 290-293, 2014.

[5] T. Gonnot and J. Saniie, "Image sensing system for navigating visually impaired people," *2013 IEEE SENSORS,* pp. 1-4, 2013.

[6] "GPS Standard Positioning Service (SPS) Performance Standard," Office of the Secretary of Defense, September 2008. [Online]. Available: http://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf.

[7] G. Tuna, K. Gulez, V. Gungor and T. Veli Mumcu, "Evaluations of different Simultaneous Localization and Mapping (SLAM) algorithms," *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society,* pp. 2693-2698, 2012.

[8] M. Kurisu, Y. Yokokohji and Y. Oosato, "Development of a laser range finder for 3D map-building in rubble," *Mechatronics and Automation, 2005 IEEE International Conference,* vol. 4, pp. 1842-1847, 2005.

[9] K. Schreve, "How accurate can a stereovision measurement be?," *2014 15th International Workshop on Research and Education in Mechatronics (REM),* pp. 1-7, 2014.

[10] J. Han, L. Shao, D. Xu and J. Shotton, "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review," *IEEE Transactions on Cybernetics,* vol. 43, no. 5, pp. 1318-1334, 2013.

[11] A. Davison, I. Reid, N. Molton and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 6, pp. 1052-1067, 2007.

[12] K. Abdulrahim, C. Hide, T. Moore and C. Hill, "Aiding MEMS IMU with building heading for indoor pedestrian navigation," *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010,* pp. 1-6, 2010.

[13] G. Panahandeh and M. Jansson, "Vision-Aided Inertial Navigation Using Planar Terrain Features," *2011 First International Conference on Robot, Vision and Signal Processing (RVSP),* pp. 287-291, 2011.

[14] T. Gonnot and J. Saniie, "Design flow for complex scene visual object tracking and avoidance system," *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),* pp. 1408-1411, 2013.

[15] M. Boysen, C. de Haas, H. Lu, X. Xie and A. Pilvinyte, "Constructing indoor navigation systems from digital building information," *2014 IEEE 30th International Conference on Data Engineering (ICDE),* pp. 1194-1197, 2014.