IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

VISIONBRIDGE: Enabling Independence through Object, Face and Currency Recognition for the Blind

Mrs. J. Fahamitha[1], K. Kalaivanan[2] , S. Karikalan[3]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science Engineering [2, 3,]
Dhanalakshmi Srinivasan University, Samayapuram, Thiruchirapalli, Tamilnadu, India

Abstract: *Navigating daily life poses significant challenges for blind and visually impaired individuals, particularly in identifying obstacles, recognizing familiar faces, and handling currency transactions. These everyday tasks often require external assistance, leading to a dependency on others and a reduced sense of autonomy. Traditional tools like white canes and guide dogs offer limited functionalities and cannot address the dynamic and complex challenges faced by visually impaired individuals in real-time environments. This study introduces an innovative interface designed to empower blind and visually impaired individuals by enhancing their independence and safety. The proposed system integrates advanced AI powered functionalities such as face detection, obstacle detection, and currency recognition. By utilizing real-time image capturing and processing, the system provides users with immediate, context sensitive audio feedback to assist them in navigating their surroundings and performing essential tasks. The development of such a system is necessary to bridge the gap left by existing assistive technologies, which are often limited in functionality, integration, or affordability. By leveraging advancements in artificial intelligence, computer vision, and wearable technology, the proposed solution addresses critical challenges, fostering greater autonomy and confidence for visually impaired individuals in their daily lives.*

Keywords: blind and visually impaired individuals

I. INTRODUCTION

For individuals who are blind or visually impaired, navigating the world presents numerous challenges. Everyday tasks such as recognizing familiar faces, avoiding obstacles, and managing currency transactions often require assistance from others. This dependency on external support reduces their autonomy and confidence in performing daily activities, leading to significant limitations in their social and personal lives. While traditional assistive tools such as white canes and guide dogs have provided valuable help, they come with limitations in functionality. These tools are primarily focused on physical navigation and safety but are unable to offer real-time information about the surrounding environment or identify specific objects, faces, and currency denominations. The inability to access comprehensive, real-time information limits the independence of visually impaired individuals.Currency handling, which is a critical part of daily life, can also be a significant challenge for blind individuals, leading to financial management difficulties. As a result, visually impaired individuals are often compelled to rely on others for assistance in these essential tasks. AI-powered solutions can combine multiple functionalities, such as face detection, obstacle detection, and currency recognition, into a single device.

II. LITERATURE SURVEY

2.1 DOMAIN RESEARCH

Javed, Sajid, et al. (2018) This paper proposes a robust method for moving object detection in complex scenes using a spatiotemporal structured-sparse Robust Principal Component Analysis (RPCA). The approach integrates spatial and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

temporal constraints into RPCA to enhance detection performance in scenarios with dynamic backgrounds, illumination changes, and camera jitter. The method models the background as a low-rank component and the moving objects as sparse components while preserving spatial structure. Spatiotemporal grouping helps to maintain the coherence of object appearance over time, making the method ideal for video surveillance applications. The algorithm is validated on several challenging datasets, demonstrating high accuracy and robustness. Its performance is compared against other state-of-the-art techniques, showcasing superior object detection capabilities.

Ren, Shaoqing, et al. (2015) This foundational paper introduces Faster R-CNN, a deep learning framework that significantly advances real-time object detection. It integrates a Region Proposal Network (RPN) with a Fast R-CNN detector, creating a unified, end-to-end trainable system. The RPN efficiently generates high-quality region proposals by learning to propose object bounds and objectness scores. This drastically reduces the computational cost associated with traditional selective search methods. Faster R-CNN improves detection speed and accuracy, making it suitable for real-time applications. It utilizes shared convolutional features between the RPN and the detection network, enhancing processing efficiency. Extensive experiments show that Faster R-CNN achieves state-of-the-art accuracy on PASCAL VOC and MS COCO benchmarks. The framework is flexible and adaptable to various CNN architectures. Its introduction paved the way for subsequent innovations in real-time object detection.

He, Kaiming, et al. (2016) The authors propose ResNet (Residual Network), a deep learning architecture that enables training of extremely deep neural networks by introducing residual learning. The core innovation is the use of skip connections (identity shortcuts) that bypass one or more layers, effectively mitigating the vanishing gradient problem. This architecture allows the construction of networks with over 100 layers, significantly improving accuracy in image recognition tasks. ResNet achieved groundbreaking results on the ImageNet and COCO datasets. It won the ImageNet 2015 classification task with a top-5 error rate of 3.57%.

Zhang, Han, et al. (2016) This work presents SPDA-CNN, a novel convolutional neural network designed for fine-grained visual recognition through semantic part detection and abstraction. The method unifies two tasks: detecting discriminative object parts and abstracting high-level features. This dual-stage approach significantly improves recognition accuracy in domains like bird species or car model classification, where subtle differences distinguish classes. The architecture uses a part detection module to localize semantic parts, and then feeds part-aware features into the abstraction module. This hierarchical processing captures fine-grained details while preserving contextual relevance. The network is trained end-to-end, allowing joint optimization of part localization and recognition. Extensive experiments on benchmark datasets such as CUB-200 and Stanford Cars demonstrate the superiority of SPDA-CNN over existing methods.

Ouyang, Wanli, et al. (2015) The authors propose DeepID-Net, a deep convolutional neural network architecture designed to enhance object detection accuracy. The model incorporates deformable part models (DPM) within the deep network to better capture object variations and poses. It introduces a novel deformation-constrained pooling layer that models part deformations during detection. The architecture also leverages multi-stage training, allowing the network to progressively refine features and improve classification. DeepID-Net is evaluated on the ImageNet object detection challenge and achieves competitive results. Its flexible design enables adaptation to different object categories and detection tasks. The paper highlights how combining domain-specific knowledge with CNNs enhances detection robustness.

III. SYSTEM REQUIREMENTS

3.1 FUNCTIONAL REQUIREMENTS

This section outlines the functional requirements essential for implementing and evaluating the **AI-powered assistive system** for visually impaired individuals, integrating face recognition, obstacle detection, and currency identification. The system should function effectively in real-time environments, providing accurate feedback and enhancing user independence and safety.

3.1.1 Face Recognition

The system must incorporate a face recognition module utilizing the **Grassmann algorithm**, capable of differentiating between known and unknown individuals. This functionality should enable the system to identify familiar faces and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

announce the person's name audibly when detected. The system must store and manage a database of registered faces, ensuring accurate and fast recognition even in dynamic and crowded environments.

3.1.2 Object Detection

The system must integrate **YOLO (You Only Look Once)** for object detection, enabling real-time identification of obstacles such as vehicles, people, or other objects in the user's surroundings. The system should detect and track moving objects, providing audio feedback to guide the user to navigate left, right, or stop to avoid potential collisions. This module must function effectively in both dynamic and static environments.

3.1.3 Currency Recognition

The system must use Convolutional Neural Networks (CNN) to identify and classify various currency denominations. The user should be able to hold a currency note in front of the wearable camera, and the system must audibly announce the recognized denomination. The currency recognition module must handle different currencies and provide high accuracy even in low-light conditions or with partially worn notes.

3.1.4 Real-Time Audio Feedback

The system must provide immediate, context-sensitive audio feedback to the user. Each detected object, face, or currency denomination must be communicated clearly and quickly, ensuring the user can make informed decisions while navigating or performing tasks. The system should have adjustable volume and tone settings to suit individual user preferences.

3.1.5 Wearable Device Integration

The assistive system must be compatible with wearable technology such as smart glasses or a small portable camera that can capture live video feed. The system should function with minimal size and weight, ensuring comfort and ease of use for visually impaired users throughout daily activities.

3.1.6 Machine Learning Adaptability

The system should incorporate machine learning algorithms to continuously improve its accuracy in recognizing faces, objects, and currency over time. It must adapt to different environments and conditions by learning from new data, improving performance in real-time as the user interacts with the system.

3.2 SOFTWARE REQUIREMENTS

- Server Side : Python 3.7.4(64-bit) or (32-bit)
- Client Side : TKINTER
- IDE : PYCHARM
- OS : Windows 10 64 –bit

Frond End: Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. In July 2018, Van Rossum stepped down as the leader in the language community. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of Python's other implementations. Python and CPython are managed by the non-profit Python Software Foundation. Rather than having all of its functionality built into its core, Python was designed to be highly extensible.

Back End:My SQL

MySQL is the world's most used open source relational database management system (RDBMS) as of 2008 that run as a server providing multi-user access to a number of databases. The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation.

Inter images

MySQL is primarily an RDBMS and ships with no GUI tools to administer MySQL databases or manage data contained within the databases. Users may use the included command line tools, or use MySQL "front-ends", desktop software and web applications that create and manage MySQL databases, build database structures, back up data, inspect status, and work with data records.

Graphical

The official MySQL Workbench is a free integrated environment developed by MySQL AB, that enables users to graphically administer MySQL databases and visually design database structures. MySQL Workbench replaces the previous package of software, MySQL GUI Tools. Similar to other third-party packages, but still considered the authoritative MySQL frontend, MySQL Workbench lets users manage database design & modeling, SQL development (replacing MySQL Query Browser) and Database administration (replacing MySQL Administrator).MySQL Workbench is available in two editions, the regular free and open source Community Edition which may be downloaded from the MySQL website, and the proprietary Standard Edition which extends and improves the feature set of the Community Edition.

3.3 HARDWARE REQUIREMENTS

- Processor : Intel processor 2.6.0 GHZ
- RAM : 4 GB
- Hard disk : 160 GB
- Compact Disk : 650 Mb
- Keyboard : Standard keyboard
- Monitor : 15 inch color monitor
- Camera

IV. SYSTEM DESIGN AND IMPLEMENTATION

4.1 PROPOSED SOLUTIONS

The proposed system introduces an advanced AI-powered assistant tailored for visually impaired individuals, integrating object detection; face recognition, and currency identification into a single, user-friendly solution. Utilizing deep learning algorithms such as YOLO (You Only Look Once), Grassmann for face recognition, and Convolutional Neural Networks (CNN) for currency detection, the system captures real-time input through an external or wearable camera. This live video feed is analyzed to detect obstacles, recognize known individuals, and identify currency notes providing critical feedback via voice output. The combination of these technologies offers a more holistic and intelligent assistive experience that goes beyond traditional tools. The object detection module, driven by YOLO, allows the system to analyze the surrounding environment and detect obstacles like vehicles, people, or other objects with high accuracy and speed. Based on the object's position and movement, the system guides the user to move left or right, enhancing navigation and safety in real-world scenarios. Simultaneously, the face recognition module leverages the Grassmann algorithm to differentiate between known and unknown individuals, which are especially useful in social or professional settings. When a known face is detected, the system announces the name of the person through audio output, fostering better interaction and awareness. In addition, the currency detection module is built using CNN, trained on datasets of various denominations. It enables blind users to identify currency notes simply by holding them in front of the camera, which the system then recognizes and announces audibly. This feature empowers users to conduct financial transactions independently and with confidence. All modules are seamlessly connected to a voice output interface, ensuring that every piece of information whether it's an obstacle, a person's identity, or a currency value is conveyed clearly to the user in real-time. Overall, this integrated, AI-driven system offers a comprehensive solution that significantly enhances mobility, awareness, and autonomy for visually impaired individuals.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

4.2 SYSTEM ARCHITECTURE

The architecture diagram illustrates the functional flow of the AI-based Blind Assistant System, which combines multiple deep learning algorithms to assist visually impaired individuals. The system begins with the activation of a camera mounted on or used by the blind person, which captures the real-time environment. This input is processed using three main algorithms: YOLO for object detection, Grassmann for face recognition, and CNN for currency detection. Each algorithm operates on its respective dataset to perform precise identification and classification tasks. The YOLO module detects nearby obstacles and objects, helping the system guide the user with directional voice commands, such as moving left or right. Simultaneously, the Grassmann-based face recognition module identifies whether a person in the frame is known or unknown. If the individual is recognized, the system confirms it via voice output.

4.3 MODULES

- DATA MANAGEMENT
    - FACE REGISTRATION
    - YOLO MODEL FOR OBJECT DETECTION
    - CNN MODEL FOR CURRENCY RECOGNITION
- FACE RECOGNITION
- OBJECT DETECTION
- CURRENCY RECOGNITION
- AUDIO FEEDBACK

4.3.1 DATA MANAGEMENT

Data management serves as the backbone of the system, ensuring that the recognition models have accurate and well-organized data to process. This module involves collecting, organizing, and storing images that the system uses for training and validation of machine learning models such as YOLO (You Only Look Once) for object detection and CNN (Convolutional Neural Network) for currency recognition.

• Face Registration

- The system captures multiple images of the user's face from different angles and lighting conditions.
- These images are stored in a database for later use in face recognition.
- Capturing diverse images helps improve the accuracy of face recognition, even in challenging environments.

• YOLO Model for Object Detection

- YOLO is a state-of-the-art object detection model that divides an image into grids and predicts bounding boxes and class probabilities for objects within each grid.
- This model is highly efficient, making it suitable for real-time object detection in dynamic environments.

• CNN Model for Currency Recognition

- A labeled dataset of images of various currency notes is used for training the CNN model.
- The dataset includes images of multiple denominations and variations in conditions such as folded, torn, or partially visible notes.
- The CNN learns to identify the unique features of each currency denomination and classifies them accurately.

4.3.2 FACE RECOGNITION

Face recognition is a crucial module that enables the system to identify familiar individuals. It extracts facial features and compares them with a stored database to recognize the person.

- **Feature Extraction:** The system extracts key facial features such as eyes, nose, and mouth using a pretrained model like OpenCV.
- **Face Matching:** The extracted facial features are compared with those in the database. The Grassmann algorithm is used to compute the similarity and identify the person.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

- **Verification and Identification:** The system can either verify a person's identity (one-to-one matching) or identify them from a group of stored faces (one-to-many matching).

4.3.3 OBJECT DETECTION

Object detection allows the system to identify and track objects in real time, helping users avoid obstacles and navigate safely. This module is particularly useful in dynamic environments where obstacles may appear suddenly.

- **Bounding Box Prediction:** The YOLO model detects objects in the image and draws bounding boxes around them.
- **Object Classification:** Each detected object is classified into predefined categories, such as cars, people, furniture, or animals.
- **Real-Time Adaptation:** The system continuously monitors the environment, updating its detection and classification results in real time.

4.3.4 CURRENCY RECOGNITION

Currency recognition helps visually impaired users identify different denominations of currency notes, allowing them to manage financial transactions independently and accurately.

- **Preprocessing:** The currency image is converted into grayscale or binary format to remove noise and improve recognition accuracy.
- **Feature Extraction:** CNNs are used to extract features such as color, texture, and patterns specific to each currency note.
- **Classification:** The CNN consists of multiple layers (convolutional, pooling, and fully connected layers) that analyze the extracted features and classify the image into the correct denomination.

4.3.5 AUDIO FEEDBACK

Audio feedback is a critical component of the system, providing users with real-time information in a format they can easily understand.

- **Object Recognition Feedback:** When the system detects an object, it announces the object's name and provides relevant instructions (e.g., "Obstacle ahead. Step left.").
- **Face Recognition Feedback:** The system announces the name of the recognized individual (e.g., "John detected.").
- **Currency Recognition Feedback:** Once the system identifies a currency note, it announces the denomination (e.g., "100-dollar note detected.").

V. SYSTEM TESTING AND VALIDATION

5.1 PERFORMANCE METRICS

The performance of the Blind Assistant system will be evaluated based on several key factors to ensure its accuracy, reliability, and real-time response. The primary metrics include accuracy, response time, detection rate, usability, and cost-effectiveness. For face recognition, accuracy will be measured by how well the system identifies familiar faces, with a focus on distinguishing between known and unknown individuals in various lighting and environmental conditions. Similarly, object detection accuracy will assess the system's ability to detect and classify obstacles like vehicles, people, and static objects. Currency recognition will be evaluated based on its ability to correctly identify different currency denominations.

Accuracy

Accuracy measures the system's ability to correctly identify and classify the key elements in the environment, such as faces, obstacles, and currency notes. For face recognition, it refers to the system's ability to correctly distinguish between familiar and unfamiliar faces, even under varying lighting conditions or different angles. For obstacle

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

detection, accuracy ensures that the system detects both static and dynamic objects with minimal false positives or false negatives. In currency recognition, accuracy evaluates the system's ability to correctly identify various currency denominations from different countries or different formats, thus ensuring users can confidently handle their finances.

Precision

Precision in this context refers to the system's ability to identify relevant objects or faces with minimal false positives. For example, when detecting obstacles, a high precision score would mean that the system correctly identifies a true obstacle (e.g., a moving vehicle or a pedestrian) and avoids incorrectly classifying non-obstacles as potential threats. In face recognition, precision ensures that the system only identifies individuals it recognizes as known faces, reducing the chances of false identification.

Recall

Recall measures the system's ability to identify all relevant objects, faces, or currency denominations within its environment, without missing any. A high recall score indicates that the system successfully detects nearly all obstacles in its vicinity, recognizes all familiar faces in crowded environments, and identifies all visible currency notes without missing any. For visually impaired users, recall is particularly important for obstacle detection, as missing even a single moving object could lead to a safety risk.

F1 Score

The F1 score is the harmonic mean of precision and recall and serves as a balanced measure of both metrics. It is particularly useful for evaluating the overall performance of the system, especially in cases where both false positives (e.g., misidentifying an object) and false negatives (e.g., failing to recognize an object) need to be minimized. In the context of the Blind Assistant system, a high F1 score indicates that the system is not only good at identifying relevant objects or faces but is also efficient in avoiding errors that could hinder the user's navigation or task completion.

Response Time

Response time refers to the time it takes for the system to process the video input and provide the appropriate audio feedback to the user. For a blind assistive system, the response time is crucial because real-time, accurate feedback is needed to help the user navigate safely and efficiently. If the system takes too long to process the data, it could lead to delays in reacting to obstacles or missed opportunities for interaction, which could affect the user's confidence and autonomy. Short response times are essential for maintaining a seamless and effective user experience.

Detection Rate

The detection rate measures how often the system correctly detects and responds to the relevant objects, faces, or currency notes within its environment. A higher detection rate indicates that the system is effectively monitoring the surroundings and providing real-time feedback without missing critical elements. For example, in dynamic environments, such as crowded places or busy streets, a high detection rate ensures that the system detects moving obstacles or recognizes familiar faces amidst distractions. This metric is key to ensuring that visually impaired users can trust the system to reliably assist them in daily activities.

5.2 TESTING

Testing is a set activity that can be planned and conducted systematically. Testing begins at the module level and work towards the integration of entire computers-based system. Nothing is complete without testing, as it is vital success of the system.

Testing Objectives:

There are several rules that can serve as testing objectives, they are

- Testing is a process of executing a program to finding an error

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

- A good test case is one that has high probability of finding an undiscovered error.
- A successful test is one that uncovers an undiscovered error.

5.3 TYPES OF TESTING

The development process involves various types of testing. Each test type addresses a specific testing requirement. The most common types of testing involved in the development process are:

- Unit Test
- Functional Test
- Integration Test
- White box Test
- Black box Test
- System Test
- Acceptance Test

Unit Testing:

The first test in the development process is the unit test. The source code is normally divided into modules, which in turn are divided into smaller units called units. These units have specific behavior. The test done on these units of code is called unit test. Unit test depends upon the language on which the project is developed.

Functional Testing:

Functional test can be defined as testing two or more modules together with the intent of finding defects, demonstrating that defects are not present, verifying that the module performs its intended functions as stated in the specification and establishing confidence that a program does what it is supposed to do.

Integration Testing:

In integration testing modules are combined and tested as a group. Modules are typically code modules, individual applications, source and destination applications on a network, etc. Integration Testing follows unit testing and precedes system testing. Testing after the product is code complete.

White Box Testing:

Testing based on an analysis of internal workings and structure of a piece of software. This testing can be done sing the percentage value of load and energy. The tester should know what exactly is done in the internal program. It includes techniques such as Branch Testing and Path Testing. White box testing also called as Structural Testing or Glass Box Testing.

Black Box Testing:

In block box testing without knowledge of the internal workings of the item being tested. Tests are usually functional. This testing can be done by the user who has no knowledge of how the shortest path is found.

System Testing

System testing is defined as testing of a complete and fully integrated software product. This testing falls in black-box testing wherein knowledge of the inner design of the code is not a pre-requisite and is done by the testing team. It is the final test to verify that the product to be delivered meets the specifications mentioned in the requirement document.

Acceptance Testing

This is a type of testing done by users, customers, or other authorised entities to determine application/software needs and business processes. Acceptance testing is the most important phase of testing as this decides whether the client approves the application/software or not. It may involve functionality, usability, performance, and U.I of the application. It is also known as user acceptance testing (UAT), operational acceptance testing (OAT), and end-user testing.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

5.4 SOFTWARE TESTING STRATEGIES

In order to make sure that the programme satisfies its intended criteria and performs as expected, software testing strategies are detailed blueprints that direct the testing procedure throughout a software development project. Various testing approaches are used, depending on the particulars of the project:

Waterfall Testing Strategy: The requirements, design, implementation, testing, and deployment phases of the waterfall model are sequentially approached, with each stage being finished before going on to the next. With a focus on validation against predetermined requirements, this strategy's testing is thorough and well-documented.

V-Model Testing Strategy: The V-Model, which is a development and testing phase relationship model, is an extension of the Waterfall model. To guarantee that requirements are verified early in the process, there is a testing phase that corresponds with each development phase.

VI. CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

This project introduces a cutting-edge solution designed to empower visually impaired individuals by integrating face detection, obstacle detection, and currency recognition into a single wearable device. By combining these essential functionalities, the system provides real-time support, helping users overcome various challenges in their daily lives and enhancing their independence. The system utilizes advanced artificial intelligence (AI) technologies, including the Grassmann model for face recognition, YOLO (You Only Look Once) for object detection, and Convolutional Neural Networks (CNNs) for currency recognition. Each of these models has been meticulously implemented to ensure accuracy and real-time performance. The audio feedback module plays a critical role in delivering immediate and context-specific auditory information, guiding users through complex environments. This feature ensures that users can respond to obstacles, recognize familiar faces, and identify currency notes without needing visual assistance. One of the major strengths of this project is its focus on usability and practicality. Beyond its technical achievements, the project also addresses the social and emotional impact of visual impairment. By reducing the need for constant external assistance, the system promotes greater self-reliance and confidence among users.

REFERENCES

[1] Javed, Sajid, et al. "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA." IEEE Transactions on Image Processing 28.2 (2018): 1007-1022.

[2] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).

[3] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[4] Zhang, Han, et al. "Spda-cnn: Unifying semantic part detection and abstraction for fine-grained recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Ouyang, Wanli, et al. "Deepid-net: Deformable deep convolutional neural networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.