International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 106-112, November-2023

https://doi.org/10.47001/IRJIET/2023.711015

Visual Assistance Glasses for Visually Impaired

[1]S.M.D.N.S.Senarath, [2]P.M.Kekulandara, [3]K.D.H.N.D.A.T.Divarathna, [4]Y.M.W.H.C.Samarasekara

[1,2,3,4]Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

Abstract **- Technology has led to advances that improve the daily lives of visually impaired people. B-Fit Glasses, smart glasses with sophisticated machine learning algorithms and sensors, cameras, microphones, and processing units orchestrated by a server, are the driving forces behind this technological evolution. The glasses automatically detect and alert users to surrounding obstacles, improving spatial awareness. Their skills go beyond this. Voice commands can transition to facial recognition mode to recognize and name familiar people or document recognition mode to read printed or handwritten information. Currency denomination detection and verbalization are other notable features. Onboard processors rapidly analyze and store data, ensuring real-time responsiveness and accuracy. The comprehensive, user-centric B-Fit Glasses give the visually impaired an enriched, autonomous, and safer method to interact with their surroundings, revolutionizing wearable assistive technology.**

Keywords: B-Fit, Smart Glasses, Machine Learning Algorithms, Visually Impaired, Voice Detection, Currency Detection, Face Recognition, Document Recognition, Obstacle Detection, Sign Detection, Health.

I. INTRODUCTION

Globally, 285 million individuals suffer from vision impairments, with 39 million classified as blind. In the US alone, there are 7 million people with visual challenges [2]. These statistics underscore the pressing need for technological advancements to bolster mobility and social connections for the visually impaired. While Braille displays, screen readers, and magnifiers remain crucial, there's a rising interest in smart glasses integrated with machine learning (ML) and state-of-the-art hardware [1]. These next-gen glasses are equipped with sensors, cameras, and microphones and leverage machine learning to execute critical functions. Tasks such as object identification and distance approximation are pivotal for navigation, and facial recognition is central to social interactions. These ML algorithms are finely tuned with extensive datasets of visual and auditory cues, leading to a technology that transcends mere" smartness" to deliver profound insights. Furthermore, these glasses employ optical character recognition (OCR) to audibly present text [5]. This capability empowers those with visual impairments to independently explore printed resources like books and documents and even discern currency denominations. The integration of voice commands augments the user's experience, with vocal prompts negating the need for touch-based interactions and ensuring the glasses remain functional when hands-on interactions become challenging [10]. When paired with smartphones and processors, these smart glasses elevate data processing capabilities. Such seamless integration enables rapid data exchanges and tailored feedback based on users' histories and preferences. As we witness an era of swift digital progression, these ML-driven smart glasses with enhanced hardware herald promise. They foreshadow a future where visual constraints won't hinder movement or social interactions, fostering greater independence for those with visual impairments.

II. LITERATURE REVIEW

Recent progress in artificial intelligence, computer vision, and the miniaturization of hardware has fueled the enhancement of assistive devices for the visually impaired [7]. From canes to advanced electronic travel aids (ETAs), assistive solutions have undergone significant evolution. This review delves into the advancements in obstacle detection, facial recognition, voice command detection, and document recognition that paved the way for B-Fit glasses.

2.1 Obstacle detection and sign recognition

While traditional tools like white canes and guide dogs have served the visually impaired well, their adaptability and convenience remain limited, as highlighted by [14]. This prompted the development of technological travel aids. A notable advancement was proposed by [4], using ultrasonic waves for obstacle detection [9]. A major stride was made when deep learning techniques were merged with ETAs, as demonstrated by [9]. Utilizing convolutional neural networks (CNNs), they greatly enhanced real-time obstacle detection. This trajectory underscores the relentless quest for optimizing assistive tools through technology, particularly deep learning, for safer mobility and greater autonomy.

2.2 Face Recognition Technology

The inability of visually impaired individuals to recognize acquaintances can lead to social isolation, as outlined by [6]. Addressing this requires reliable technology. Facial recognition has increasingly incorporated deep learning, analyzing more than mere facial features for enhanced

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 106-112, November-2023

https://doi.org/10.47001/IRJIET/2023.711015

precision. [11] were pioneers with their Face Net model, setting new benchmarks in the field. By embedding such models in wearables, real-time recognition of acquaintances becomes feasible, trans-forming social interactions.

2.3 Voice command and Detection

Voice command systems offer a more natural, hands-free interface with technology, enhancing the experience for the visually impaired. Traditional touch-based systems may be challenging for this demographic, as observed by [2]. long-short-term This, in turn, facilitates more interactive and user-tailored assistive devices.

2.4 Document, Bill and Currency Recognition

Reading printed information, like bills or currencies, has always been a challenge for the visually impaired. Though tools like braille provided some relief, they weren't holistic solutions, as mentioned by [12]. However, the advent of Op-tical Character Recognition (OCR) systems such as Tesseract brought about a paradigm shift. With the infusion of machine learning and deep learning, OCR capabilities have expanded significantly, as noted by [5]. Simultaneously, advancements in banknote recognition, as illustrated by [15], have empowered the visually impaired to handle money more autonomously.

2.5 Challenges in Integration of Technology

While the progression in individual technologies is commendable, integrating them into a single device like the B-Fit Glasses presents challenges. The need for instantaneous responses is critical, as underlined by [10] and [3]. Furthermore, such devices must adeptly multitask, swiftly toggling between different functionalities, a complexity highlighted by [8]. Factors like rapid battery consumption and ergonomics, as discussed by [13], further complicate the integration. Yet the pursuit continues, driven by the aim to furnish the visually impaired with an all-encompassing tool for enhancing their daily lives.

III. METHODOLOGY

3.1 Data Collection

For the B-Fit Glasses' success, comprehensive data collection is essential. It's crucial to gather a wide variety of images and videos capturing common and dynamic obstacles, from uneven pavements and poles to moving vehicles and people. The collection should also include crucial navigation and safety signs like stop signs and traffic lights. Additionally, the dataset should encompass the faces of people frequently encountered by the user and cater to diverse user voices for effective voice recognition. A robust database reflecting real world textual and monetary elements is also vital.

3.2 Data Processing

To optimize recognition models, images undergo several quality enhancement steps regardless of the capture conditions. Techniques include lighting correction, noise reduction, and image enhancement for clearer subjects. For face recognition, the process entails face alignment and cropping, illumination normalization, and histogram equalization to highlight distinct facial features. To focus solely on a user's voice command, the system employs noise cancellation and voice enhancement, ensuring clear and accurate voice command recognition. Lastly, to improve text and feature extraction, images are re-scaled for uniform resolution, converted to grayscale for simpler processing, and enhanced for contrast, making text more visible and distinguishable.

3.3 Feature Extraction

Feature extraction isolates essential details from images, videos, or audio to enhance recognition accuracy. For general objects, this involves shape, size, texture, and color analyses, helping identify items like poles, signs, and traffic signals. When processing facial data, techniques dlib facial landmarks, analyze dimensions between these landmarks, and study skin textures and tones to distinguish individuals. Voice recognition relies on evaluating pitch, frequency, and tone, and establishing unique voice prints for users. Lastly, for text and currency discernment, methods like edge detection, corner recognition, and identification of visual patterns like watermarks ensure accurate content recognition.

3.4 Recognition

The B-Fit Glasses utilize advanced methods for efficient recognition. Using CNNs, they process images to detect obstacles and signs. They employ models like FaceNet for facial identification and alert users via audio when known faces are spotted, with customization options. Voice prompts are handled by neural models like RNNs and are turned into actionable commands, with text-to-speech technology facilitating user communication, adaptable to surroundings. OCR engines transform visual text into readable formats, emphasizing spatial text positioning and currency value recognition. The culmination of these techniques transforms the B-Fit Glasses into an interactive guide for the visually impaired.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 106-112, November-2023

https://doi.org/10.47001/IRJIET/2023.711015

3.5 User Alert Mechanism

The final step of the detection process involves promptly and efficiently notifying the user. This is achieved through vibration-based alerts within the glasses or a wearable, with the vibration's characteristics denoting the obstacle's closeness and type. Voice alerts further specify obstacles, like warning of a nearby staircase. Users can also tailor these notifications to their liking, ensuring optimal user-friendliness.

3.6 Object Detection and Obstacle Recognition

The primary aim of this component of the research is to ensure the safety of visually impaired individuals by offering real-time detection and recognition of potential obstacles and important signs they might encounter. This detection mechanism facilitates better spatial awareness and offers guidance, thereby enhancing their independent mobility.

3.7 Face Recognition

Pre-trained Deep Learning Models: Leverage models like FaceNet or VGG-Face which have been previously trained on extensive facial datasets. These models have architectures fine-tuned for facial recognition tasks.

Convolutional Filter Map:

$$a_{1,m}^{(k)} = \text{max-pool}\left(\max\left(0, W_{l-1,l}^{(k)} * h_{1,(l-1)} + b_l\right), 2\right) \quad (1)$$

$$a_{2,m}^{(k)} = \text{max-pool}\left(\max\left(0, W_{l-1,l}^{(k)} * h_{2,(l-1)} + b_l\right), 2\right) \quad (2)$$

Where:

$a_{1,m}^{(k)}$: the output value of the k-th filter for the first twin image after max-pooling.

$a_{2,m}^{(k)}$: the same for the second twin image.

$W_{l-1,l}^{(k)}$: matrix of numbers (weights) that the convolutional filter uses to process the previous layer's feature maps.

h1, (l−1): hidden vectors from the previous layer for the first twin image.

h2, (l−1): hidden vectors from the previous layer for the second twin image.

bl: number called bias that is added to the calculation.

Predicting Similarity:

$$p = \sigma\left(\sum_j \alpha_j \cdot \left|h_{1,L-1}^{(j)} - h_{2,L-1}^{(j)}\right|\right)$$

Where:

p: prediction score.

σ: sigmoid activation function, which squashes values between 0 and 1.

αj: parameters learned during training that weigh the importance of different components of the distance between the images.

$h_{1,L-1}^{(j)}$ and $h_{2,L-1}^{(j)}$: the hidden vectors of the two images after processing by the last hidden layer.

$\left|h_{1,L-1}^{(j)} - h_{2,L-1}^{(j)}\right|$: the absolute difference between the two vectors.

3.8 Voice Command and Detection

The aim of this component of the research is to make the glasses technologically adept in recognizing and acting upon voice commands, offering a hands-free experience for visually impaired users. Additionally, it ensures that the glasses track the user's voice, thereby enhancing interactivity and ease of use. This voice command and detection research facet ensures that the B-Fit Glasses are not just passive visual aids but interactive tools that users can communicate with and control using just their voice, making the technology both empowering and user-friendly.

3.9 Document, Bill and Currency Recognition

In this research phase, the B-Fit Glasses evolved from merely assisting with vision to actively helping users decipher documents, bills, and currency notes. This enhancement is crucial, as it grants visually impaired users the ability to independently navigate their financial and textual surroundings, fostering a greater sense of independence, awareness, and empowerment.

- Document Recognition: Gray Scale Conversion

$$\text{Gray\_intensity} = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (4)$$

- Bounding Box Manipulation:

Given four points (x1, y1), (x2, y2), (x3, y3), and (x4, y4), the array of points is defined as:

$$\text{points} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{bmatrix}$$

- Money Note Recognition: L2 Distance (Euclidean Distance)

Given two descriptor vectors, x and y, the L2 distance (Euclidean Distance) between them is defined as

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 106-112, November-2023
https://doi.org/10.47001/IRJIET/2023.711015

$$D(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2}$$

Where xi is the value of the i-th element in the first descriptor vector, and yi is the value of the i-th element in the second descriptor vector.

- Intersection given two histograms represented by descriptors, x and y, their intersection is defined as:

$$I(\mathbf{x}, \mathbf{y}) = \sum_i \min(x_i, y_i)$$

Where:

- xi is the value of the i-th bin in the histogram of the first descriptor.
- yi is the value of the i-th bin in the histogram of the second descriptor

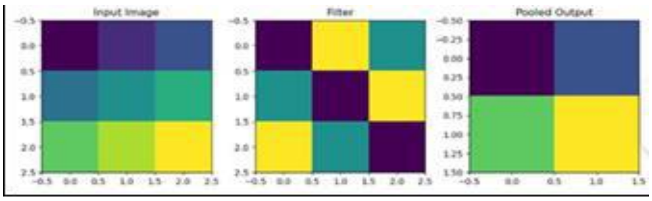
IV. RESULTS AND DISCUSSION

The 3D model employed in this research likely contributed to the high accuracy rate by allowing the system to capture depth and perspective, crucial factors in recognizing faces, object detection, and document detention, from different angles and in varying conditions.

4.1 Face Recognition

Twin networks are employed in the Siamese CNN architecture, which uses facial recognition to concurrently evaluate and compare two images. ReLU activation functions at the beginning and sigmoid at the end, convolutional processes, and max pooling are used in these twin models, which are symmetrical. When learning image similarities rather than representations, the network uses fully connected layers to compute picture similarity. Identical-labeled image pairs are used in the training process, and backpropagation is used to update the network parameters based on learning gradients. Bayesian optimization aids in determining the ideal parameters, while batch training boosts productivity and stability. To avoid overfitting, real-time modifications are performed to the learning rate, momentum, and regularization strength. The experiment used a video clip to identify faces, and it did it with astounding 100% accuracy. Convolutional filters are used to highlight picture features, and after that, processing is done to do recognition. The Siamese CNN is a top option.

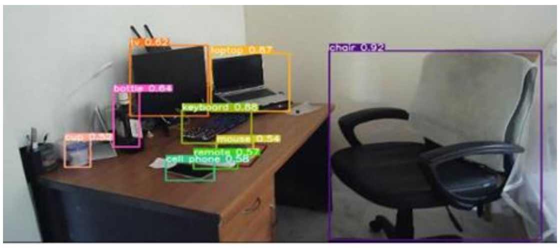
This process is broken down into two primary steps to emphasize their roles in feature extraction. Initially, the con-volitional filter is applied. This filter, a mathematical computation, is designed to accentuate specific features in an image, bringing out characteristics that are most pertinent for the subsequent steps.



After the convolutional filtering, the max-pooling operation takes precedence. Explicitly termed the" max-pool", this operation meticulously selects the highest value from a set of values. This process ensures that the most dominant feature from the set is retained, potentially reducing dimensionality and computational requirements. For a comprehensive understanding, it's noteworthy that these calculations are performed for every k-th filter and are applied at every m-th position within the feature map, ensuring a detailed and nuanced extraction of features from the original image.

4.2 Object Detection

The object detection model detected many room objects with amazing adaptability. As shown below, the model could identify laptops, computers, televisions, bottles, cups, cell phones, remotes, keyboards, mice, and seats. This vast range of object detection highlights the model's potential to be integrated into B-Fit Glasses, giving visually impaired people an enhanced feel of their surroundings and potentially revolutionizing independent movement and navigation.



International Research Journal of Innovations in Engineering and Technology (IRJIET)

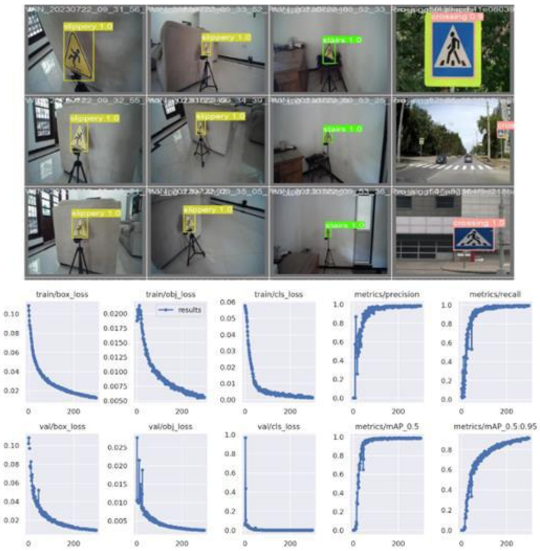
ISSN (online): 2581-3048

Volume 7, Issue 11, pp 106-112, November-2023

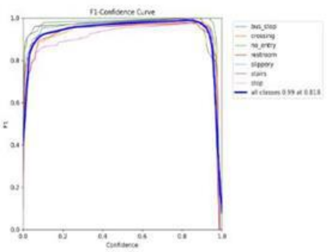
https://doi.org/10.47001/IRJIET/2023.711015

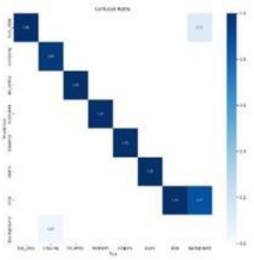
4.3 Sign Detection in Daylight

The model showcased impressive capabilities for detecting and distinguishing various signs in broad daylight. Notably, signs like slippery zones, stairs, crossings, and other specific indicators like bus stops, no-entry zones, and restrooms were detected with high confidence.



The Fi-Confidence curve and Confusion Matrix for the designated classes - bus stop, crossing, no-entry, restroom, slippery, stairs, and stop - demonstrated a notable score of 0.99 at a threshold of 0.818. This result underscores the high reliability and accuracy of the model in recognizing and distinguishing between these classes.

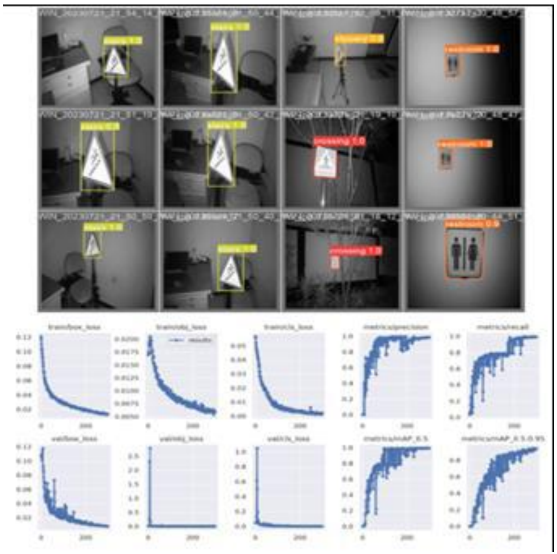




4.4 Sign Detection in Night Vision

Even under low-light conditions, our model demonstrated a robust ability to recognize signs, including slippery zones, stairs, and crossings. The angle of the sign did not deter the model's performance, which is crucial for real-world applications.

B-Fit glasses are able to be used for both day and night assistance, however adding night vision increases hardware requirements and necessitates careful power management. For the model to operate at its best, handling real-world problems like bright city lights and moving cars at night must also be included in the training process.

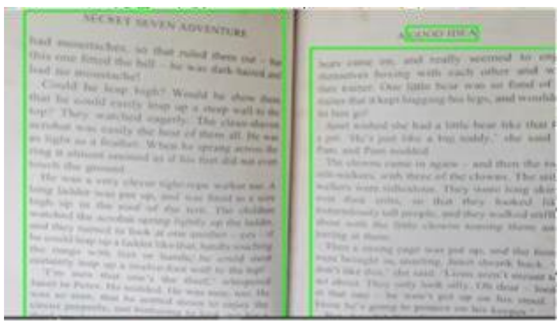


4.5 Voice Detection

The voice detection model designed for B-Fit Glasses is highly accurate in recognizing various user commands even in noisy real-world environments. It effectively distinguishes user voices from background noise, ensuring the proper execution of commands. However, there are some challenges to address: Environmental Noise and Adaptability: The model performs well in tests but may face difficulties in dynamic outdoor situations due to background noise. Implementing adaptive noise suppression technology that learns from the environment could enhance speech clarity. Multiple keywords for a command: The voice detection system recognizes commands by different keywords with similar meanings. While the prototype model was trained on the dataset, ongoing feedback mechanisms are needed to continually improve its accuracy across keywords and accents. Integration with Other Systems: It's crucial for the voice detection system to seamlessly work with other features of B-Fit Glasses, especially in cases where the glasses provide audible feedback after identifying objects or signs. To prevent overlapping commands or feedback, the voice detection system should pause or reduce its sensitivity during these interactions.

4.6 Document, Bill, and Currency Recognition

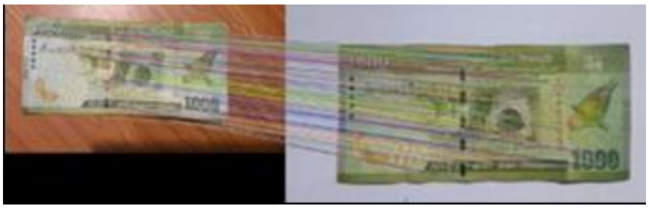
• Documentation Recognition



ISSN (online): 2581-3048
Volume 7, Issue 11, pp 106-112, November-2023
https://doi.org/10.47001/IRJIET/2023.711015

The document recognition system successfully processed a variety of books and was proficient at identifying chapters, topics, and subtopics within the text. Through rigorous testing, the system displayed a consistent ability to break down content hierarchically, providing organized auditory feedback to the user.

• Currency Recognition



The money recognition model was designed to identify various currency notes with high precision. Throughout testing, it showcased an ability to recognize and audibly inform the user about different denominations across various global currencies.

V. CONCLUSION

B-Fit Glasses research has successfully delved into the transformative potential of wearable technology to aid the visually impaired. Through the integration of various AI models and modules, this study demonstrated the tangible benefits such devices can offer in enhancing the daily lives of users. From face recognition to bill deciphering, the technological advancements were not only effective but also indicated a significant improvement over traditional aid. The synergy of multiple technologies in a singular wearable platform showcased the impressive strides that interdisciplinary research can achieve.

VI. FUTURE WORK

In considering the future work for the B-Fit Glasses, it's crucial to expand on the promising initial results through extended user testing with diverse groups, ensuring broader real-world applicability and refining the device. The rising prevalence of Augmented Reality (AR) technology offers potential avenues for its integration, which could significantly enhance navigation and provide immersive feedback to the visually impaired. There's also a noted necessity to expand the glasses' capability for wider sign recognition, accounting for the dynamic urban environment and including a broader range of international and local signs. Feedback indicates a need for improving the ergonomics of the glasses, making them more comfortable, durable, and aesthetically appealing. The introduction of collaborative features could offer carers or family members a way to interact or monitor the device remotely, enhancing safety and communication. A forward-thinking approach would incorporate advanced adaptive learning, enabling the B-Fit Glasses to evolve based on user habits thus offering a more tailored experience. Finally, with a global shift towards eco-consciousness, it would be prudent to delve into sustainable materials and energy-efficient components for the device, aligning with a more sustainable future.

ACKNOWLEDGEMENTS

We want to express our sincerest gratitude to our supervisor Dr. Dharshana Kasthurirathne, co-supervisor Ms. Hansi De Silva for their unwavering support and guidance during this journey, especially during the challenging periods. We would like to take this chance to express Our sincere gratitude to Ms. Sanjeevi Chandrasiri, Ms. Narmada Gamage, and Ms. Karthiga Rajendran for their suggestions and comments, which helped us to successfully complete this research. Finally, we would like to express Our sincere gratitude to Our families for their tremendous support and encouragement.

REFERENCES

[1] D. Brown et al., "User interfaces for visually impaired individuals: A comparative analysis," Journal of Visual Impairment and Accessibility, vol. 108, no. 6, pp. 509-520, 2014.

[2] JA. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 6645-6649, 2013.

[3] P. Hansen, R. Jones, and M. Tuchman, "Importance of real-time feed-back in wearable devices for the visually impaired," Visual Disability Research, vol. 28, no. 1, pp. 15-23, 2016.

[4] A.Johnson and P. Roberts, "Progress in wearable electronics and their role in modern society," Journal of Modern Electronics, vol. 12, no. 4, pp.77-89, 2015.

[5] K. Jaidka et al., "Deep learning in optical character recognition: Challenges and opportunities," Int. J. Advanced Computer Science and Applications (IJACSA), vol. 11, no. 3, pp. 42-49, 2020.

[6] A.Khan et al., "The social implications of face recognition systems: Exploring the impact on visually impaired individuals," Journal of Visual Impairment and Blindness, vol. 112, no. 5, pp. 583-588, 2018.

[7] Y. Kim and P. Gupta, "Battery life in AI-powered wearables: A contemporary review," Electronics Today, vol. 57, no. 3, pp. 143-149, 2020.

[8] T. Lin, M. Maire, and S. Belongie, "Deep learning in wearable devices: Opportunities and challenges," in Proc. 14th Wearable Tech Symposium, 2018.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 106-112, November-2023

https://doi.org/10.47001/IRJIET/2023.711015

[9] Y. Liu et al., "Deep learning for real-time obstacle detection and safe navigation for the visually impaired," Digital Signal Processing, vol. 70, pp.75-86, 2017.

[10] R. Patil and U. Kulkarni, "Challenges and opportunities in AI-integrated wearables," Journal of Wearable Technologies, vol. 3, no. 2, pp. 45-51, 2019.

[11] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 815-823, 2015.

[12] R. Smith, "An overview of the Tesseract OCR engine," in Proc. Ninth Int. Conf. Document Analysis and Recognition (ICDAR), vol. 2, pp. 629–633, 2007.

[13] J. Turner, H. Van der Loos, and K. Salem, "The ergonomics of wearable computing," Int. J. Human-Computer Studies, vol. 24, no. 6, pp. 486-497, 2017.

[14] R. Velazquez, "Wearable assistive devices for the blind," in Wearable and Autonomous Biomedical Devices and Systems, 2010.

[15] B. Zhou et al., "Currency recognition using a smartphone: It's all about the blur," IEEE Trans. Image Processing, vol. 25, no. 6, pp. 2718-2730, 2019.

AUTHORS BIOGRAPHY

[1]S.M.D.N.S.Senarath,
Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.

P.M.Kekulandara,
Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.

K.D.H.N.D.A.T.Divarathna,
Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.

Y.M.W.H.C.Samarasekara,
Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.

Citation of this Article:

S.M.D.N.S.Senarath, P.M.Kekulandara, K.D.H.N.D.A.T.Divarathna, Y.M.W.H.C.Samarasekara, "Visual Assistance Glasses for Visually Impaired" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 11, pp 106-112, November 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.711015

****

