




ALO: AI for Least Observed People

Shamim Al Mamun¹(✉) , Mohammad Eusuf Daud², Mufti Mahmud³,
M. Shamim Kaiser¹, and Andre Luis Debiase Rossi⁴

¹ Institute of Information Technology, Jahangirnagar University,
Savar, Dhaka 1342, Bangladesh
shamim@juniv.edu

² Consultant, Ministry of finance, Naples, Italy

³ University of Nottingham, Nottingham, England

⁴ Universidade Estadual Paulista, Itapeva, Brazil

Abstract. In recent years, visual assistants of humans are taking place in the consumer market—the eye-line of humans equipped with a see-through optical display. Computer Vision Technology may play a vital role in visually challenged people to carry out their daily activities without much dependency on others. In this paper, we introduce ALO (AI for Least Observed) as an assistive glass for blind people. It can listen as a companion, read from the internet on the fly, detect surrounding objects and obstacles for freedom of movement, and recognize the faces he is communicating with. This glass can be a virtual companion of the users for social safety from unknown people, reduce the dependency of others. This system uses the camera for identifying human faces using MTCNN deep learning technique, bone conduction microphone, and google API (Application Programming Interface) for translating voice to text and text to bone conduction sound. A Market Valuable Product (MVP) has already been developed depending on our survey of over 300 visually impaired persons in Europe and Asia.

Keywords: Smart glass · Blind vision · Face recognition · Object detection

1 Introduction

Globally, at least 2.2 billion people have a vision impairment or blindness. The majority of these populations face moderate or severe distance vision impairment or blindness due to refractive error, cataracts, glaucoma, corneal opacities, diabetic retinopathy, and trachoma. In terms of regional differences, blindness in low and middle-income countries projects to be four times higher than in developed countries. In the Asian region, the blindness problem is 10% lower than the developed region [1].

The healthcare industry is in the midst of a transformative period, thanks to the emergence of sensor technology and the rapid implementation of the Internet of Things (IoT) [2–4]. These sensors and IoT devices provide researchers with

several opportunities to develop assistive products for people with special needs in a variety of ways [5–8].

In recent years, researchers have attempted to develop innovative gadgets that will be of assistance to visually impaired people in their daily lives. Nevertheless, since the end-users are real blind people, finding a solution must be based on empathy for the situation. The standard survey method will never be effective in identifying their life’s most difficult challenges. The process of prototyping hardware and software involves a great deal of thinking and iteration to be successful. There are unique problems in each and every fundamental domain of Artificial Intelligence – Machine Learning, Cognitive Vision, and Natural Language Processing, for example – from data preparation to attaining high accuracy levels using performance measurements [18, 19]. Though the challenges, researchers and startup industries are trying to develop vision assistance to blind people. We are proposing an intelligent eyeglass for blind people that can assist them in indoor environments. In indoor, the blind or low visioned person needs daily activity like pouring water into the glass for drinking. In this scenario, users need to find the glass and water bottle first and know its location in indoor premises [16]. Therefore, they need a companion or ask help from family members to serve them. In addition, in the developed country, blind people have many community services in a working hour, but in the Asian region, that kind of service is rare. So, they need to help themselves to find someone by oral communication. If glass demonstrate someone around them like in [20, 21], it would be great. In this context, researchers and startup companies come up with some solutions with the help of AI. One of the widespread products provided by google’s X lab. This intelligent glass uses primary navigation and localizes information based on the mobility of the mobile user on the road. It uses an optical head-mounted display connected via wi-fi or pairing with an android mobile phone to assist the user in getting information surrounding them. It also uses its Speech to Text/Text to Speech (GTTS) for communication with users. According to our study, researchers try to assist real blind people by innovating technology for oral communication (OC). Moreover, gives observability of the surrounding through the camera, object detection (OD), face detection/identification (FDI), optical character recognition, and bone conduction (BC) for ear free hearing. The summary of our findings from different research groups describes in tabular form in Table 1.

Table 1. System review for blind people assistive product.

Product	Developer	OC	NAV	OD	FDI	BC
Google glasses [9]	Google Inc	Yes	Yes	Yes	Yes	No
Aira [10]	Suman Kanuganti	Yes	Yes	Yes	No	No
eSights [11]	CNETs	No	No	Yes	Yes	No
[12]	CCES, PMU	Yes	Yes	No	No	No
ALO	Yes	Yes	Yes	Yes	Yes	Yes

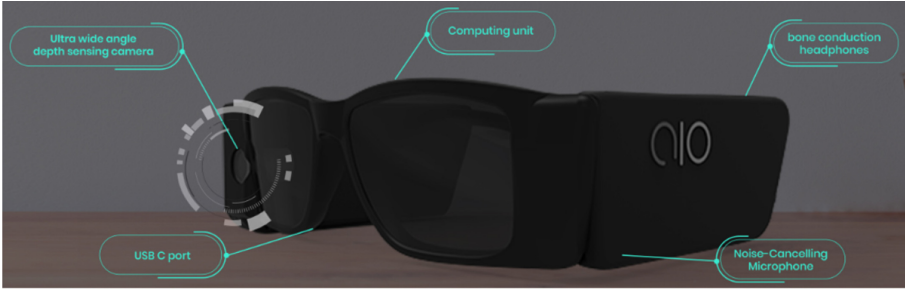


Fig. 1. Overall system specification

In this paper, we are proposing an intelligent goggle named ALO (AI for Least Observed) that can communicate with blind people or user with the help of the internet. It can able to read text from any newspapers. We use “google text to speech (gTTS)” service API for reading the text and send it to our bone conduction microphone because we want to free the user’s ear from hearing bud. It can recognize the object around the user so that s/he can find their necessary things without the help of companions or family members. Our glass also trained to recognize the user’s family members or acquaintances with whom they communicate in their daily lives. These features make them self-dependent when they go outside. When blind people are in their comfort zone, users want to do their work like finding water glass and bottle, TV remote, finding the dress to wear, etc. Our system acts as a complete companion in their life to give them freedom of movement in an indoor scenario. Figure 1 illustrated the overall system features and real market valuable product (MVP) unit. In the subsequent Sects. 2, 3 will describe ALO briefly with our survey over 300 blind participants from Asia and European regions.

2 AI for Least Observed (ALO)

We are proposing ALO in collaboration with the Ministry of Economic Development, Italy, and the ICT division of the Bangladesh government by conducting a survey of 300 participants from both countries. Among them, 158 are male participants, and 142 are female in 6 different age groups illustrated in Table 2. Our survey illustrated that 60% male and 40% female are suffering from low vision due to accident or disease. We have also interviewed them personally by social interaction method in the park, road, or home. 88.2% of the sample population feels lack of autonomy in aspect to

- Dependency on assistance,
- Not understanding the Surrounding,
- Inability to recognize the facial expression of others,
- Inability to find Objects, and,
- Inability to understand Distance and Direction.

Table 2. Survey sample size of 300 participants

Age	Sample size	Male	Female
<20	18	12	6
21–30	25	14	11
31–40	64	34	30
41–50	67	35	32
51–60	76	38	38
61–70	50	25	25

Moreover, 81.2% feels worried in the aspect of 1: Lack of visual information, 2: Unable to determine risks, 2: person/kids/pets in the vicinity, Obstacles/risks around the path, 3: Unable to see the facial expression of others and 4: Inability to understand texts/reading materials. Moreover, we found that lack of self-esteem, Inability to develop a desired life or career, Forced to follow a routine without any exception, Psychological health condition of the subject, Gradual disconnect from society (example: Friends or colleagues), and Self-isolation by the subject [22]. Therefore, Fig. 1 ALO gives them a chance to lead their own life without dependency, communicate with other persons, remember acquainted persons in the next meet, measure the object’s distance to avoid or pick up, and read contents from the internet.

3 System Overview

Our goal is to make a Market Valuable Product (MVP) for the slightest vision or blind user. So that, they can read text from internet resources like Wikipedia or any daily newspaper, detect and recognize the faces of family members or acquainted person at home or roadside walking, detection of an object for using or avoiding hazardous. ALO uses Raspberry pi zero W model with 8 Mega Pixel camera module (Sony IMX219 image sensor) supported by 850 mAh battery for 2 h of continuous energy supply. It also uses a 2.4 GHz wi-fi module for home use. Noise cancellation MIC and used Remax earphones wireless audio driver circuit running a bone conduction module. The dimension of ALO is 154 mm × 43 mm and 147 mm from the side view angle and 154 mm × 40 mm from the front view. Total weight is average 80 gm only, which will give comfort for the users (Fig. 2).

3.1 Real Time Messaging

Our system also proposes a high level of service integration ecosystem to support blind users through this glass illustrated in Fig. 4. ALO uses a contextual chatbot for a live conversation with users. Users give voice commands through noiseless MIC using a real-time messaging protocol like person detection or object detection using Google speech API services. Moreover, Contextual chatbot API then

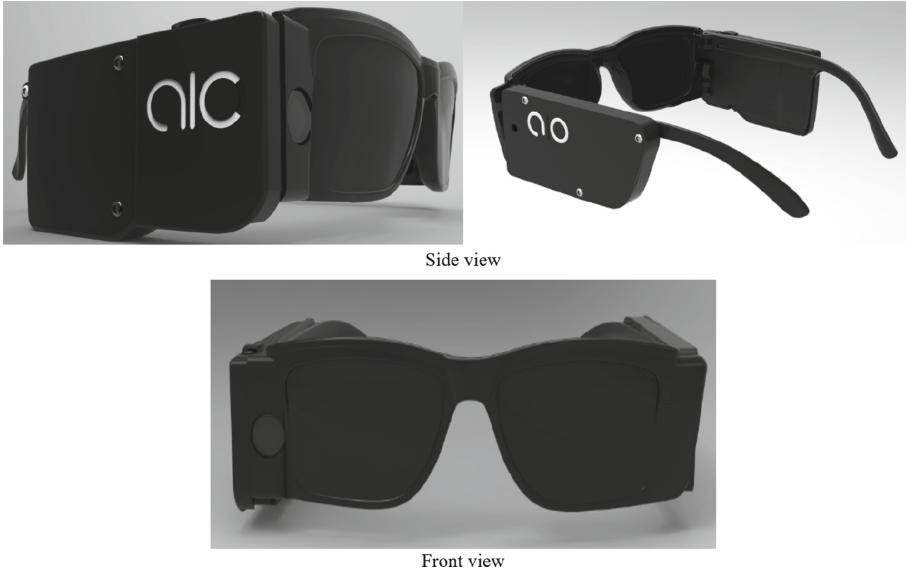


Fig. 2. Side and front view of ALO

hooks the scheduler model update service to get the API for other services. Model builder and model service layer are separated because of continuous update of the model when ALO gets new objects or persons. ALO assists the different users simultaneously from the web. Hence it will open API services for real-time messaging among targeted users and storage of ALO. For real-time messaging, ALO uses Mesibo [13] frameworks for building on system FnF APP which can open instantaneous replies from the chatbot. The working procedure of Mesibo illustrates in Fig. 3. Moreover, the chatbot model continuously updates its model depends on the conversations. Users to ALO real-time communication makes extremely simple by Mesibo. mesibo know about each of ALO end user. Mesibo will create an access token for each user and give it to the unit using the internet (using mesibo Server-side Admin API) to give respective access tokens to the users. Therefore, users use this access token in Mesibo SDK to create a real-time connection with the mesibo server to send and receive real-time messages. Using this system, users can communicate with the glass by voice command. To implement this task, we use Google Speech-to-Text (gSTT), which interfaced with a python library to get the message into our Mesibo architecture for real-time execution of the command. gSTT has a multi-language support API for convert the ultimate length of voice to text. This API recognizes over 120 languages and variants to support the user base. gSTT uses deep natural language machine learning techniques to identify the languages for instant translation to text. In our system, we use only English and Bangla for MVP purposes. We have used a simple questionnaire in fig that the user may ask to find an object or person.

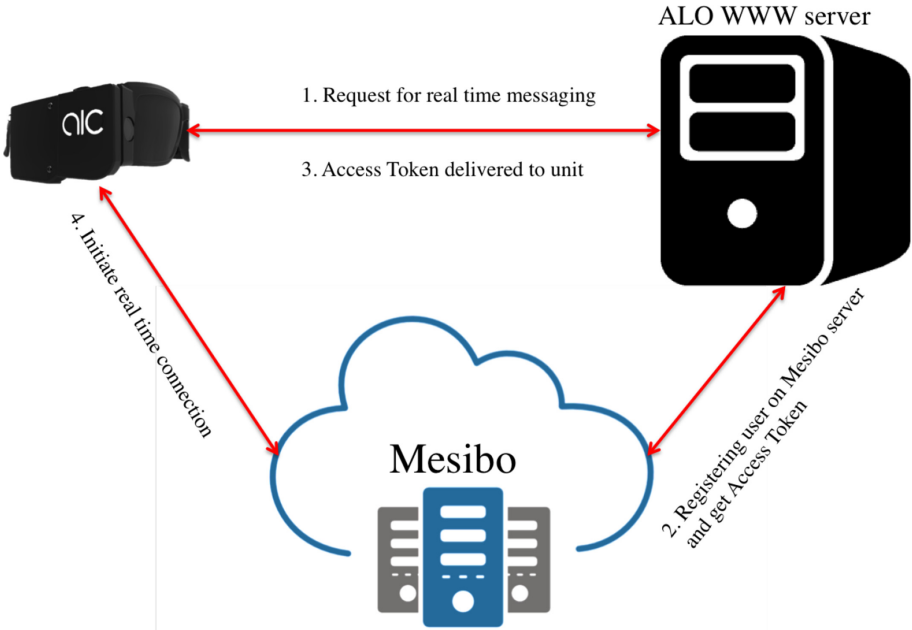


Fig. 3. Mesibo works flow

```

ques1 = "Where are we"
ques2 = "Why you think so"
ques3 = "No thanks"
ques4 = "Let us get introduced with these person"
ques5 = "What is in front of me"
b_ques5 = 'সামনে কি আছে' #'ওকে'
ques6 = "what are the items on the table"
b_ques6 = "টেবিলে কি আছে"
    
```

Fig. 4. Question ask by the user to the system

3.2 Bone Conduction Unit

The regular headphone uses wired or wireless passes through the ear canal to the eardrum for making a spectrum of voice. The bone conduction unit of our system works by vibrating against the bones in our cheeks or upper jaw, passing the ear canal to direct hit on the eardrum. Therefore, users are independently listening to system output as well as surrounding events. We included HBQ-Q25C TWS Wireless Bluetooth Headphones Ergonomic Waterproof Earbuds Ear Hook Bone Conduction (BC) Earphones module. This particular module has 10 m of transmission distance with a sensitivity range 50–180 KHz. In Fig. 5 exhibit our BC unit attached to ALO, modified to be compatible with our system for connecting it. We opened the bought bone conduction product and tried to

reset up it. We have changed the battery and circuit of the bone conduction unit and attached it with our glass handle. So, this BC module is no more wireless Bluetooth, and it now acts as a System on a chip (SOC) (Fig. 6).

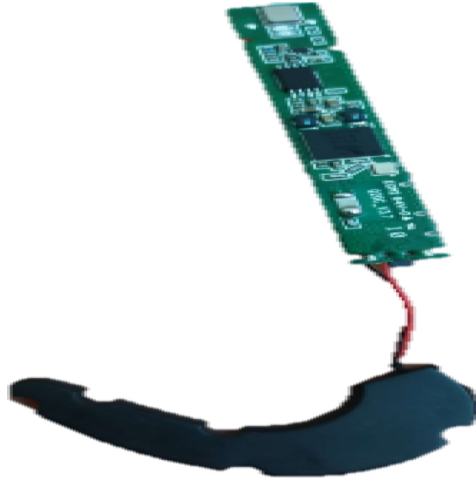


Fig. 5. ALO Bone Conduction Module

3.3 Object Detection

Our system can successfully detect objects in an indoor situation to find out the objects like water bottles, cups etc., for daily uses. We used Look Only Once (YOLO) [16] for object detection based on VGG-16 deep learning architecture. Though most Image Recognition Systems (IRS) use GPUs, we use the tiny version of YOLO's cpuNet that runs on a CPU at 15 FPS, which is quite good for detecting only 80 classes of objects using COCO dataset [14]. COCO dataset contains 82,783 training, 40,504 validation, and 40,775 testing images split into train, validation, and test data. There are nearly 270k segmented people and 886k segmented object instances in the 2014 train and validation data alone. The cumulative 2015 release will contain 165,482 train, 81,208 val, and 81,434 test images. Moreover, we use VGG-16 net to train the model developed in Keras and the google TensorFlow framework. The system is identifying an object with 98% of accuracy. In addition, when a user gives a voice command like "What is in front of me?" using mesibo framework, it takes a camera feed. It feeds into our IRS model to identify object and labeling as categories and name it. The user gets feedback sound through a bone conduction microphone which is attached to our unit. Figure 7 shows the accuracy of detecting objects using our system.

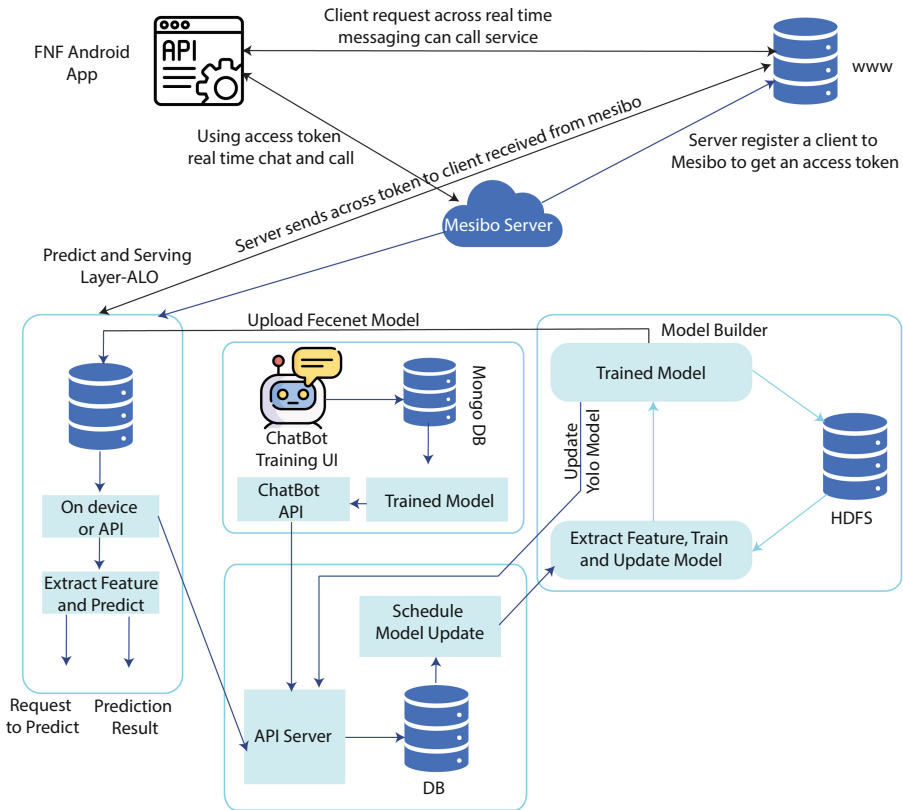


Fig. 6. System Architecture of ALO

3.4 Face Detection and Identification

Humans are using feature-based techniques to solve vision problems. The machine also uses the same process, such as the cascade classifier. Nowadays, deep learning methods have achieved state-of-the-art results on standard benchmark face detection datasets like Multi-task Cascade Convolutional Neural Network (MTCNN). In the traditional feature engineering method, we need thousands of features or kernels to detect the facial point to identify the face. But, in MTCNN, face detection and face alignment are done jointly, in a multi-task training fashion. Face alignment allows the model to detect better faces that are initially not aligned. MTCNN model structure uses three networks; at the first stage of this system, the image is resized to a range of different sizes called an image pyramid. And, then the first model (Proposal Network or P-Net) proposes candidate facial regions, the second model (Refine Network or R-Net) filters the



Fig. 7. System performance of ALO for object detection.

bounding boxes, and the third model (Output Network or O-Net) offers facial landmarks. There are three types of prediction uses in MTCNN; face classification, bounding box regression, and facial landmark localization. Three models are not directly connected and act as a lap of a sprint race. When one round finishes, the next one starts, and so on until the 3rd lap finished. The additional processing is performed between stages; for example, non-maximum suppression (NMS) is used to filter the candidate bounding boxes proposed by the first-stage P-Net, the input feeder R-Net model. Figure 8 shows the camera feed of faces and extract faces from the feed, and Fig. 9 illustrated that our system could recognize those two faces using MTCNN framework. In our system, we use OpenCV library with python to implement the model. As we are considering multiple models in our architecture to work simultaneously and recognize the acquainted person’s faces, we have developed multi-layer trained deep CNN for cascading the facial images in training and validation. Our system also teaches individual models with a single image instead train the whole model for classification. Moreover, multiple users can update their model on-the-fly mode to store and prepare their model without dormant the previous weight file rather than update it concurrently.



Fig. 8. Face detection using MTCNN

```

Say something!
You said: what is in front of me

1080 1920
{'person'}
Start Recognition!
Face Detected: 1
[534 378]
yes
[[0.03438325 0.04981384 0.88569597 0.03010695]]
best_class_probabilities [0.88569597]
['Akshay Kumar', 'Salman Khan', 'Tushar', 'kazi']
Tushar
Start Recognition!
Face Detected: 1
[492 359]
yes
[[0.04501446 0.04126357 0.0085204 0.90520156]]
best_class_probabilities [0.90520156]
['Akshay Kumar', 'Salman Khan', 'Tushar', 'kazi']
Tushar kazi
Tushar kazi 2 person
Say something!

```

Fig. 9. System output for MTCNN face detection

4 Conclusion

In this paper, we propose communicative assistive tools for blind people found by the survey of 300 participants of real blind users. Our system gives level up the user's confidence level in their daily life. ALO also gives them a comfort zone to communicate with another person whether they are unknown to them. Hence, also improve their security when they are communicating with an unknown person at home or road. We will do more experiments in natural

environment scenarios shortly and find their comfort level for using ALO. Additionally, implement improved circuitry for resizing the glass at the minimum size and weight. We will train our system with different objects, languages and enhance the voice-to-text accuracy also.

Acknowledgment. This work was financial supported by the Information and Communication Technology Department of the Bangladesh Government through startup Bangladesh Program and ALO Limited.

References

1. Blindness and Vision Impairment. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>. Accessed on 24 Feb 2020
2. Noor, M.B.T., Zenia, N.Z., Kaiser, M.S.: Challenges ahead in healthcare applications for vision and sensors. In: Ahad, M.A.R., Inoue, A. (eds.) *Vision, Sensing and Analytics: Integrative Approaches*. ISRL, vol. 207, pp. 397–413. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75490-7_15
3. Esha, N.H., Tasmim, M.R., Huq, S., Mahmud, M., Kaiser, M.S.: Trust IoHT: a trust management model for internet of healthcare things. In: *Proceedings of International Conference on Data Science and Applications*, pp. 47–57. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-7561-7_3
4. Farhin, F., Kaiser, M.S., Mahmud, M.: Secured smart healthcare system: blockchain and bayesian inference based approach. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) *Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing*, vol. 1309. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_36
5. Kaiser, M.S., et al.: 6G access network for intelligent internet of healthcare things: opportunity, challenges, and research directions. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) *Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing*, vol. 1309. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_25
6. Kaiser, M.S., Al Mamun, S., Mahmud, M., Tania, M.H.: Healthcare robots to combat COVID-19. In: Santosh, K., Joshi, A. (eds.) *COVID-19: Prediction, Decision-Making, and its Impacts. Lecture Notes on Data Engineering and Communications Technologies*, vol. 60. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-9682-7_10
7. Farhin, F., Kaiser, M.S., Mahmud, M.: Towards secured service provisioning for the internet of healthcare things. In: *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1–6. IEEE (2020)
8. Jesmin, S., Kaiser, M.S., Mahmud, M.: Artificial and internet of healthcare things based Alzheimer care during COVID 19. In: Mahmud, M., Vassanelli, S., Kaiser, M.S., Zhong, N. (eds.) *Brain Informatics. BI 2020. Lecture Notes in Computer Science*, vol. 12241. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59277-6_24

9. Google Inc.: Google glasses. Retrieved from https://en.wikipedia.org/wiki/Google_glasses. Accessed on 28 Feb 2020
10. Aira: By Your Side Throughout Life's Journey. Accessed on 28 Feb 2020
11. Electronic Glasses for the Blind. <https://esighteyewear.com/>. Accessed on 28 Feb 2020
12. AlSaid, H., et al.: Deep Learning Assisted Smart Glasses as Educational Aid for Visually Challenged Students. In: 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS). IEEE (2019)
13. <https://www.mesibo.com/what>. Accessed on 27 Feb 2020
14. Al Mamun, S., Lam, A., Kobayashi, Y., Kuno, Y.: single laser bidirectional sensing for robotic wheelchair step detection and measurement. In: Huang, D.S., Hussain, A., Han, K., Gromiha, M. (eds.) Intelligent Computing Methodologies. ICIC 2017. Lecture Notes in Computer Science, vol. 10363. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63315-2_4
15. Lin, T.-Y., et al.: Microsoft coco: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol. 8693. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
16. Al Mamun, S., et al.: Autonomous bus boarding robotic wheelchair using bidirectional sensing systems. In: Bebis, G., et al. (eds.) Advances in Visual Computing. ISVC 2018. Lecture Notes in Computer Science, vol. 11241. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03801-4_64
17. Zhang, K., et al.: Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Let.* **23.10**, 1499–1503 (2016)
18. Kaiser, M.S., et al.: iWorkSafe: towards healthy workplaces during COVID-19 with an intelligent pHealth App for industrial settings. *IEEE Access* **9**, 13814–13828 (2021)
19. Kaiser, M.S., Al Mamun, S., Mahmud, M., Tania, M.H.: Healthcare robots to combat COVID-19. In: Santosh, K., Joshi, A. (eds.) COVID-19: Prediction, Decision-Making, and its Impacts. Lecture Notes on Data Engineering and Communications Technologies, vol. 60. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-9682-7_10
20. Rahman, M.M., Mamun, S.A., Kaiser, M.S., Islam, M.S., Rahman, M.A.: cascade classification of face liveness detection using heart beat measurement. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol. 1309. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_47
21. Tabassum, T., Tasnim, N., Nizam, N., Al Mamun, S.: anonymous person tracking across multiple camera using color histogram and body pose estimation. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol. 1309. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_52
22. Asif-Ur-Rahman, M., Afsana, F., Mahmud, M., Kaiser, M.S., Ahmed, M.R., Kaiwartya, O., James-Taylor, A.: Toward a heterogeneous mist, fog, and cloud-based framework for the internet of healthcare things. *IEEE Internet Things J.* **6**(3), 4049–4062 (2018)