Embedded Navigation and Classification System for Assisting Visually Impaired People

Antonio Miguel Batista Dourado[1,2] and Emerson Carlos Pedrino[1]
[1] *Department of Computing, Federal University of São Carlos, Rod. Washington Luís - Km 235, São Carlos-SP, Brazil*
[2] *Federal Institute of São Paulo - Boituva Campus, Av. Zélia de Lima Rosa - 100, Boituva-SP, Brazil*

Keywords: Visually Impaired People, Electronic Travel Aids, Computer Vision System, Navigation System, Object Recognition, Object Classification.

Abstract: Loss of vision has a large detrimental impact on a person's mobility. Every day, visually impaired people (VIPs) face various challenges just to get around in the most diverse environments. Technological solutions, called Electronic Travel Aids, help a VIP with these challenges, giving greater confidence in the task of getting around in unfamiliar surroundings. Thus, this article presents an embedded navigation and classification system for helping VIPs indoors. Using stereo vision, the system is able to detect obstacles and choose safe ways for the VIP to walk around without colliding. A convolutional neural network using a graphics processing unit (GPU) classifies the obstacles. Acoustic feedback is transmitted to the VIP. The article also features a wearable prototype, to which the system hardware is docked for use. Using the system, the prototype could detect and classify obstacles in real time defining free paths, all with battery autonomy of about 6 hours.

1 INTRODUCTION

Electronic Travel Aids (ETAs) have gained prominence in the last decade in the area of visual impairment. The World Health Organization (WHO, 2014) reported that there are at least 285 million visually impaired people (VIPs), considering partial and total loss of vision, thus making ETAs important tools that can be built into the day-to-day life of a visually disabled person.

In terms of hardware, ETAs can be developed in different ways depending mainly on the type of input sensors and how information is transmitted to the visually impaired person (VIP). Input sensors commonly found in the literature are GPS cameras, IMUs, RFID readers, infrared lasers and others (Fajarnes et al., 2010; Katz et al., 2012; Mehta et al., 2011; Tapu et al., 2016). Each input sensor can be used alone or in conjunction with others to provide safety information to the user via acoustic audio feedback (Schauerte et al., 2012), where sounds represent the information, or the user receives some other physical stimulus, such as vibrations (Bourbakis, 2008).

A device to help safe navigation for a VIP must, however, allow the user good mobility, in terms of the weight and size of the device (Pissaloux, 2002). This premise makes it unfeasible that an ETA device has a PC or a notebook as the processor, since the former is not mobile, (although some PC based systems use cloud processing), and the latter, while offering a certain degree of mobility, still tends to be quite heavy. One possible solution for local processing is to make an ETA with an embedded system, albeit with more limited processing power.

The segmentation of free paths and classification of obstacles in embedded platforms have been partially explored in some studies, as seen in Section 2, however there is no approach that considers presenting multiple paths to the user, nor any work using recent object classification techniques using deep neural networks with graphics processing units (GPUs).

This paper presents an embedded ETA that helps the VIP with navigation, through the recognition and classification of indoor obstacles, providing acoustic feedback. The system was built in an NVIDIA Jetson Tegra X1 module (NVIDIA, 2017) that has a CUDA 256-core video card for parallel processing and offers high performance with low power consumption. The images are captured through the sensors of an RGB-D camera and processed. The use of a RGB-D camera imposes an indoor environment limitation, knowing that external environments have a high incidence of infrared light which may interfere with the camera's image capture. For the classification of obstacles, the

Dourado, A. and Pedrino, E.
Embedded Navigation and Classification System for Assisting Visually Impaired People.
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, pages 516-523
ISBN: 978-989-758-290-5

ETA will use Convolutional Neural Networks (CNN), a category of Deep Neural Network that has produced great results since its first appearance in 2012 in the ILSVRC competition (Krizhevsky et al., 2012).

The article is divided as follows: Section 2 presents studies related to this one that have been reported in the literature. In Section 3, the proposed system is described along with the architecture in which it was developed. Section 4 describes the developed prototype and the experiments performed using the system and the results are shown in section 5. Finally, in Section 6, the conclusions and future work are given.

2 RELATED WORK

The literature about recognition and classification navigational systems for VIPs describes several different approaches to address the issues of safe paths and the classification of obstacles.

Mocanu et al. (2015) used the classic features extraction technique to classify obstacles with an adaptation of the Histogram of Oriented Gradients (HOG) and they represented images globally using techniques like the Vector of Locally Aggregated Descriptors (VLAD) and the Bag of Visual Word (BoVW) framework. The images are ranked in conjunction with Support Vector Machines (SVM). The implementation is done on a smartphone and the processing reaches a maximum of 5 fps. It is important to emphasize that, although 5 fps is an acceptable performance for VIPs, the authors did not suggest safe paths and they do not make it clear how the VIP is informed about the obstacles.

The idea proposed by Deb et al. (2013) aims predominantly at the safe navigation of the VIP by estimating the safe direction in which they can walk. In the project, authors use a simple camera for image acquisition and apply techniques such as edge detection and pyramidal segmentation to then define a safe zone using Template Matching and transmit via musical tones whether the safe path is to the right or left. The area analyzed is center-bottom of the image. So the VIP is not alerted to possible (and unclassified) obstacles above the ground and the proposed system disregards much of the left and right sides of the image which, in some cases, could have alternative routes. Finally, the authors also state that, in the case of two safe paths, the system will always choose the left side and not even present the alternative to the VIP periodically.

The use of embedded platforms for VIPs was explored by Bangar et al. (2013), using recognition of objects and their colors. In their research, the authors extract the images from a coupled video sensor in a pair of glasses and apply procedures like background extraction, edge detection and pixel clustering to define an object and also its color, using a standard color scheme. The system then passes the result of processing to the VIP via stereo sound. Color detection after object detection is an important factor for classification, but the authors do not classify the object completely and, although mentioning stereo sound, do not mention in their work if the audio is used in a way that transmits the direction of the object. The work also does not aim for or try to explain the segmentation of free paths.

Poggi and Mattoccia (2016) also use embedded technology for the implementation of an intelligent system for VIPs. In their work, the authors use an Odroid U3 device in communication with a smartphone, headphones, tactile feedback glove, and deep camera goggles as an embedded platform and the system classifies obstacles using convolutional neural networks, but without the acceleration of a Graphics Processing Unit (GPU). There is a navigation stage but only via the GPS of the smartphone and the system proposed by the authors does not define safe paths, only dealing with the classification of obstacles and their positioning and informing the user of the GPS directions.

3 PROPOSED SYSTEM

The system developed here has been divided into five main stages: image acquisition, preprocessing, free path segmentation, obstacle detection and acoustic feedback as shown in Figure 1. Image acquisition is a simple and obvious process in which two 640 x 480 images are received from the RGB-D sensor, one depth and one color.

Preprocessing only happens for the depth image. The original depth image captured by the RGB-D sensor contains several faults with indefinite depths, generating chromatic irregularities in the depth image. The preprocessing corrects such flaws using mathematical morphology. The image is dilated and then reduced again to the same 21 by 21 element square structure. Then it's submitted to a free path segmentation, obstacle detection and acoustic feedback stages, which are presented in the subsections below.

3.1 Free Path Segmentation

To determine one or more safe paths for the user, the RGB-D camera depth image has been divided

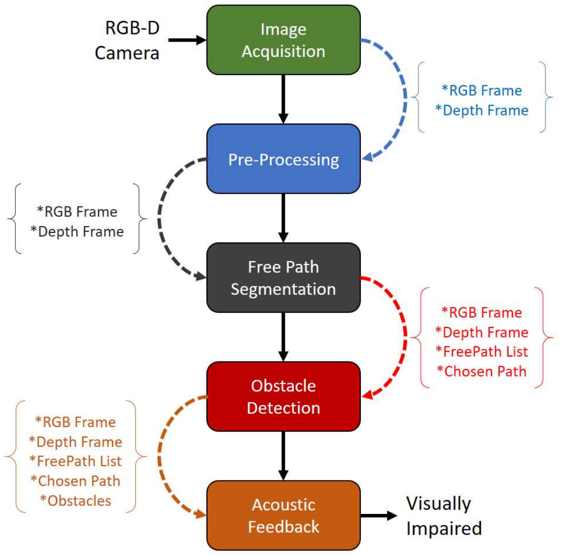

Figure 1: System Overview.

into four horizontal parts and, in this stage only, the bottom of the image has been taken, i.e. the bottom 25% of the image is processed. The reason for this is that the VIPs will carry the camera in their belly region during the use of the prototype and, considering the field of vision of the camera, the lowest quarter of the image contains more elements that limit the navigation of the VIP. The top 75% of the image, however, is not discarded but processed for obstacle detection after the indicated safe path has been chosen. Figure 2 shows the selected lowest area of the color and depth images but, as mentioned, only the depth image is used for path segmentation.

Figure 2: Splitting process. RGB Image (left) and Depth Image (right).

When the lowest 25% of the image is separated, the Template Matching technique (Brunelli, 2009) is applied to define which areas of the image are safe for navigation. Before the template matching is done, a simple thresholding operation is done on the bottom slice of the image to separate the possible free paths from those with obstacles. To define the free paths, a template, 50 pixels wide by the height of the bottom slice, performs the template matching operation throughout the length of the slice, as shown in Figure 3. Regions next to free paths are joined together and considered as one region and its centroids added to a list of coordinates that, in the end, is used by the system to give the direction to the VIP.

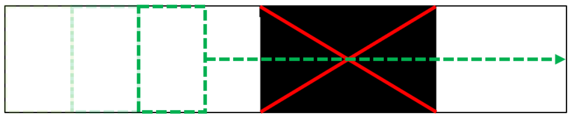

Figure 3: Sliding template (Green) for free path segmentation. Dark areas indicate obstacles.

If more than one path is separated, the options will be sent via acoustic feedback and then the path closest to the center of the image will be chosen. The system considers five possible directions of orientation for VIPs in the free path segmentation: right, slightly right, front, slightly left and left (Figure 4). Each region corresponds to 20% of the image so 'slightly left' and 'slightly right' mean the user should turn slightly in the specified direction, while the right and left directions define the user should turn between 30° and 45° in that direction.

Figure 4: Directions defined by image regions.

Algorithm 1 contains the pseudocode for the free path segmentation stage. The code determines one or more free paths and chooses the closest to the center, if any. A list with the X coordinates of the centers of the free paths and also the path chosen are those resulting from the code.

With one free path determined, the system may then look at verifying the existence of an obstacle in that direction.

3.2 Obstacle Detection and Classification

Although free path segmentation occurs in the bottom 25% of the image, the detection and classification of obstacles occurs in the upper 75% of the image as shown in Figure 5. However, this stage only occurs if and when a free path is chosen, since obstacle checking occurs only in the direction of such a chosen path. The areas of the image relative to any not-selected paths are ignored and not processed in search of obstacles. This search for obstacles in the upper part of the image is because several objects, such as an open drawer or a cupboard door, could block the space above the free path and prevent the user from walking forward.

Algorithm 1: Free Path Segmentation Algorithm

```
Data: DepthFrame
Result: List of Free Paths, Chosen Path
bottomSlice = imageYSplit(DepthFrame,4)[4]
bottomSlice = threshold(bottomSlice,
  distanceThreshold)
template = Rectangle(50,bottomSlice.height)
while Sliding template through bottomSlice do
    if template matches then
        if adjacent region then
            updateLastFreePath(freePaths,
              XCenterCoordinate)
        else
            freePaths = XCenterCoordinate
        end
    end
end
if freePaths >0 then
    if freePaths >1 then
        sendAcousticFeedback(freePaths)
    end
    chosenPath = pickCentralPath(freePaths)
else
    sendAcousticFeedback("No path found.")
end
```

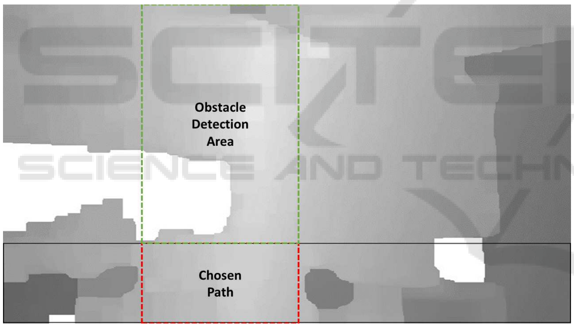

Figure 5: Obstacle detection area (green).

The upper slice of the image, as well as the lower one, is also subjected to a thresholding operation (with the same value as the lower slice) to delimit closer obstacles, and then a new pixel operation defines, if any, the obstacle closest to the defined area. Any obstacle detected is defined by a Canny edge detector, its contours drawn and a bounding box is created for that obstacle. For best results, the bounding box area is increased by 25% to avoid cropping the borders of the obstacle.

Obstacle classification uses a convolutional neural network, using the Caffe framework (Jia et al., 2014). A 22-layer GoogLeNet (developed by Google) model (Szegedy et al., 2015) was trained to classify twenty classes of objects: animal, cabinet, vacuum cleaner, bag, chair, bed, basket, stove, refrigerator, window, table, backpack, person, sink, door, sofa, monitor and toilet. About 60,000 images were used for the training, reaching 82% classification accuracy for Top-1 prediction and 95% for Top-5 prediction (the correct class is one of the top five ranked classes) after 35 epochs. Although many of the trained objects are not usually detected in the upper 75% of the image, especially in the experiments of this study, the model will also be used for a future study mentioned in Section 6. However, the model is able to classify hanging backpacks and table corners, for example. In cases where the obstacle cannot be classified, the system will emit a warning sound about the obstacle but not identifying it.

When an obstacle has been detected and its region determined, the defined bounding box is used to extract the sub-image from the RGB image containing the obstacle. This sub-image is sent to the classifier which returns a list of the five most likely classes of the image and their accuracy percentages. The confidence of the classification of the obstacle is divided into three bands according to the percentage of accuracy:

$$Conf(img) = \begin{cases} \mathbf{100\%}, & \text{if } Predict(img)_1 \geq 90\% \\ \mathbf{50\%}, & \text{if } Predict(img)_1 \geq 60\% \\ & \text{and } Predict(img)_2 \leq 30\% \\ \mathbf{0\%}, & otherwise \end{cases}$$

The confidence defined by *Conf(img)* directly informs the feedback that the VIP will receive from the system, as shown in Section 3.3, , and ranges from total certainty that the obstacle is of a given class to unable to identify the obstacle. With the information provided, the function $Predic(img)_x$ is called to classify the image (img) and its index ($\boldsymbol{x}$) is the position of the result in the list, with 1 being the most accurate, 2 the second most accurate and so on. Algorithm 2 shows the pseudocode for the detection and classification of an obstacle.

3.3 Acoustic Feedback

Acoustic feedback implies in processed system information of interest to the user passed to them via sound. The system developed here performs two types of acoustic feedback: via voice, in which feedback consists of words informing the user about a free path or an obstacle; and via tones, in case a detected obstacle cannot be correctly classified or only classified with low accuracy. Thus the types of feedback that the system can provide are:

- **Directions** (voice), informing the direction in which the defined free path is. In this case the phrases were defined as: "left", "slightly left", "front", "slightly right" and "right";

Algorithm 2: Obstacle Detection and Classification Algorithm

```
Data: DepthFrame, RGBFrame, ChosenPath
Result: Acoustic Feedback
upperSlice = imageXSplit(DepthFrame, ChosenPath)
obstacles = threshold(upperSlice, distanceThreshold)
obstacle = getNearestObstacle(obstacles)
if nearestObstacle is not null then
    obstacleImg = cropImage(RGBFrame, obstacle)
    obstacleClass = classify(obstacleImg)
    obsInfo = {obstacle, obstacleClass}
    sendAcousticFeedback(chosenPath, obsInfo,
      RGBFrame, DepthFrame)
else
    sendAcousticFeedback(chosenPath, NULL,
      RGBFrame, DepthFrame)
end
```

- **No direction** (voice), when the system is not able to find a safe path, the sentence informed the user is ”No path”;
- **Obstacle class** (voice), identifying the obstacle after its classification only by its class and position via 3D audio. Two types of messages can be sent to the user. If the prediction reliability is 100%, the system will simply send the obstacle class as feedback. For 50% reliability, the system will report that the obstacle appears to be of the informed class, by means of the sentence ”It appears to be <class>”;
- **Unclassified obstacle** (tone), if the obstacle can not be classified, will be issued a tone towards the obstacle, rather than say the class of the obstacle.

In order to preserve the contact of VIPs with the sounds of the environment around them, this design uses a non-invasive bone-conducting stereo headset that allows both system feedback and ambient sounds to be heard simultaneously. The system also converts the sounds of obstacle positions into 3D positioning, indicating by sound if an obstacle is to the right, left or center of the path. The text-to-speech synthesis of the above phrases was done with the *eSpeak* library (ESpeak, 2007) and reproduced by the OpenAL library (Hiebert et al., 2017), which was also used reproducing the tone of an unclassified obstacle.

4 EXPERIMENT

The proposed system aims at improving mobility of VIPs and their relative comfort without detriment to the system performance. So, before experimenting with the system, a functional, wearable prototype for navigating had to be developed.

4.1 Wearable Prototype

The prototype hardware seeks to minimize the impact on both the movement and also the hearing of VIPs, avoiding their total sensory immersion in the system. The prototype consists of: a Kinect for XBOX 360 RGB-D camera, an NVIDIA Jetson TX1 board, Aftershokz Sportz M2 bone conduction headphones with built-in microphone, a LiPo 3S 2.200MAh battery to power the RGB-D camera and a 2.2200MAh 4S LiPo battery as the Jetson TX1 power source.

The Kinect camera is surrounded by a fabric cover. This cover is used to attach the camera to a special dress for VIPs tailored from a tactical (airsoft) vest. The vest is used not only to attach the camera to the front, but also to attach to three other compartments on the back that carry both batteries and the Jetson TX1 card. Figure 6 shows a person wearing the prototype vest with the attached equipment.

Figure 6: Prototype equipment for developed system. Front (left) and back (right).

4.2 Practical Experiment

The prototype was tested in the corridors of the department to which the project is linked. Although a partnership with a non-governmental organization (NGO) for VIPs already exists, the first testing was done with a fully-sighted person and a future version of the system will be tested by the VIP members of the NGO.

The experiment consisted of following several paths through the corridors, shown in Figure 7, where the green dot indicates the beginning of the route, the red dots indicate the end and each line linking them indicates a path. This sought to check if the system would prevent collisions between the participant and any object, wall or person, in addition to indicating the free paths including the most central. The system was configured and tested for detection of free paths from three maximum distances: 80cm, 120cm and 160cm. All the test routes were followed for each of these three configurations.

None of the surroundings in the department were adapted for the experiments nor was the movement of people restricted at any time the routes were being followed, in order to encourage chance encounters with people and the response of the system to such an event.

During the experiment, the authors also looked at the questions of the speed and time of the feedback in relation to the user's reaction, i.e. if the time spent sending the feedback is adequate so that the user can, for example, avoid bumping into a wall. The classification of obstacles was also checked during the experiment, including its impact on the performance of the system.

5 RESULTS

For all the routes followed, the system was able to help with safe navigation, since there were no collisions with walls, people or obstacles in general.

Each maximum distance produced different system behavior. The 80cm distance could inform the user in a timely manner about changing a path or obstacle but the authors observed that the user reaction time and the speed of their steps can be a problem if they walk fast and have slower reactions. A maximum distance of 150cm was enough to detect obstacles and we believe that a slow reaction from user for system feedback is not a problem for such distance. However, there was an issue found for this maximum distance, when moving in narrow places where any minimal change of user angle (i.e.: user slightly turning right) potentially generates a free path feedback in the opposite direction to that minimal turn. Finally, the ideal maximum distance tested was 120 cm. This distance maintains user safety even with a possible low post-feedback reaction time. It still gives the VIP freedom to choose their direction and there is no need for constant slight adjustments as with the 160cm maximum. Figure 8 shows the free path detected by the system in red.

Figure 8: Path chosen by the system (red).

In scenarios with multiple paths, the system detected each of the paths and chose the one closest to the center to indicate as the safe path. In this case, the system took up to three seconds to transmit all the directions to the user but there was no impact on its reliability, considering the decision time and reaction time that the user has to carry on in or change to the direction that the system chose. Multiple path detection is shown in Figure 9. The system also de-

Figure 7: Department map with routes followed.

tected false paths when there was a partial or full glass door. These false paths are explained by the fact that the glass in the door does not reflect the infrared emitted by the RGB-D camera, an event which the authors already expected. Although the total direction reproduction time was up to three seconds, the frame rate (fps) was 15fps, on average, after subtracting the feedback time.

Figure 9: Multiple path indications. Red indicates the chosen path.

Obstacle detection by the system went as expected and all the obstacles that appeared in the upper portion of the chosen free path were properly detected. The obstacle classification was also tested in the experimental stage, but to a lesser extent than the free path segmentation. The classifier was able to correctly classify people, tables, chairs, and computer monitors. However, there is an important limitation that the authors found in the experiments: Partial images of obstacles are difficult to classify and generally lead the system to only emit a tone indicating lack of accuracy. The results therefore show the need for a classification approach that considers partial images of an obstacle, such as an arm being classified as a person. The performance of the classification was satisfactory, on average 240 milliseconds to predict the class of an obstacle, without optimizations.

Finally, the wearable prototype was shown to be an alternative that works but the authors will study other designs, since clothing tends to be hot and cumbersome after a long period of use. The total distributed weight of the prototype is about 1 kg, and this was not a problem during the experiment. The measured energy autonomy of the prototype is about 6 hours of continuous use, considering both batteries.

6 CONCLUSIONS AND FUTURE WORK

This article presented an embedded navigation and classification system that can assist VIPs in their daily lives, as well as a wearable prototype that includes such an embedded system. Table 1 compares the differences between the system proposed here and the approaches presented in Section 2.

The system developed here showed itself totally capable of indicating free paths so that a VIP does not bump into obstacles such as walls, tables, chairs, etc. in addition to detecting multiple path choices. This is the first step in a larger project of which this system is part and the next step is to include simultaneous localization and map (SLAM) techniques (Leonard and Durrant-Whyte, 1991) so that a VIP can be guided to a specific location in an environment. The detection and classification of obstacles achieved the expected results but new classes of objects must be added to the current model, in order to contemplate other objects commonly found in closed public environments. As previously mentioned, the model used in this project was also trained for a future project with classification of objects in a residential environment (see the list Section 3.2 in above).

Soon, the prototype will be tested in an experiment involving VIPs, who will be able to offer their feedback on the experience with the system and prototype. The implementation of a voice interface for direct interaction between the user and the system is also planned.

ACKNOWLEDGEMENTS

We would like to thank the Federal Institute of São Paulo, the Federal University of São Carlos (UFSCar), the Brazilian NGO ‘PARA-D.V’ (non-profit organization for the inclusion of VIPs). And also CNPq, (Project No. 2015/23297-4), for their immense support.

Table 1: Comparison between our proposed system and related works presented in Section 2.

	Embedded	GPU	Multipath Detection	Multipath Alert	Obstacle Detection	Obstacle Classification	Stereo Vision	Acoustic Feedback	3D Audio Feedback
Our	x	x	x	x	x	x	x	x	x
Mocanu	x				x	x		x	
Deb			x		x				
Bangar	x				x	x		x	x
Poggi	x				x	x	x	x	

REFERENCES

Bangar, S., Narkhede, P., and Paranjape, R. (2013). Vocal Vision for Visually Impaired People. *The International Journal Of Engineering And Science (IJES)*, 2(3):1–7.

Bourbakis, N. (2008). Sensing surrounding 3-D space for navigation of the blind. *IEEE Engineering in Medicine and Biology Magazine*, 27(1):49–55.

Brunelli, R. (2009). *Template Matching Techniques in Computer Vision*. John Wiley & Sons, Ltd, Chichester, UK.

Deb, S., Reddy, S. T., Baidya, U., Sarkar, A. K., and Renu, P. (2013). A novel approach of assisting the visually impaired to navigate path and avoiding obstacle-collisions. In *2013 3rd IEEE International Advance Computing Conference (IACC)*, pages 1127–1130. IEEE.

ESpeak (2007). eSpeak text to speech. Avaiable at http://espeak.sourceforge.net/.

Fajarnes, G. P., Dunai, L., Praderas, V. S., and Dunai, I. (2010). CASBliP - a new cognitive object detection and orient ation system for imp aired people. In *Proceedings of the 4th International Conference on Cognitive Systems*, volume 7.

Hiebert, G., Charley, K., Harrison, P., Jot, J.-M., Peacock, D., Trivi, J.-M., and Vogelsang, C. (2017). OpenAL Programmer ' s Guide.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Katz, B. F. G., Kammoun, S., tan Parseihian, G. e., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., and Jouffrais, C. (2012). NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality*, 16(4):253–269.

Krizhevsky, A., Sutskever, I., and Geoffrey E., H. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pages 1–9.

Leonard, J. and Durrant-Whyte, H. (1991). Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(3):376–382.

Mehta, P., Kant, P., Shah, P., and Roy, A. K. (2011). VI-Navi : A Novel Indoor Navigation System for Visually Impaired People. *International Conference on Computer Systems and Technologies – CompSysTech'11 VI-Navi:*, pages 365–371.

Mocanu, B., Tapu, R., and Zaharia, T. (2015). An Obstacle Categorization System for Visually Impaired People. *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 147–154.

NVIDIA (2017). NVIDIA Embedded Systems and Developer Kits.

Pissaloux, E. (2002). A characterization of vision systems for blind people mobility. In *IEEE International Conference on Systems, Man and Cybernetics*, volume vol.4, page 6. IEEE.

Poggi, M. and Mattoccia, S. (2016). A Wearable Mobility Aid for the Visually Impaired based on embedded 3D Vision and Deep Learning. In *First IEEE Workshop on ICT Solutions for eHealth (IEEE ICTS4eHealth 2016) in conjunction with the Twenty-First IEEE Symposium on Computers and Communications*.

Schauerte, B., Martinez, M., Constantinescu, A., and Stiefelhagen, R. (2012). An assistive vision system for the blind that helps find lost things. *Computational Science and Its Applications–ICCSA 2007*, 7383 LNCS(Chapter 83):566–572.

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE.

Tapu, R., Mocanu, B., and Zaharia, T. (2016). A computer vision-based perception system for visually impaired. *Multimedia Tools and Applications*, pages 1–37.

WHO (2014). Visual impairment and blindness. Technical report.