Sensors and Materials, Vol. 34, No. 5 (2022) 1963–1970
MYU Tokyo

S & M 2946

Development of a Haptic Viewfinder Module for Locating Objects Using Vibrotactile Feedback

Aaron Raymond See* and Roberto Jacob Policar

Department of Electrical Engineering, Southern Taiwan University of Science and Technology,
Tainan 71005, Taiwan

(Received September 29, 2021; accepted November 22, 2021)

Keywords: haptic viewfinder, vibration, Eulerian motion magnification

Visually impaired and blind people use several mobile phone applications that enable them to socially interact in social media platforms. However, photography is challenging for them since it requires focusing on a target object. To address this issue, we developed a haptic viewfinder to assist the user in locating an object within the image frame of a mobile phone application. The system is initialized through speech-to-text conversion to provide the system an input on what item the user wants to focus on. To guide the user, the device uses artificial intelligence and image processing to locate objects, and then provides a vibrotactile feedback to localize a target object's location. Moreover, we evaluated the effectivity of the vibration through Eulerian motion magnification.

1. Introduction

According to a study conducted by Wu and Adamic,(1) visually impaired people have the same number of social activities as people without vision impairment in Facebook, such as posting daily status and photographs. Such activities are possible through the different assistive devices geared towards blind photography. To name a few assistive applications, on-screen readers and aim-assisting photography are among the popular technologies available today.(2,3) However, some, if not all, aim-assisting photography applications reduce the creativity of a visually impaired and blind (VIB) user by removing a few features that make the application easy to use.(3) Therefore, in this study, we would like to improve current aim-assisting applications while maintaining the ease of using such applications.

In studies conducted by Radecki *et al.*, sonification was used to let VIB students ages 10 to 16 years old understand shapes and images with only a few sessions of training.(4) Sonification maps the characteristics of an image into a certain frequency band. Moreover, the conversion process has been tested in audio modulation studies. In this study, we explore the possibility of its usage in vibrotactile displays to convey information. Previous research studies on vibrotactile displays and devices include navigation assistance,(5,6) shape determination,(7–9) and screen location searching.(10–12) With vibrotactile displays, the researchers used sonification to guide

---

*Corresponding author: e-mail: aaronsee@stust.edu.tw
https://doi.org/10.18494/SAM3668

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

their users in aiming their mobile devices. In this article, we will include the methods used to create the device, the test used to evaluate the overall system, and related theories that affect the overall design of the system. We will demonstrate the possibility of creating a device capable of controlling vibration through the location of the subject relative to that of the user's finger, but would not delve deeper in the creation of the text-to-speech algorithm, object detection models, and vibration-dampening materials. The proposed external hardware is a vibrotactile device, called the haptic viewfinder, which is paired with modern Android devices.

2. Materials and Methods

Figure 1 shows the system flowchart of the proposed haptic viewfinder. The haptic viewfinder acts as a location feedback system built on an Android operating system platform. Initialization is performed through speech control wherein the user would make a request to search for a specific object. Speech input is taken in through an Android mobile phone. The input would then trigger the camera of the mobile phone to initialize the search for the object within the frame of the application. Subsequently, through object detection and the use of artificial intelligence, the haptic viewfinder would use the context of the input speech to locate the object. Given that object detection models localize multiple objects within a frame, the haptic viewfinder uses the mobile application to search for the sought-out object and concurrently returns a certain magnitude value to scale the corresponding vibration feedback of the hardware. The vibration system guides the user in localizing the specified object. The system uses a pre-trained object detection model in combination with the coordinates of the finger to control the strength of the vibration feedback of the external device. If the desired object were not found, the system would have no feedback.

2.1 Natural language processing and object detection model

Natural language processing (NLP) is implemented to select the target that will be searched by the mobile application. The NLP was made possible by the Android OS' SDK. It allows the

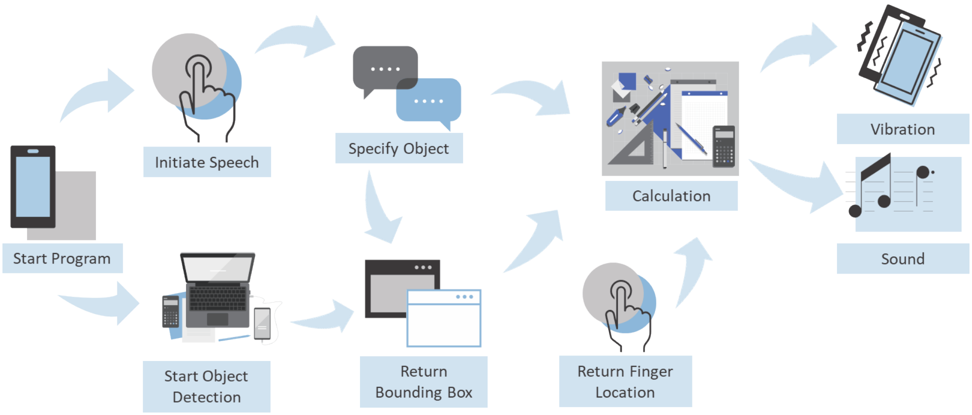


Fig. 1. (Color online) System flowchart of proposed haptic viewfinder.

speech-to-text function called "android.speech.SpeechRecognizer" to invoke the speech-to-text capability of the Android device. This process is used to remove the common limitation of current aim-assisting mobile applications that only focus on human faces.(2) On another note, the risk of using two hands for the device is also lessened as one of the stated possible problems by Vázquez and Steinfeld, which is commonly used by VIBs for other assistive devices (white canes) or companions (guide dogs).(2) However, as a limitation to the concept, the researchers assume that the user has sufficient knowledge that the target object is within the phone's peripheral vicinity.

After determining that the target object is within the list of items found in the frame, the overall system uses a pretrained mobile net model taken from the TensorFlow API and was significantly reduced by compressing the model using TensorFlow Lite (Table 1).(13) The model was pre-trained with the COCO dataset, which already has 90 classes. As a base prototype of the device, note that the object classes, although more than needed, are already capable as a place-holder model for the current experimentation procedure.

2.2 Motor frequency algorithm

Upon determining the location of the object, the information is then converted for feedback. A previous study by Tekli *et al.* indicated that simple shapes are easier to understand for vibrotactile displays.(8) By using the bounding box as a basis for the simple shape, the vibration frequency is arranged in ascending order as the user's finger runs closer to the shape. The method is based on the experiment performed by Radecki *et al.* in determining the different elevations in a map.(4) Moreover, another study showed that a gradient descent of vibration amplitude for haptic perception to locate the center is easily understood by examinees.(14)

Table 1
Classes of TensorFlow Lite model used.

person	bicycle	car	motorcycle	airplane
bus	train	truck	boat	dog
traffic	backpack	light	fire hydrant	stop signv
parking meter	bench	bird	cat	horse
sheep	cow	elephant	bear	zebra
umbrella	handbag	tie	suitcase	frisbee
kite	baseball bat	baseball glove	skateboard	surfboard
glass	cup	fork	knife	spoon
sandwich	orange	broccoli	carrot	hot dog
couch	potted plant	bed	dining table	toilet
remote	keyboard	cell phone	microwave	oven
book	clock	vase	scissors	teddy bear
toothbrush	sink	toaster	refrigerator	hair drier
TV	donut	pizza	cake	chair
bowl	banana	apple	laptop	mouse
skis	snowboard	sports ball	tennis racket	giraffe
bottle	wine			

To properly convey the information, vibration amplitudes of the four vibration motors are controlled through phantom sensation. Vibrations occurring in solid objects affect each other, making it difficult to determine the place of vibration, in comparison with out-of-body and on-skin phantom sensations.(15) However, if the vibration is dynamic or moving, the vibration from solid objects could inform the user with its location. Therefore, by using a finger on the capacitive screen sliding throughout the frame, the vibration would have a dynamic gradient depending on the Euclidean distance of the finger to the target object center (Fig. 2). The saltation intensity is determined through the equation

$$A = \frac{d_{f-t}}{D_{max}}, \tag{1}$$

where d_{f-t} is the Euclidean distance of the finger to the target object center and D_{max} is the maximum distance of the difference, which in this case is the length of the frame's diagonal.

The vibration frequency is distributed to four different vibration motors constrained at the four corners behind the video frame. With phantom haptics in mind, all vibration motors would affect each other on the basis of their frequency and distance from the finger. By the method of Park and Choi, the vibration strength of each motor is determined as(11)

$$A_i = A\left(1 - \frac{x_i}{x_{max}}\right)\left(1 - \frac{y_i}{y_{max}}\right), \tag{2}$$

where x_i is the horizontal distance of the finger to the vibration motor i, x_{max} is the maximum distance of the horizontal difference or the frame width, and y_i is the vertical distance of the finger to the vibration motor i.

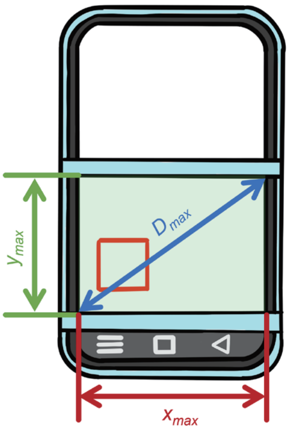


Fig. 2. (Color online) Representation of distances and slopes considered to solve for vibration frequencies.

2.3 Hardware

To build the prototype, three main components, namely, an ESP32 development kit (devkit), four eccentric rotating mass (ERM) vibration motors, and a buzzer, are used. The main controller of the device is the ESP32 devkit. The microcontroller unit has the basic functions required for the device such as Bluetooth (BLE) communication with the Android mobile device and pulse width modulation for the ERM vibration motors. The system would be handheld and coupled with the user's mobile phone. The device uses three Li-ion batteries (3.7 V each), each having a storage capacity of 2300 mAH. Finally, the device also uses a small 3 V buzzer to determine the moment the finger is already inside the bounding box. The device is built into a mobile device casing 3D-printed using polylactic acid (PLA).

3. Results and Discussion

3.1 Haptic viewfinder and mobile application

Figure 3 shows the prototype with its height, length, and width of 182, 84.4, and 56.4 cm, respectively. The total device weight is approximately 0.4 kg without the mobile phone. The dimensions are fitted for the use of a Samsung A71 Android mobile phone. Communication with the ESP32 devkit is established through BLE communication. Subsequently, the object detection responds to multiple objects in real time and will only vibrate once the user initializes the target object through speech input as shown in Fig. 4.

3.2 Experimental results

The haptic viewfinder is tested through Eulerian motion magnification (EMM). It magnifies minute changes in the pixels of a video and amplifies them to be visible to the naked eye.(16) It uses the video feed to capture vibrations by separating the high-frequency and low-frequency components of the video. Then, the low-frequency video motion is magnified and later on

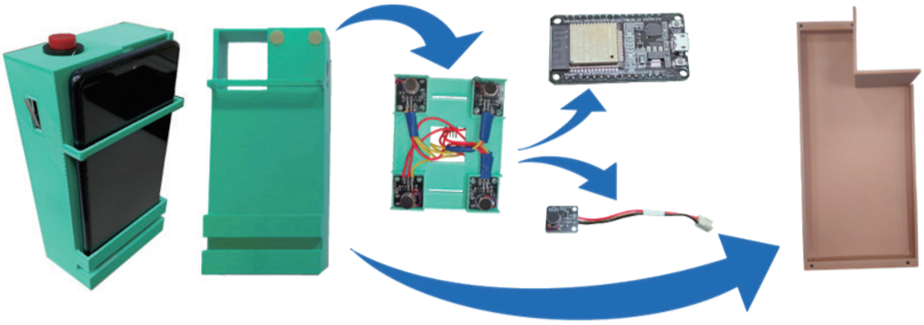
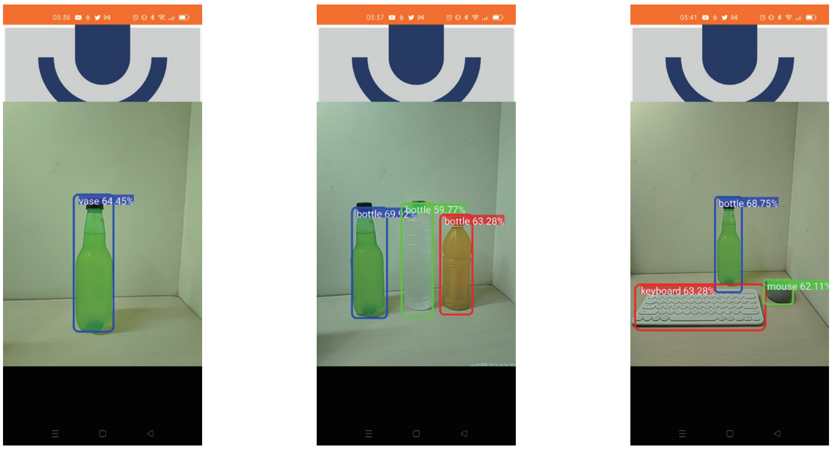


Fig. 3. (Color online) Haptic viewfinder prototype



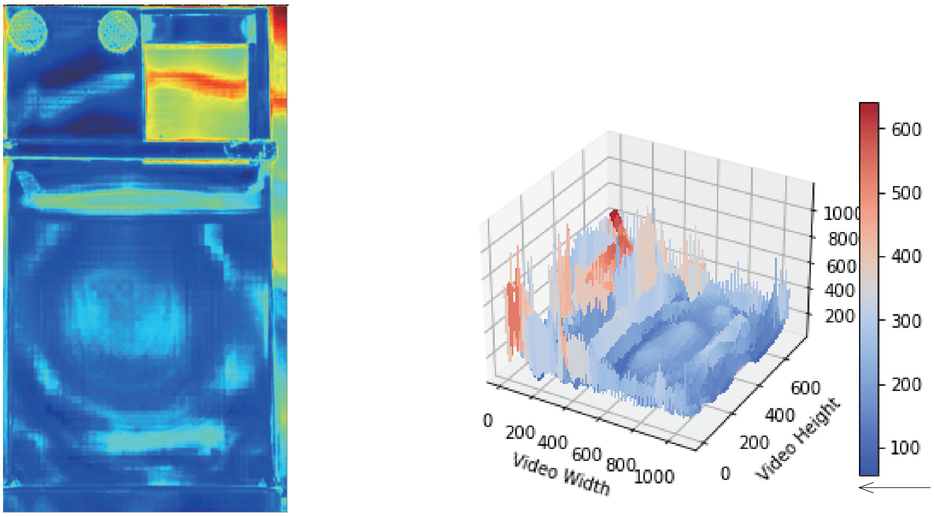
(a) (b) (c)

Fig. 4. (Color online) Mobile applications with (a) one object, (b) multiple similar objects, and (c) multiple different objects.

combined back to the high-frequency component of the video. The resulting video enhances the motion of vibration, which is typically indistinguishable to the naked eye. In the case of the test, a short 2 min video is shut in 240 frames per second to capture only the notable differences in frequency mentioned in previous studies.(17,18) The video would only contain the vibrations of the external device to remove the light variation from the mobile phone. On the other hand, the mobile phone is still operating and continuously sending Bluetooth commands to the device. A target is observed using the camera on the lower left part of the frame while the user's finger is located on the upper right as shown in Fig. 5(a). The video is then cut to 1 min segments, removing interruptions from the start and end of the experiment. The video is then run through the EMM algorithm with the magnification value set to 50. The unmagnified video is then subtracted from the magnified video and a pseudo-color is imposed to clearly observe the vibration throughout the device. Figure 5(b) shows a 3D graph of the vibration signals.

3.3 Discussion

In the completion of the first prototype, some improvements could be addressed in future works. One improvement is to study what objects should be included in a camera-oriented mobile application. In the current works stated above, most camera apps only include text finders, face finders, and pedestrian objects(2). However, studies for commonly posted social media images have not been discussed for camera assistive mobile applications. Another point of improvement is the use of a device with higher isolation performance and a different haptic actuation. In the papers of Yang *et al.*(6) and Seo and Choi,(11) linear resonant actuators are used to separately control the amplitude and vibration strength. Subsequently, owing to the current pandemic, we are not able to perform tests on VIB users. We plan to work with them to test the system in the near future.



(a) (b)

Fig. 5. (Color online) Vibration isolation test of prototype by Eulerian motion magnification of recorded video where (a) shows an overview of magnified video difference with pseudo-color where red displays location with highest vibration frequency and (b) is 3D graph of total vibration that occurred throughout a video in pixel location.

4. Conclusions

A haptic viewfinder is an assistive device for assistive aiming without limiting the creative freedom of VIB users. Its design and algorithm are based on the theory of sonification and phantom haptic studies that map images or shapes to different frequencies. These frequencies are used as the basis for vibration modulation in guiding the user to the desired target object. Through Eulerian motion magnification, the desired vibration location and modulation are visible and have been tested. The proposed device can be a potential assistive device for improving VIB photography and we will test it after pandemic restrictions are eased.

Acknowledgments

This research was funded by the Ministry of Science and Technology of Taiwan, grant numbers MOST 109-2222-E-218-001-MY2 and MOE 1300-108P097. We would like to thank Resources for the Blind in the Philippines for their inputs that gave us insights and inspiration for the project.

References

1 S. Wu and L. Adamic: Proc. Conf. Human Factors in Computing Systems (ACM, 2014). https://doi.org/10.1145/2556288.2557415
2 M. Vázquez and A. Steinfeld: Trans. Comput.-Human Interact. **5** (2014) 25. https://doi.org/10.1145/2651380
3 C. Jayant, H, Ji, S. White, and J. Bigham: Proc. 13th Int. ACM SIGACCESS Conf. Computers and Accessibility - ASSETS '11 (ACM, 2011). https://doi.org/10.1145/2049536.2049573
4 A. Radecki, M. Bujacz, P. Skulimowski, and P. Strumi: Br. J. Edu. Technol. **51** (2019) 2. https://doi.org/10.1111/bjet.12852

5 A. R. See, L. V. Costillas, W. D. Advincula, and N. Bugtai: Sens. Mater. **33** (2021) 6. https://doi.org/10.18494/SAM.2021.3221
6 G. H. Yang, M. S. Jin, Y. Jin, and S. Kang: Int. Conf. Int. Robots and Systems (IEEE, 2010). https://doi.org/10.1109/iros.2010.5651759
7 Y. Ryu and E. Ryu: Sens. Mater. **29** (2017) 7. https://doi.org/10.18494/SAM.2017.1594
8 J. Tekli, Y. B. Issa, and R. Chbeir: Int. J. Hum. Comput. Stud. **110** (2018) 1. https://doi.org/10.1016/j.ijhcs.2017.10.009
9 N. Giudice, H. Palani, E. Brenner, and K. Kramer: Proc. 14th Int. ACM SIGACCESS Conf. Comput. and Acc.-ASSETS (ACM, 2012). https://doi.org/10.1145/2384916.2384935
10 J. Seo and S. Choi: 2015 IEEE World Haptics Conf. https://doi.org/10.1109/whc.2015.7177686
11 G. Park and S. Choi: Proc. Conf. Human Factors in Computing Systems (ACM, 2018). https://doi.org/10.1145/3173574.3173832
12 G. Park, H. Cha, and S. Choi: IEEE Trans. Haptic **12** (2019) 1. https://doi.org/10.1109/toh.2018.2859955
13 TensorFlow Developers: Zenodo (2021) 1. https://doi.org/10.5281/zenodo.5095721
14 F. Maurel, G. Dias, W. Safi, J. M. Routoure, and P. Beust: Micromachines **11** (2020) 4. https://doi.org/10.3390/mi11040376
15 S. Choi and K. Kuchenbecker: Proc. IEEE **101** (2013) 9. https://doi.org/10.1109/jproc.2012.2221071
16 H. Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman: ACM Trans. Graphics **31** (2012) 4. https://doi.org/10.1145/2185520.2185561
17 N. Cao, H. Nagano, M. Konyo, S. Okamoto, and S. Tadokoro: World Haptics Conf. (IEEE, 2017). https://doi.org/10.1109/whc.2017.7989957
18 J. Tennison, P. M. Uesbeck, N. Giudice, A. Stefik, D. Smith, and J. Gorlewicz: ACM Trans. Appl. Percept. **17** (2020) 2. https://doi.org/10.1145/3383457

About the Authors

Aaron Raymond See is an assistant professor in the Department of Electrical Engineering at Southern Taiwan University of Science and Technology. He is concurrently the Director of Student Affairs in the Office of International Affairs. He was born in Manila, Philippines, and received his B.S. degree in electronics and communications engineering from De La Salle University, Manila, Philippines, in 2006. He obtained his masters and Ph.D. degrees in electrical engineering with a major in biomedical engineering at Southern Taiwan University of Science in Technology in 2010 and 2014, respectively. He handles courses in assistive technology device design and engineering education curriculum transformation. His research interest is in assistive device design and development, biomedical image and signal processing, and engineering education.

Roberto Jacob Policar received his B.S. degree in manufacturing engineering and management with a specialization in robotics and mechatronics from De La Salle University, Manila, Philippines, in 2019, and is currently pursuing his M.S. degree in electrical engineering at Southern Taiwan University of Science and Technology, Tainan, Taiwan. His current research involves machine vision, prosthetics, and haptic-enhanced assistive devices.