A Crowd-Sourced Obstacle Detection and Navigation App for Visually Impaired

Edward Kim, Joshua Sterner, and Afra Mashhadi(✉)

Computing Software System, University of Washington, Bothell, WA, USA
{edkim84,sternerj,mashhadi}@uw.edu

Abstract. Individuals with sight impairments rely heavily on various types of travel-aid when navigating their ways across their neighborhoods. Recently, there have been many breakthrough technologies that focus on the visually impaired by providing solutions such as wearable bands and optical wearable devices. However, such technologies are costly and not suited for the general market. Others have started investigating smartphone applications as a much more widely available solution but with limited applicability on outdoor barriers and obstacles that these groups of people face in their day to day journeys. In this work, we propose GeoNotify, a smartphone application which is tailored to detect unexpected temporary obstacles that could cause injury to visually impaired people. We present how advances in Convolutional Neural Networks merged with crowd-sourcing methodologies could be used to build more accurate models capable of recognizing wide representations of the real-world obstacles.

Keywords: Crowd-sourcing · Navigation system · Federated learning

1 Introduction

People with *partial* sight loss face daily challenges getting around their local area for accessing local shops or visiting their local general practitioners. Currently in the UK alone over two million people live with sight loss, and this number has been forecasted to reach to four million by 2050. In a recent report published by *Royal Institute of Blind People* [1], sidewalk obstacles have been identified as the biggest concern of visually impaired when walking in their local neighborhood. Indeed in 95% of cases, blind and partially sighted people have collided with an obstacle over a three month period with nearly third of those being seriously injured. The most common obstacles are reported to be bins of all kinds (e.g., household), street furniture (e.g., restaurants outdoor table and chairs), advertising boards, and hanging baskets.

Recently, smartphone-based navigation systems for the blind has attracted much attention with some commercial key players. For example Blindways[1] is

[1] https://www.perkins.org/access/inclusive-design/blindways, last accessed June 5, 2020.

S. Paiva et al. (Eds.): SmartCity360° 2020, LNICST 372, pp. 571–579, 2021.
https://doi.org/10.1007/978-3-030-76063-2_38

an application that helps users get to their bus stop once they are within 30 ft of the bus stop location. This application provides the concept of using 'clues' to help users find the bus stop post through a crowd-sourcing methodology. Eye-D[2] provides speech feedback to communicate to the user where they are currently located. "Seeing AI"[3] uses artificial intelligence to assist the user with varities of tasks and tools including object recognition and optical character recognition, face recognition, but does not offer navigation information to the user. Blind-square[4] is a navigation application for the blind/visually impaired that provides detailed points of interest and intersections. These details are gathered from Foursquare and OpenStreetMaps and they use algorithms to determine what information would be useful for the user.

In this paper we propose Geo-Notify, a crowd-sourcing smart-phone application that is designed to assist visually impaired with their walking navigation by notifying them of the obstacles that are placed in their path. Geo-Notify relies on the participation of users to create a dataset of hazardous obstacles. We demonstrate that the current image recognition models fail to recognize the ordinary obstacles due to the lack of data available for this specific task and the real-world diversity in the presentation of obstacle images. We show that by using transfer learning, we can achieve a high accuracy of improving obstacle recognition by relying on as little as 20 images per class of the obstacles. Finally, we report the viability of our approach in a distributed setting where the model is trained locally on the user's smartphone device by leveraging the Federated Learning paradigm. The contributions of this paper are as follows:

– We design an outdoor smart-phone navigation application that enables visually impaired to receive notifications of hazardous obstacles on their route, and re-route them accordingly.
– We propose a crowd-sourcing mechanism whereby images of obstacles can be contributed to our convolutional neural network model so to learn the task specific and context dependent representations of images. We report the result of our experimental analysis under both centralized and distributed setting.
– We produce the first curated dataset of images of sidewalk obstacles that are reported as main sources of injuries by the Royal Institute of Blind People, and share it with future researchers.

2 Related Work

Recently, the navigation assisting systems for the visually impaired has attracted much attention [6,8]. Navigation assisting technologies try to guide the user to a destination or provide in-situ information about user's current surroundings. For example in [8] authors present a campus navigation system for visually impaired people that includes a smartphone and a sonar device. Smartphone

[2] Last accessed June 5, 2020 https://eye-d.in/.
[3] https://www.microsoft.com/en-us/ai/seeing-ai, last accessed June 5, 2020.
[4] https://www.blindsquare.com/, last accessed June 5, 2020.

is used to obtain GPS data by the built-in GPS receiving module, and sonar device is employed to detect obstacles on the road. A cloud based server is queried and updated to the user's location. The sonar device however is unable to distinguish the types of obstacles as it functions independently of the mobile device. In [19], authors proposed a guidance system for the blind by relying on the OpenStreetMap [9] through vibration feedback. Several applications are designed to provide information regarding nearby POIs [4]. A main challenge that is faced in providing contextual information in outdoor environment is managing the quantity of audio information by balancing the quantity of information against distracting or overwhelming the user [16].

Some research efforts try to complement these limitations by detecting elements that are relevant for navigation, such as crosswalks and bus stops using computer vision [6,10,14,17]. However these approaches are limited to the identification of the fixed structures as opposed to temporary obstacles. Moreover, previous studies [3,5] on uncovering the information needs of visually impaired have identified *interactivity* and the ability to *pull* information on surrounding environment as an important component that often goes missing in spatial in-situ navigation systems.

The closest work to ours is by Chen et al. [6] who integrated obstacle detection and GPS navigation into one system. The authors study the obstacle detection based on MobileNetV2 [12]. To this end, the authors manually collect a set of 4500 images from four specific class of *cars, pedestrians, bicycles* and *electric bicycles.* They present that the trained model performance ranges between 69%–96% accuracy across various categories, with cars identified as the most accurate classification class and people as the least. Our work differs, as we perform not only object detection on the mobile device but also use the crowd-sourced images for training of the convolutional neural network model. Indeed, most available approaches assume that an accurate model is previously trained on the representation of images of the possible obstacles and thus the model is used only for inference on the device. In contrast, in our work we illustrate how such assumption does not hold and propose a continuous learning model which is capable of learning new obstacles through a crowd-sourced mechanism.

3 Application Design

In this section we briefly describe the application design and system architecture of GeoNotify and functionalities of its various components.

3.1 User Interface

To design for accessibility, we followed the previous guidelines by [2,11,18]. In so doing, we utilize the whole screen to maximize on the space for each button. With the maximized space, we can provide large letters for each button as demonstrated in Fig. 1. We use a black background and white text to create contrast and provide large descriptive icons to help moderately visually impaired users find the right button. We also use dark mode as suggested by [11,18], to make it easier for mild to moderate visually impaired to read the text. There is also an option to provide speech feedback to let the user know which screen they are currently on. To navigate the application, the user just taps a button to choose an option, or swipes right to go back to the previous screen. We also provide a voice navigation button where the user can tell the application the screen they want to navigate.

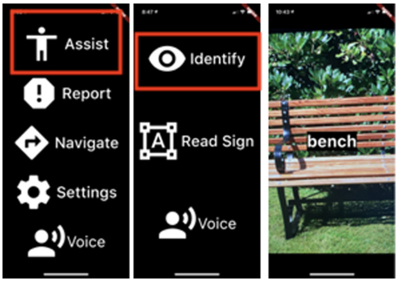

Fig. 1. Work flow (highlighted in red) of the Assist functionality of our application which is designed to help users to contextualize their surrounding environment. (Color figure online)

3.2 System Architecture

GeoNotify is composed of three main components (Navigate, Assist, and Report) with the overarching goal of providing a sense of additional safety for visually impaired users. Figure 2 presents the overall architecture of our system. The functionalities include object recognition, safety/obstacle monitor during navigation through notifications, as well as providing means for reporting issues through crowd-sourcing images. All these functions play a role in gathering data on the users' surroundings. The data gathered are stored into AWS relational database, which acts as proxy between the client and server. AWS Lambda[5] is a serverless compute service that acts as the access point to read and write to the database. For client application to be able to execute these lambda functions, we use AWS API Gateway. API Gateway creates an endpoint made available to the public where clients can make CRUD (Create, Read, Update, Delete) requests. API Gateway also seamlessly connects to AWS lambda functions by passing the data from the client to the lambda function[6].

3.3 Navigate

Our application provides the user with live information on the reported obstacles that are within their proximity. We rely on Google Maps API tool set to create

[5] https://aws.amazon.com/lambda/, last accessed June 5, 2020.
[6] https://aws.amazon.com/api-gateway/, last accessed June 5, 2020.

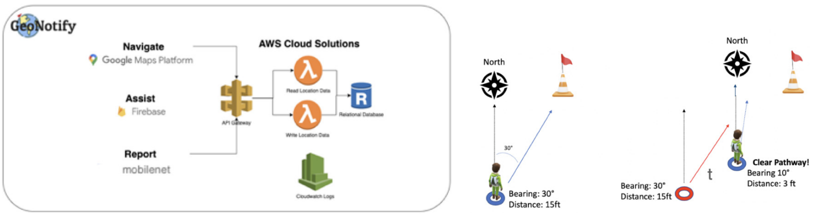

Fig. 2. The system architecture including the three main component of the GeoNotify application: Navigate, Assist and Report (left). Situating position of the obstacle by calculating the delta bearing angle (right).

a navigation path. The Navigate component takes the users current location and makes a get request to the API Gateway. When approaching any obstacle, the Navigate function uses push notification to warn the user about the obstacle and create an alternative route. In order to avoid obstacles, we calculate the bearing angle of the user with the nearby reported obstacles at every t time interval (Fig. 2). If the bearing stays the same for two consecutive time intervals we can infer that the user is heading towards the obstacle and reinforce the audio notification.

3.4 Assist and Report

The main underlying foundation of the Assist and Report component of our system is image recognition. The Assist component enables user to enhance their understanding of their surrounding environment. This module allows user to either request information about their surrounding by using Voice module which describes the nearby POIs to the user through audio, or alternatively to take a photo of their surrounding in order to read a sign or describe the objects in their proximity (Fig. 1). The Report module enables the users to voluntarily report obstacles thus contributing to the crowd-sourced dataset. The Report module also supports both audio and visual data. Through microphone integration, users are able to dictate the type of encountered obstacle on their path, which is then stored along with the GPS coordinate into our crowd-sourced database. Alternatively users can report any obstacles by take a photo of the nearby object. The photo is then classified using the obstacle detection component (as we describe in details next) and is recorded along with the GPS information.

4 Obstacle Detection

We use advances in Convolutional Neural Networks (ConvNet) to train a classifier that is able to identify the obstacles in the side walks. In particular we focus on two existing models that are lightweight and suitable for mobile device computation, MobileNet [12] and SqueezeNet [13]. MobileNetV2 proposed in 2018 by Google is a lightweight ConvNet that replaces standard convolution with depthwise separable convolution. With the improvement of convolution structure, the convolution computational complexity has been shown to be reduced significantly. SqueezeNet [13], is another state of the art lightweight ConvNet which has shown to achieve the same accuracy as AlexNet but 50 times faster.

In doing so we follow the user study by [1], which interviewed 500 visually impaired participants over the period of three months. This report identifies four most common obstacles that impact visually impaired daily: cars parked on the pavement, advertising-boards, bins and recycling boxes on pavements, street furniture (such as chair and tables). Some of these classes (e.g., chair and table) are common objects that are labeled in image recognition datasets such as ImageNet [7] and thus are detectable using existing ConvNet models. However, the presentations of these common objects across the world can vary significantly. Others such as boards, sidewalk signs and potholes are specific to the context of our study. We thus curated a dataset corresponding to 5 aforementioned common obstacles. Figure 3 presents the sample images in each category as well as the average in-class inference accuracy when the pre-trained MobileNet and SqueezeNet are applied on 100 test images from each category.

Fig. 3. Five classes of the obstacles that are identified as most common source of injuries for partially impaired.

As we rely on user contributions to construct our obstacle dataset, it is important to quantify the number of images that are required to train the model to learn the representation of an object. To this end, we rely on transfer learning, where we use the pre-trained MobileNet and SqueezeNet models and freeze the learning on the earlier layers where the weights are transferred across the network, and retrain the fully connected (FC) layer to learn the representations of our domain specific images. We evaluated these two models on a manually curated dataset. Our dataset consists of 5 classes where each class has x number of training images that we vary in our experiments, and 100 validation images. Figure 4 presents the validation accuracy and loss results over the 10 training epochs. This result suggests that using transfer learning both pre-trained networks are able to accurately learn the representation of the new obstacle objects (maximum validation accuracy of 0.9) when relying on as few as 20 training images per category.

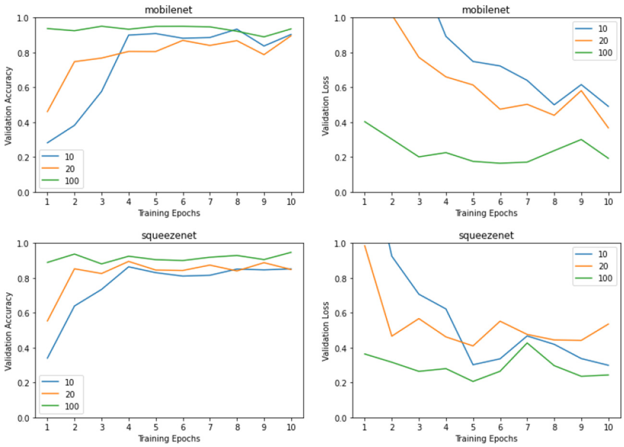

Fig. 4. Validation accuracy and loss for MobileNet and Squeezenet when applied under transfer learning setting to datasets of varying $x = 10, 20$ and 100 images per class.

To further assess the viability of relying on a crowd-sourced approach for obstacle detection, we also test both MobileNet and SqueezeNet using a Federated Learning [15] methodology. Federated learning is a recent machine learning paradigm where participants volunteer to collaboratively train a model under the orchestration of a central server, while keeping the training data on their own devices at all time. By following the principles of focused data collection and minimization, we rely on the users to contribute to the pre-trained global models (i.e., MobileNet and SqueezeNet that are pre-trained on ImageNet). In so doing, each user contributes a set of images across the five categories of obstacles to the retrain the final fully connected layer of the global model. As the training is done locally on the devices, the updated network weights are shared with the federated server and aggregated to Table 1 presents the average maximum validation accuracy and the average minimum loss for 5 cross-fold evaluation of the MobileNet using 10 epochs and batch size 20 (same hyper-parameters as the centralized setting).

Table 1. Maximum average validation accuracy and validation loss for varying number of images and contributors.

	No. images	No. users	$ValAccuracy$	$ValLoss$
MobileNet	20	20	0.86	1.18
		10	0.86	1.17
	100	20	0.93	0.9
		10	0.93	0.72

5 Conclusion

In this paper we proposed GeoNotify, a navigation and obstacle detection system for visually impaired users. We examined the viability of using crowd-sourcing approach to improve the obstacle detection model to recognize domain specific representation of new objects. We evaluated and reported the result of two ConvNet models, MobileNet and Squeezenet, when applied in a centralized setting using transfer learning, as well as preliminary results on when using a federated learning methodology.

We believe applications such as ours which rely on the crowds to build and improve the underlying models are needed as cities around the world go through rapid transformations. An example of such transformation is apparent across the world in the current pandemic where previous sidewalks and footpaths have been transformed to accommodate for businesses, thus creating a major hurdle for visually impaired. In our future work, we will expand our dataset to include more categories of obstacles and more number of images from across the world.

References

1. The royal institute for blind people (2016). https://www.rnib.org.uk/sites/default/files/who
2. Antunes, A.C., Silva, C.: Designing for blind users: guidelines for developing mobile apps for supporting navigation of blind people on public transports. In: User-Centered Software Development for the Blind and Visually Impaired: Emerging Research and Opportunities, pp. 1–25. IGI Global (2020)
3. Banovic, N., Franz, R.L., Truong, K.N., Mankoff, J., Dey, A.K.: Uncovering information needs for independent spatial learning for users who are visually impaired. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 1–8 (2013)
4. Blum, J.R., Bouchard, M., Cooperstock, J.R.: What's around Me? Spatialized audio augmented reality for blind users with a smartphone. In: Puiatti, A., Gu, T. (eds.) MobiQuitous 2011. LNICST, vol. 104, pp. 49–62. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30973-1_5
5. Brock, A.M., Truillet, P., Oriola, B., Picard, D., Jouffrais, C.: Interactivity improves usability of geographic maps for visually impaired people. Hum. Comput. Interact. **30**(2), 156–194 (2015)
6. Chen, Q., Wu, L., Chen, Z., Lin, P., Cheng, S., Wu, Z.: Smartphone based outdoor navigation and obstacle avoidance system for the visually impaired. In: Chamchong, R., Wong, K.W. (eds.) MIWAI 2019. LNCS (LNAI), vol. 11909, pp. 26–37. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33709-4_3
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
8. Dutta, S., Barik, M.S., Chowdhury, C., Gupta, D.: Divya-Dristi: a smartphone based campus navigation system for the visually impaired. In: 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), pp. 1–3. IEEE (2018)

9. Haklay, M., Weber, P.: Openstreetmap: user-generated street maps. IEEE Pervasive Comput. **7**(4), 12–18 (2008)
10. Hara, K., Azenkot, S., Campbell, M., Bennett, C.L., Le, V., Pannella, S., Moore, R., Minckler, K., Ng, R.H., Froehlich, J.E.: Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis. ACM Trans. Accessible Comput. (TACCESS) **6**(2), 1–23 (2015)
11. Hersh, M.: Mobility technologies for blind, partially sighted and deafblind people: design issues. In: Mobility of Visually Impaired People, pp. 377–409. Springer (2018)
12. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications (2017). arXiv:1704.04861
13. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡0.5 mb model size (2016). arXiv:1602.07360
14. Mascetti, S., Ahmetovic, D., Gerino, A., Bernareggi, C.: Zebrarecognizer: pedestrian crossing recognition for people with visual impairment or blindness. Pattern Recogn. **60**, 405–419 (2016)
15. McMahan, H.B., Moore, E., Ramage, D., y Arcas, B.A.: Federated learning of deep networks using model averaging. corr abs/1602.05629 (2016). arXiv:1602.05629
16. Panëels, S.A., Olmos, A., Blum, J.R., Cooperstock, J.R.: Listen to it yourself! evaluating usability of what's around me? for the blind. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2107–2116 (2013)
17. Shangguan, L., Yang, Z., Zhou, Z., Zheng, X., Wu, C., Liu, Y.: Crossnavi: enabling real-time crossroad navigation for the blind with commodity phones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 787–798 (2014)
18. Sierra, J.S., Togores, J.: Designing mobile apps for visually impaired and blind users. In: The Fifth International Conference on Advances in Computer-Human Interactions, pp. 47–52. Citeseer (2012)
19. Velazquez, R., Pissaloux, E., Rodrigo, P., Carrasco, M., Giannoccaro, N.I., Lay-Ekuakille, A.: An outdoor navigation system for blind pedestrians using GPS and tactile-foot feedback. Appl. Sci. **8**(4), 578 (2018)