FEDERAL UNIVERSITY OF TECHNOLOGY, AKURE
JOURNAL OF ENGINEERING AND ENGINEERING TECHNOLOGY

A WEARABLE FACIAL RECOGNITION DEVICE FOR THE VISUALLY IMPAIRED

[1]Ubochi, B. C., [2]Olawumi, A. E., [3]Shuiabu, A. O. and [4]Akingbade, K. F.

[1]Department of Electrical and Electronics Engineering, Federal University of Technology Akure, Ondo State, Nigeria
[2]Department of Mathematics, Physics, Electrical Engineering and Computer Science, Yokohama National University, Yokohama, Japan
[3]Department of Computer Engineering, Federal University of Technology Akure, Ondo State, Nigeria
[4]Department of Information and Communication Technology, The Federal University of Technology Akure, Ondo State, Nigeria
Corresponding author email:bcubochi@futa.edu.ng

Ubochi, B. C., Olawumi, A. E., Shuiabu, A. O. and Akingbade, K. F. (2024): A Wearable Facial Recognition Device for the Visually Impaired *Journal of Engineering and Engineering Technology /18(1), 48-55*

Received Date: 12.3.24 **Accepted Date: 7.5.24**

Abstract
Among other challenges, the visually impaired face problems related to a real-time human facial identification and recognition. In this paper, the design of a wearable real-time facial recognition device is presented. The system uses an acquired video feed that is processed to identify and recognize a face in the feed. An accurate face identification of a selected video frame was achieved using the Haar cascade algorithm. The system performs a similarity test using the cosine similarity function by which the system compares a detected face to the other faces saved in the database, and then proceeds to recognize the face in the video feed. Subsequently, an audio output of a name of the identified person is generated. The experiments performed with the head-mounted wearable system show that it can function effectively within a wide illumination level (11 lux to 1039 lux) but fails to detect a face in the image for distances greater than 90cm. Additionally, the recognition rate for both known faces and labelled faces in the wild obtained at different times of the day showed the highest recognition rate in the afternoon of 88% and 90%, respectively and the lowest recognition rate at night of 58% and 66%, respectively.

Keywords: Real-time face recognition, open CV, deep (Metric) learning, machine learning, wearable device

Introduction

Visual impairment is a common health problem in different age groups. According to the World Health organization (WHO), around thirty-six (36) million people from the world population are blind while 217 million suffer from moderate to severe visual impairment (WHO, 2020). The impact of visual impairment is felt across several sectors and can lead to a significant loss of productivity. For example, the time spent on caring for or assisting visually impaired persons was found to be related to the degree of visual impairment, with blind persons requiring the most assistance (Wójcik *et al*., 2016).

Action and cognition are the two major challenges that most visually impaired people experience in their daily lives. The corresponding solutions are navigation and image recognition (Wójcik *et al*., 2016). The difficulties faced by visually impaired people could potentially be alleviated by integrating navigation and image recognition technology into a wearable device (Jafri and Ali, 2014). Most of the assistive systems for the visually impaired collect visual information through various sensors, and then convert to auditory or tactile information (Akinsiku *et al*., 2020; Pissaloux *et al*., 2017). However, the research in assistive systems is more oriented towards navigation, thereby providing information about the location of the user and also guiding the user around obstacles and wet floors (Pissaloux *et al*., 2017; Patil *et al*., 2018). Most obstacle detection and avoidance systems rely on the use of ultrasonic sensors and feedback mechanisms either in the form of vibration or voice (Akinsiku *et al*., 2020; Yi and Dong, 2015).

Global Positioning System (GPS)-based receivers can also be used to improve the location accuracy, but this system can only be used for outdoor applications (Lapyko *et al*., 2014; Kuriakose, 2022). In addition to an ultrasonic sensor, Radio Frequency Identification (RFID) sensors, RGB-D camera and Kinect depth camera are also available as sources of visual information (Owayjan *et al*., 2015). However, RFID systems are expensive and require indoor installation in advance, and therefore are not adaptable to unfamiliar environments. On the other hand, images captured by RGB-D cameras and Kinect cameras are converted into deep images, which consume a lot of computing resources during processing. This makes image acquisition using these cameras unsuitable for low-power devices.

Generally, wearable devices that do not support object recognition are difficult to meet the daily needs of the visually impaired. They cannot know the object information of the target appearing in front of them, that is, whether it is a person or a thing, and what the specific name is. In view of this, the current challenges include the heavy computational requirement when performing face recognition (Krishna *et al*., 2005; Zhang *et al*., 2022), thus requiring external computers for image processing and detection; inability to recognize multiple faces (Utsumi *et al*., 2013) and a requirement for the camera to be still, in order to detect, capture and identify images, which is a problem with the Samsung Galaxy Gear that boasts of its wearable nature and low processing power (Bai *et al*., 2017; Adjabi *et al*, 2020).

This paper presents the design and development of a wearable real-time facial recognition device that is capable of facial image extraction, processing and recognition obviating the need to use external computing resources for image processing. Section 2 provides an insight into related works, while section 3 gives detailed explanation about the system architecture. In section 4, the findings obtained upon experimentation are discussed and section 5 concludes .

Related Work

Different approaches have been proposed to help visually impaired people in the navigation process. The most used are image processing-based techniques. These techniques include filtering, deep learning, and convolutional neural networks (CNN). Several research studies showcasing the application of these techniques can be found in the literature.

Jin *et al*. (2015) designed a system that provides tactile feedback to users through vibration patterns that corresponds to a particular person known by the user. Modified census transform (MCT) descriptor for feature extraction and L2-norm was used for real-time classification of images.

There are many cases where the classified image does not match any of the images contained in the database. Francis *et al*. (2020) developed an approach to helping visually impaired people keep up with the dynamic mix of people they might encounter in their daily life. This model is an assistive system with a text-to-speech approach on recognized objects that allows the user to decide if to include an unrecognized face in the device's database for future reference. It uses extracted Oriented Gradient Histogram (HoG) features and compares with the database which already has the extracted HoG feature. For a face image 4680 HoG features are extracted. If the input image is matched, it pronounces the person's name to the blind person by audio signal through a text-to-voice converter. Suppose the image is not matched, it asks whether to store it in the database. For this work, Balabolka software was used for text to voice conversion.

On a further approach, an electronic-based walking stick with text image recognition and echolocation capability was designed in Akinsiku *et al*. (2020). The system can guide the user to avoid obstacles and perform text recognition. Text-to-speech technology was used to provide aural feedback to the user. The work employed deep learning models which include EAST convolutional neural network and the Tesseract OCR model for text image detection and recognition respectively.

Researchers have also been burdened with the need for a system that not only detects but can also track. Mohana *et al*. (2018) used a combination of Google's Single Shot Detector (SSD) and MobileNets algorithm to achieve an efficient implementation of detection and object tracking. The authors used this algorithm to perform, in sequence, frame differencing, optical flow, and background subtraction for the object detection task. For object tracking, they make use of a sequence of detection to lock in on an object and track it using a python program for the algorithm, which is then implemented in Open CV.

Most object detection systems use Convolutional Neural Networks because of its effectiveness (Sinhal, 2017). In Szegedy *et al*. (2013), the use of Deep Neural Networks (DNNs) for object detection was investigated by researchers from Google Inc. In the paper, the authors not only address the problem of object detection using DNNs but also precisely localizing objects of various classes. They do this by using the network to identify every object which has been trained in a single frame. This ensures that every object which can be detected by the algorithm will not only be detected but will as well be localized by drawing a box around it. They presented a simple and yet powerful formulation of object detection as a regression problem to object bounding box masks.

Some researchers proposed a software application, Accessibility Bot, which utilizes the powerful face detection and recognition algorithms of Facebook (Zhao *et al*., 2018). It can scan the environment of the user and aurally produce the names of the people in front of its camera in real time. The developed system is not wearable since it uses the mobile phone's camera and was plagued with various usability issues such as difficulty in aiming the camera, low recognition accuracy etc.

System Overview and Architecture

The operation as shown in Figure 1 can be divided into two: (i) identification of face/s present in front of the user, and (ii) recognition of the face/s if they are present among the known or stored faces by producing a corresponding audio feedback of their names. The wearable facial recognition device was implemented using a Raspberry Pi with a camera module embedded in a wearable device (see Figure 2). The system uses a camera to transmit the video feed of the person to the Raspberry pi. The video feed is cut into 10 frames per second with the aim of reducing the computational cost. This frame rate is suitable for the use case since it is desired to detect slow-moving faces which would be interacting with the user of the software. Each frame potentially contains images of the faces to be identified in real time. The system then recognizes a face by matching a detected face in a frame with images in the database. Upon recognition, an audio output of the name of the identified person is received by the user. Different illumination conditions were used in testing the system and it produced good results.

The two major modes for face recognition are face verification which involves a one-to-one mapping of a given face against a known identity (e.g. is this the person?) and face identification that involves a one-to-many mapping for a given face against a database of known faces (e.g., who is this person?) (Wheeler *et al*., 2011).This work uses the concept of face identification for implementation of the proposed system. The system comprises of the hardware and software integrated together.

Figure 1: Face identification and recognition procedure

The hardware consists of a Raspberry Pi 4 Model B, 2GB RAM, fourth version of the low-cost single board computer which houses a Broadcom SoC (system-on-a-chip) containing an ARMv8-A72 quad-core CPU, a VideoCore VI 3D unit, and some video decoders and encoders. The Raspberry Pi has proven to be an immensely useful piece of hardware for numerous tasks. It is often used for educational purposes, and by makers (Koul et al., 2019). The 8-megapixel sensor based Raspberry Pi camera module can be used to take high-definition video, as well as still photographs (Chaithra and Vadiraja, 2015). It is attached via a 15 cm ribbon cable to the camera serial interface (CSI) port on the Raspberry Pi.

A 5 V DC fan and heat sinks were used to regulate the temperature of the system. An earphone is connected directly to the audio jack of the Raspberry Pi to send audio feedback to the user. SDHC MicroSD Card 32 GB is used to provide permanent storage for the Operating System and files.

Power is supplied to the Raspberry Pi using a mini power bank that is connected to the Pi using a USB cable. The software is written in Python. Python Face Recognition Library wraps around the distribution library thereby making our software program more compact. Imutils and Flite form the important parts of the software. Imutils software contains a series of functions which make basic image processing activities easier with OpenCV and Python. Flite is a powerful offline text-to-speech (TTS) software which provides decent voice audio and is relatively light-weight. The software maps the name of the face recognized by the system to an audio output for the user.

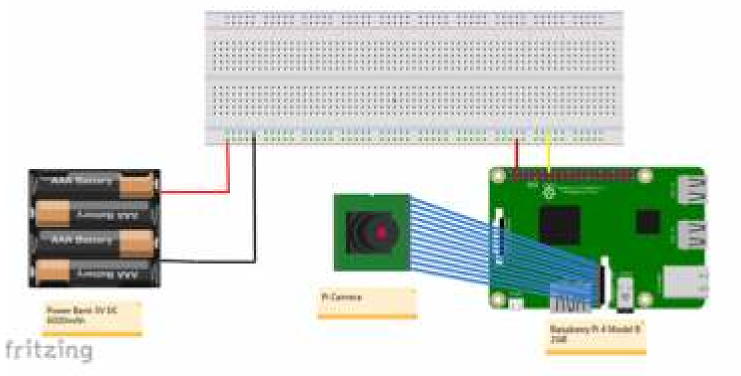

Figure 2: Layout diagram of the wearable facial recognition device

System Implementation and Operation

A series of images were captured using the camera, and then further resized to an image size of 320x240 in the window, in 10 frames per second. A face from the input image is identified by segregating the captured image into frames with each frame converted to gray image and then into a binary image. The Haar cascade algorithm, a localization algorithm which is known for high accuracy and fast real-time computational performance (Yustiawati *et al*., 2019), is used on the binary image to localize and identify a face from the background or the entire image. Not all the images were captured under ideal conditions (which is the case for many other data sets) to provide real-world conditions where various challenges will be present.

The most reliable and efficient way to recognize a face is to compute 128-d measurements (embedding) for each face and use a k-NN classification to compare distance between pairs of face embedding in order to determine whether a face belongs to same person or not (Schroff *et al*., 2015). Embedding are computer-generated features that help to characterize an object or image for recognition. This approach of using deep metric learning (Kaya and Bilge, 2019) for face recognition is more effective than traditional methods of training a network to recognize pictures or objects (Sinhal, 2017). After the faces have been localized from the images, if present, the window containing the faces would be extracted as a multidimensional vector, and then converted to a lower dimension, using unsupervised machine learning algorithm Principal Component Analysis (PCA).

Face Recognition

Figure 3 shows the processes involved in developing the face recognition system. The Haar cascade technique for face detection and image classification algorithms is used. The Raspberry Pi camera feeds into the application a real-time video of a person in front of it. The video feed is processed into multiple frames and passed into the face recognition system where each sframe is compared with faces stored in the database.

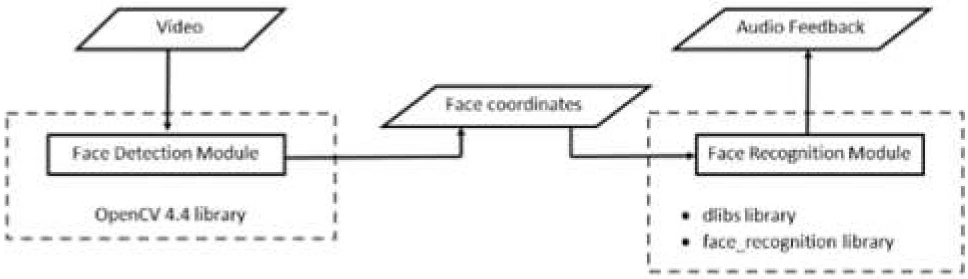

Figure 3: Face recognition system software framework.

The new vector obtained from dimensionality reduction is then compared to the images that are currently in the database (DB). A similarity test is performed, using the cosine similarity function expressed in Equation (1), and a 'similarity score' is calculated for all the images present in the DB. If the highest similarity is above a certain threshold (0.3/1), the corresponding saved name of the matching image is passed to the text-to-speech algorithm. The text to speech algorithm is responsible for transcribing the text version of the name into an audio format.where *D*1 is the vector representation of reduced frame, *D2* is the vector representation of the images in the DB.

The name of the person whose face is identified, as shown in Figure 4, is then transmitted through the audio feed connected to the audio jack of the Raspberry Pi.

$$similarity\,(D1, D2) = \frac{D1.D2}{\|D1\|\ \|D2\|} \qquad (1)$$

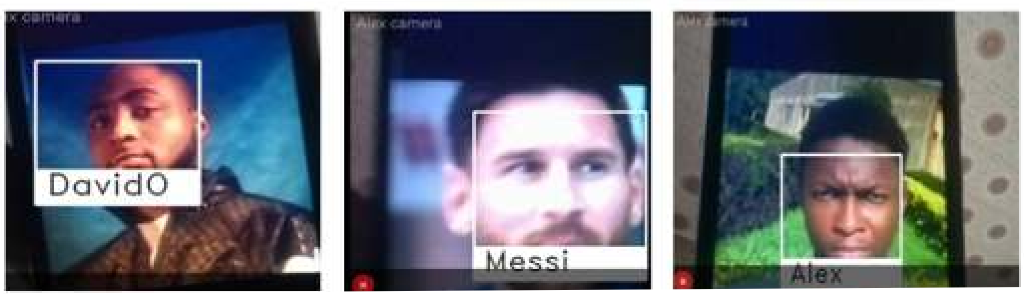

Figure 4: Face detection and recognition on selected images

Figure 5 shows the workflow of the face recognition pipeline. Live video stream from the Pi Camera relays the faces of the individual to the Raspberry Pi which contains the software. Upon reaching the Raspberry Pi, the video would be separated into several frames and a facial detection and localization algorithm, the Haar cascade classifier, would be run on every frame to identify if there is a face in the frame or if it is an empty image. The cosine similarity function is used to calculate a 'similarity score' for all the images present in the database. Through a deep-learning enabled image matching decision, the system determines which image is most similar to the one in the captured frame (i.e., the image in the database with the highest similarity that exceeds the threshold (0.3/1)). The saved name corresponding to the matching image is passed to the text-to-speech algorithm. The name is displayed and relayed via audio to the visually impaired individual.

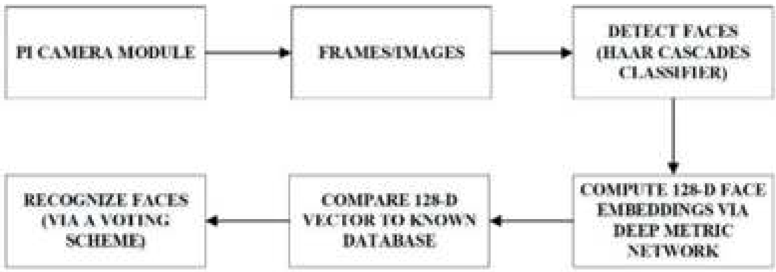

Figure 5: Workflow of the face recognition pipeline

Discussion

The experiments show that the wearable device can accurately recognize faces in the database within the wide range of illumination levels that characterize the differences between dawn and dusk, conditions that the user will most likely need the device. First, the device was tested over a range of illumination levels from as low as 11 lux to as high as 1039 lux. Figures (6-7) show instances where a face is detected at high illumination of 1039 lux and 49 lux, respectively.

(a) (b)

Figure 6: Face detection at a high illumination level (a) image at 1039 lux light intensity (b) face detected

Figure 7: Face detection at a high illumination level (a) image at 49 lux light intensity (b) face detected

Figure 9 shows the results of the performance and accuracy tests carried out at the different times of the day. It depicts the performance of the model on known faces (KF) and labeled faces in the wild (LFW). It can be depicted that the device worked more effectively during the day than at night. Figure 10 shows the developed facial recognition system. The complete system is lightweight, wearable and head-mounted resulting in a reduced camera movement and improved accuracy. The head-mounted system is a more convenient solution when compared to systems built on a walking stick (Akinsiku *et al*., 2020; Rahmad *et al*., 2020) and a preferred location for people who are blind or have vision impairments (Gamage *et al*., 2023).

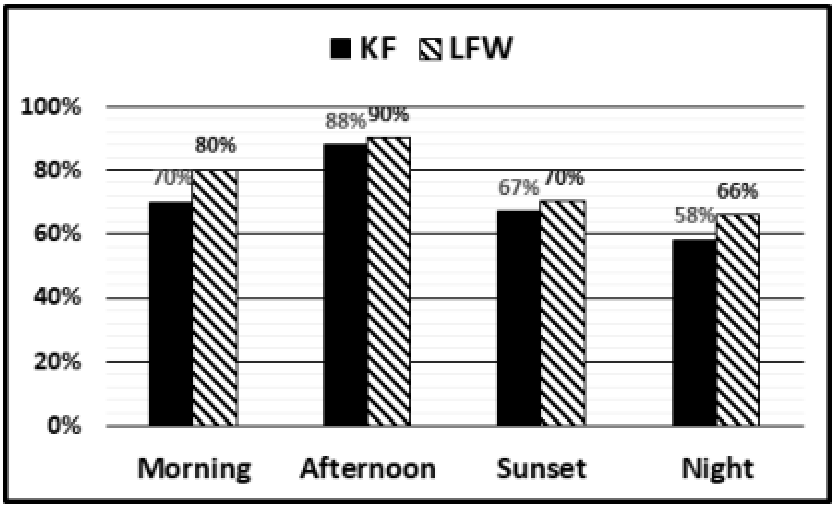

Figure 9: Recognition rate as a function of time of day

Figure 10: The developed wearable device installed on a hat

Conclusion

This paper has presented the design of a real-time wearable face recognition system for the visually impaired. The developed system uses OpenCV framework, and a pre-trained Deep Learning model built on a low-cost embedded AI device. This portable, wearable device extracts images captured by the camera module of Raspberry pi. A series of image processing operations are performed on the image. A similarity test using cosine similarity function is performed on the captured image and the saved images in the database. For similarity scores above a certain threshold (0.3/1), the name that corresponds to the matched image is passed to the text-to-speech algorithm. The results indicate that the device is suitable for applications having illumination conditions between 49 lux and 1039 lux. Also, if the illumination goes below 11 lux or the distance from the device is greater than 90 cm, the device fails to detect the image.

References

Adjabi, I., Quahabi, A., Benzaoui, A. and Taleb-Ahmed, A. (2020) Past, present, and future of face recognition: A Review. Electronics, 9 (8), 1-52.

Akinsiku, M., Ubochi, B. and Apena, W. (2020). Smart Walking Stick for the Blind with Text Image Recognition. Journal of Electrical Engineering, Electronics, Control and Computer Science, 6(22), 17–22.

Francis, G. A., Karthigaikumar, P. and Kumar, G. A. (2020). Face Recognition System for Visually Impaired People. Journal of Critical Reviews, 7 (17), 2760-2764.

Bai, J., Liu, D., Su, G., and Fu, Z. (2017). A Cloud and Vision-Based Navigation System Used for Blind People. Proceedings of the 2017 International Conference on Artificial

Intelligence, Automation and Control Technologies, 1-6.

Chaithra, P. and Vadiraja A. (2015). I Recognise U Raspberry Pi based Wearable Face Recognition Device with Image Tagging Facility for Blind. International Journal Of Engineering Research & Technology (IJERT) NCRTS, 3(27), 1-5.

Sinhal K. (2017). ResNet, AlexNet, VGGNet, Inception: Understanding various architectures of Convolutional Networks. https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/

Gamage, B., Do, T-T.,Price, N. S. C., Lowery, A., and Marriott, K. (2023). What do Blind and Low Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. Proceedings of the 25th International ACM SIGAACCESS Conference on Computers and Accessibility, 30, 1-21.

Jafri, R. and Ali, S. A. (2014). Exploring the Potential of Eyewear-Based Wearable Display Devices for Use by the Visually Impaired. 2014 3rd International Conference on User Science and Engineering (i-USEr), 119-124.

Jin, Y., Kim, J., Kim, B., Mallipeddi, R., and Lee, M. (2015). Smart Cane: Face Recognition System for Blind. Proceedings of the 3rd International Conference on Human-Agent Interaction, 145–148.

Kaya, M., and Bilge, H. Ş. (2019). Deep Metric Learning: A Survey. Symmetry, 11(9), 1066.

Koul, A., Ganju, S., and Kasam, M. (2019). Practical Deep Learning for Cloud, Mobile, and Edge: Real-World AI & Computer-Vision Projects Using Python, Keras & TensorFlow (Ist). O'Reilly Media, 679-680.

Krishna, S., Little, G., Black, J., and Panchanathan, S. (2005). A Wearable Face Recognition System for Individuals with Visual Impairments. Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility, 106–113.

Kuriakose, B., Shrestha, R., and Sandnes, F. E. (2022). Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review. IETE Technical Review, 39 (1), 3-18.

Lapyko, A., Tung, L.-P., and Lin, B.-S. (2014). A cloud-based outdoor assistive navigation system for the blind and visually impaired. 2014 7th IFIP Wireless and Mobile Networking Conference, WMNC 2014, 1–8.

Mohana C., Ayush J. and Harsh J. (2018). Real Time Object Detection and Tracking Using Deep Learning and OpenCV. The 2013 International Conference on Information and Knowledge Engineering (IKE '13), 1305–1308.

World Health Organization. (2020). WHO Facts Sheet 2020 on Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

Owayjan, M., Hayek, A., Nassrallah, H., and Eldor, M. (2015). Smart Assistive Navigation System for Blind and Visually Impaired Individuals. 2015 International Conference on Advances in Biomedical Engineering (ICABME).

Patil, K., Jawadwala, Q., and Shu, F. C. (2018). Design and Construction of Electronic Aid for Visually Impaired People. IEEE Transactions on Human-Machine Systems, 48(2), 172–182.

Pissaloux, E., Velázquez, R., and Maingreaud, F. (2017). A New Framework for Cognitive Mobility of Visually Impaired Users in Using Tactile Device. IEEE Transactions on Human-Machine Systems, PP, 47 (6), 1040-1051.

Rahmad, C., Asmara, R. A., Putra, D. R. H., Dharma, I., Darmono, H., and Muhiqqin, I. (2020). Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection. IOP Conference Series: Materials Science and Engineering, 732(1), 12038.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 7, 815–823.

Szegedy, C., Toshev, A., and Erhan, D. (2013). Deep Neural Networks for Object Detection. In Advances in Neural Information Processing Systems 26. Curran Associates, 2553-2561.

Utsumi, Y., Kato, Y., Kunze, K., Iwamura, M., and Kise, K. (2013). Who are you? - A wearable face recognition system to support human memory. 4th Augmented Human International Conference, AH 2013, 150–153.

Wheeler, F.W, Liu, X. and Tu P. H. (2011). Face Recognition at a Distance in Handbook of Face Recognition. Springer London, 353-381.

Wójcik, W., Gromaszek, K., and Junisbekov, M. (2016). Face Recognition: Issues, Methods and Alternative Applications, Face Recognition- Semisupervised Supervised Classification Subspace Projection and Evaluation Methods, 7-28.

Yi, Y., and Dong, L. (2015). A design of blind-guide crutch based on multi-sensors. Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2288–2292.

Yustiawati, R., Husni, N. L., Evelina, E., Rasyad, S., Lutfi, I., Silvia, A., Alfarizal, N., and Rialita,

A. (2019). Analyzing of Different Features Using Haar Cascade Classifier. Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018, 129–134.

Zhang, P., Zhao, F., Liu, P., and Li, M. (2022). Efficient Lightweight Attention Network for Face Recognition. IEEE Access, 10, 31740-31750.

Zhao, Y., Wu, S., Reynolds, L., and Azenkot, S. (2018). A Face Recognition Application for People with Visual Impairments: Understanding Use Beyond the Lab. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–14.