JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Voice-Guided Object Detection: A Comprehensive Survey

Pooja Patil[1], Kishor Mane[2]

[1]M. Tech Student, Department of Computer Science & Engineering, D. Y Patil College of Engineering & Technology, Kolhapur
[2]Assistent Professor, Department of Computer Science & Engineering, D. Y Patil College of Engineering & Technology, Kolhapur

Abstract – Object detection with voice guidance is an emerging domain which integrates the computer vision, human computer interaction and NLP. This can enable to scan the surrounding, interpret the objects and inform to the users through voice command. This system provides assistance to the blind and visually impaired person for navigating on the road or within indoor environment. It also used in the variety of applications including robotics, autonomous vehicle driving, context aware computing, etc. There are few challenges that need to be address including context understanding, real time processing of the surrounding data and object identifications from variety of domains. This paper provides detail review of the various techniques used for the object detection and voice guidance. It reviews various techniques based on the DL, speech to vision modelling, fusion of multimodal data, etc. The various challenges have been identified related to generalization, noise reduction, etc. for design and development of the robust system for voice guided object detection.

Keywords: Voice-guided object detection, Multimodal learning, Speech–vision integration; Human–computer interaction, Assistive technology, Deep learning, Natural language processing, Visual grounding, Audio–visual perception, Transformer models, Context-aware systems, Real-time object detection, Multimodal fusion, Accessibility, Autonomous systems.

1. Introduction:

In the recent, the use of the Ai with computer vision and speech recognition has been raised to makes it possible to analyse the surrounding and understand the scenes almost same as human eyes. Among this voice guided object detection (VGOD) has been evolving faster and enables the visually impaired or blind person to do assistance for navigation with providing the facility of locate, understand and speak about the scene objects. With the integration of the NLP with visual perception, it offers HCI interactions with ease.

Traditionally object detection depends on the manual user Interaction based on the visual inputs. Now with the integration of voice commands it gives detail content related to

the object detected. This will enable the development with the various DL and ML mechanism for the object detection automatically. It is much suitable for industry automation, robot movements, self-driving vehicles, drones and visually impaired persons. Despite these advancements still it faces various challenges including synchronization, real-time environment with speed of recognition of objects, semantic ambiguity in context understanding, lack of availability of datasets for training based on multimodal data and lightening conditions while capturing the scenes.

This paper reviews various ML and DL based object detection and voice guidance techniques. It provides the various details regarding the techniques of object detection for development of intelligent detection of the object with speed. Section II shows review of the ML and DL mechanism used for object detection. Section III presents the challenges in-front of the voice guided detection of objects.

2. Literature Review:

The researchers are worked on the various mechanisms used for the object detection and voice guidance specially used for the blind or visually impaired persons. The following techniques haves been proposed by various researchers as –

R. R. Subramanian, L. Ravikiran [1] proposed a model for guidance of the visually impaired people. This model is based on the ESP32 mechanism which can assist the person through their cap. It detects the objects and provides the audio-based description of that object which can be more helpful for the person who is not able to see. This model is trained with various objects and can able to add any objects for training. This will enhance the navigation of the blind person with ease. Person gets immediate feedback as soon as object is recognized. This can be extended further with various hybrid approaches. S. Bihani and S. Sharma [2] develop the smart glasses using the YOLOv10 mechanism and IoT camera mounted for capturing persons surrounding area. With the help of the MiDaS model the distance between person and object has been recognized and give the voice-based feedback. MiDaS performs encoding and decoding with ResNet. Smart classes tested on the object sequences. Everyday used objects are trained for the recognition. This device also tested in real time environment with the actually blind people. The results and assistance is appreciated by them. S. Kirithika, T. U. Mageswari [3] proposes the solar battery-operated model for the object detection inside the wearable devices. Using the sensors, they give the directions to the blind people for navigation. It shows the advanced sensor technology with green power for navigation. It is ideal for enhancing the blind person life and ease of happenings. L. Narendra, B. S. Babuet al. [4] proposes the smart classes with the AI based voice assistance with the feature of obstacle and object detection. OCR and GPS have been sued for tracking. It will enhance the standard of living of the blind person. The sensors detect the hazards and inform to the system with object recognition. It boosts the confidence in the person. P. A. Garcia Gaona, D. Martin Moncunill, et al. [5] proposes the overview study of the navigational search virtually using knowledge organization system. It uses the 3D structure along with mobile devices for the navigation and assistance.

N. Farheen, G. G. Jaman and M. P. Schoen [6] propose the ML based mechanism to enhance the capability of the noncommercial vehicles using object detection. The obstacles are detected using the raspberry pie and

Google coral. It classifies the images using the pretrained TensorFlow lite model. Degree of obstacles are also measured with classification. It can further be extended for building the navigation platform using the ML. W. Shi, R. Shan and Y. Okada [7] system for recognition of the surrounding environment when user navigates suing the YOLOv5 and RICOH R. The Google map has been embedded with the system for recognition and tracking. This helps to avoid he dangers in the path of the user. X. Hou, M. Zhan, C. Wang and C. Fan [8] focus on the glass object detection using transformer encoder- decoder model. This work used to detect the glass objects using mixed CNN and transformer model. It uses the pytorch environment where model has been trained using the dataset called Trans10K-v2. It shows quite impressive results. Yash Khopkar, Avantika Deshmukh, et a. [9] focuses on the voice guided solution for the assistance using mobile devices. The new solution called voice companion has been designed for visually impaired people using java on android OS. It includes the various features including commands based on voice, object recognition, text to speech capability, location and time with battery status announcement. This solution will improve the life of the blind users. T. Saleem and V. Sivakumar [10] proposes the integrated framework of DL and distance measure mechanism. It identifies the indoor objects and their distance for ease of navigation within the house. The user can wear the device for getting the details of the surrounding and estimate the objects used in the in- door environment along with voice assistant. It estimates the distance of each obstacle and suggest the safe path for navigation using DL mechanism. It is helpful for the visually impaired people. This work can be extended with hybrid DL mechanisms.

Islam, M., Rashid, M., Ahmad, et al. [11] develops the microprocessor-based glass for visually impaired person which can capable to detect the objects. When the obstacle is detected, the buzzer has been raised. This work can be used with the audio output for instructing the users. The proposed model sed the GPS, NEO-6M, camera with the R-CNN approach for the object detection and recognition. This system proves as the real assistance for the blind individuals. It can be further extended with hybrid DL approaches. K. S. Rangam, P. Ram Pukale, et al. [12] propose the smart hat which is a wearable solution with voice assistant. It also includes the object detection, sensing of obstacles, enables the feedback in contextual format for movement of blind people. It uses the edge platform for integration with the various cloud services. The Ngrok has been used for monitoring in real- time. It gives the robust performance with responsive feedback. S. Sreehari, A. R. Menon, et al. [13] proposed the VISTA for finding the objects, hazard detection for the blind users. It uses ML with TTS for independent navigation of the users. It uses the camera-based system for detection of the objects and gives the audio feedback to the user. It uses various speech to text conversion libraries. It uses QR based navigation system for tracking locations in the building. Srikanteswara, R., Reddy, M.C., et al. [14] provides the model for the solving the issues faced by blind and visually impaired people. It uses the image recognition system for detecting the objects from the surrounding areas. It gives the assistant system for the people for the navigation and finding right path. S. S, M. Kumar Thakur, K. Gowda Y, et al. [15] proposes the DL based system for object recognition for the visually impaired people suing the YOLOv5 model. It gives the alert through the speech notification to the user. This system able to identify the location of the objects using CNN. The relative distance and position have been computed using the stereo cameras and sensors.

Golla, Hemanth Kumar Yadav, et al. [16] aims to smart cane with IoT integration for detection of the obstacles. It is able to guide the users for effectively capturing the surroundings and enhance mobility of blind person. It gives the notification about surrounding by recognizing the objects. The proposed system gives better accuracy. Djinko, Issa AR, and Kacem, Thabet [17] propose the DL based object detection technique form he video input using YOLOv7. It captures the ROI and applies on the dataset. When the objects are recognized, it informs that to the user by providing the object description in audio version. The Google voice recognizer APL has been used. M. S. Kabir, S. Karishma Naaz, et al. [18] proposes the blind assistance system along with feedback in voice suing YOLO and SSD mechanisms. It uses the YOLOv4 and YOLOv7 and SS MobileNetv2 on the COCO dataset with more than 8- categories of the objects. It provides the good solution for the blind persons. The gTTS package has been used with Google to test API for voice feedback. K. K,

Y. V. R R, et al. [19] propose the stick for the blind person including AI, object recognition and GPS guided technologies. When the objects are detected, it informs to the person through the headphone and navigate the user suing GPS. It offers the novel solution which is helpful for the blind people for navigating throughout roads and inside home also. Barkovska, O., & Serdechnyi, V. [20] proposes the assistance system for the visually impaired people. It has been integrated with the other systems and having ability for dynamically object detection. It also predicts their movement and provides the feedback to the user. It includes the functional and system analysis modeling. It dynamically detects the obstacles in various environment.

3. Challenges:

Various challenges have been identified through the detail overview of the various techniques used for object detection and voice guidance as follows –

1. The combined approach of the visuals and speech remains difficult due to various data structures used, semantics and time constraints.
2. The context understanding creates the language ambiguity in the scene explanation.
3. The noisy environment degrades the performance of the system.
4. Lack of the annotated and large dataset hampers the synchronization in the speech and visuals.
5. The real time environment facing the latency in processing and dispatching the command.
6. Absence of the benchmarking standards for comparison in the VGOD field.
7. Models are failed in the generalized environment due to lack of object categories.
8. It also faces the challenges related to lightening environment and variation in objects.
9. It faces the ethical values and concerns related to privacy of the users.

4. Conclusion

Object detection from the video stream and voice guidance has been one of the evolving domains for helping the blind persons, guiding autonomous vehicles, navigating the robots, guiding drones. The ML and DL mechanisms has been used for making this process autonomous. This review presents the review of various mechanisms used for voice guided object detection. It performs various steps while detection of the various objects including the pre-processing, feature extraction, training and testing along with model evaluation. This paper highlights the various challenges occurred during the object recognition and speech conversion. It will give the further directions towards development of the intelligent system for object detection and voice guidance in the real time environment.

References:

[1] R. R. Subramanian, L. Ravikiran, K. V. P. Teja, K. V. Reddy and K. N. Reddy, "Voice Guided Object Detection: Enabling Independence for the Visually Impaired," 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Krishnankoil, Virudhunagar district, Tamil Nadu, India, 2024, pp. 1-6, doi: 10.1109/INCOS59338.2024.10527768.

[2] S. Bihani and S. Sharma, "Smart Glasses: Portable Navigation Aid for the Visually Impaired with Object Detection and Monocular Depth Estimation," *2025 7th International Conference on Intelligent Sustainable Systems (ICISS)*, India, 2025, pp. 141-149, doi: 10.1109/ICISS63372.2025.11076511.

[3] S. Kirithika, T. U. Mageswari, A. Rajasekar, D. R, R. N and M. S, "Solar Powered Smart Glasses for Smart Navigation and Object Detection for Visually Impaired," *2025 International Conference on Computing and Communication Technologies (ICCCT)*, Chennai, India, 2025, pp. 1-3, doi: 10.1109/ICCCT63501.2025.11020264.

[4] L. Narendra, B. S. Babu, B. V. Naga Durga Vinay, G. Tirumani and C. Manohar, "Smart Glasses with Voice Assistance and GPS for Independent Mobility of the Blind People," *2025 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)*, Bengaluru, India, 2025, pp. 1-7, doi: 10.1109/ICECCC65144.2025.11064275.

[5] P. A. Garcia Gaona, D. Martin Moncunill, K. Gordillo and R. Gonzalez Crespo, "Navigation and Visualization of Knowledge Organization Systems using Virtual Reality Glasses," in *IEEE Latin America Transactions*, vol. 14, no. 6, pp. 2915-2920, June 2016, doi: 10.1109/TLA.2016.7555275.

[6] N. Farheen, G. G. Jaman and M. P. Schoen, "Object Detection and Navigation Strategy for Obstacle Avoidance Applied to Autonomous Wheel Chair Driving," *2022 Intermountain Engineering, Technology and Computing (IETC)*, Orem, UT, USA, 2022, pp. 1-5, doi: 10.1109/IETC54973.2022.9796979.

[7] W. Shi, R. Shan and Y. Okada, "A Navigation System for Visual Impaired People Based on Object Detection," *2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Kanazawa, Japan, 2022, pp. 354-358, doi: 10.1109/IIAIAAI55812.2022.00078.

[8] X. Hou, M. Zhan, C. Wang and C. Fan, "Glass Objects Detection Based on Transformer Encoder-Decoder," *2022 6th International Conference on Automation, Control and Robots (ICACR)*, Shanghai, China, 2022, pp. 217-223, doi: 10.1109/ICACR55854.2022.9935562.

[9] Yash Khopkar, Avantika Deshmukh, & Prof. Gufran Ansari. (2024). "Voice-guided Mobile Assistance for the Visually Impaired". *International Journal of Information Technology & Computer Engineering*, *4*(02), 6–17. https://doi.org/10.55529/ijitc.42.6.17

[10] T. Saleem and V. Sivakumar, "Reimagining Accessibility: Leveraging Deep Learning in Smartphone Applications to Assist Visually Impaired People Indoor Object Distance Estimation", *EAI Endorsed Trans IoT*, vol. 10, Jul. 2024.

[11] Islam, M., Rashid, M., Ahmad, M., Kuwana, A., & Kobayashi, H. (2022). "Design and implementation of smart guided glass for visually impaired people". International Journal of Electrical and Computer Engineering (IJECE), 12(5), 5543-5552. doi: http://doi.org/10.11591/ijece.v12i5.pp5543-5552

[12] K. S. Rangam, P. Ram Pukale, S. P. Jitendra, D. Lee and J. Gao, "Smart Hat 2.0: An Energy-Aware

Wearable Navigation System for Visually Impaired," *2025 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, Madison, WI, USA, 2025, pp. 1-6, doi: 10.1109/COINS65080.2025.11125722.
[13] S. Sreehari, A. R. Menon, D. Malavika, C. K. Navaneeth, R. M. Devrag and M. Rashmi, "VISTA: Vision-Integrated Smart Tracking Assistant," *2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT)*, Kollam, India, 2025, pp. 1311-1317, doi: 10.1109/ICCPCT65132.2025.11176567.
[14] Srikanteswara, R., Reddy, M.C., Himateja, M., Kumar, K.M. (2022). "Object Detection and Voice Guidance for the Visually Impaired Using a Smart App". In: Shetty D., P., Shetty, S. (eds) Recent Advances in Artificial Intelligence and Data Engineering. Advances in Intelligent Systems and Computing, vol 1386. Springer, Singapore. https://doi.org/10.1007/978-981-16-3342-3_11
[15] S. S, M. Kumar Thakur, K. Gowda Y and N. Jayapandian, "Machine Learning Based Object Detection and Voice Guidance Using Yolo Algorithm," *2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG)*, Indore, India, 2024, pp. 1-6, doi: 10.1109/ICTBIG64922.2024.10911406.
[16] Golla, Hemanth Kumar Yadav, Sree Ranga Bharani Kandalam, Greeshma Chinnarapareddygari, Keerthana Kasireddy, and Balaram KrishnaYenugula. "Smart glasses for object detection and voice guidance for blind people using IoT." In *AIP Conference Proceedings*, vol. 3237, no. 1, p. 060027. AIP Publishing LLC, 2025.
[17] Djinko, Issa AR, and Kacem, Thabet. "Video-based Object Detection Using Voice Recognition and YoloV7". Retrieved from https://par.nsf.gov/biblio/10433657. The Twelfth International Conference on Intelligent Systems and Applications (INTELLI 2023)
[18] M. S. Kabir, S. Karishma Naaz, M. T. Kabir and M. S. Hussain, "Blind Assistance: Object Detection with Voice Feedback," *2023 26th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, Bangladesh, 2023, pp. 1- 5, doi: 10.1109/ICCIT60459.2023.10440977.
[19] K. K, Y. V. R R, M. B. P A A, K. B. C H and M. R, "An Artificial Eye for Blind People," *2022 IEEE Delhi Section Conference (DELCON)*, New Delhi, India, 2022, pp. 1-5, doi: 10.1109/DELCON54057.2022.9752999.
[20] Barkovska, O., & Serdechnyi, V. (2024). Intelligent assistance system for people with visual impairments. *INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES*, (2(28), 6–16. https://doi.org/10.30837/2522-9818.2024.28.006
[21] Lee Sang-Hyeon, Kang Moon-Sik. (2018). Implementation of an Object Detection and Voice Guidance System for the Visually Impaired Using Object Recognition Technology. Journal of the Institute of Electronics Engineers, 55 (11), 65-71. 10.5573/ieie.2018.55.11.65
[22] H. Liu, R. Liu, K. Yang, J. Zhang, K. Peng and R. Stiefelhagen, "HIDA: Towards Holistic Indoor Understanding for the Visually Impaired via Semantic Instance
Segmentation with a Wearable Solid-State LiDAR Sensor," *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 2021, pp. 1780-1790, doi: 10.1109/ICCVW54120.2021.00204.
[23] P. Akkapusit and I. -Y. Ko, "Task-oriented Approach to Guide Visually Impaired People During Smart Device Usage," *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju Island, Korea (South), 2021, pp. 28-35, doi: 10.1109/BigComp51126.2021.00015.
[24] Rohith Raghavan, Vishodhan Krishnan, Hitesh Nishad, Bushra Shaikh, "Virtual AI Assistant for Person with Partial Vision Impairment", ITM Web Conf. 37 01019 (2021), DOI: 10.1051/itmconf/20213701019
[25] Ou, Soobin, Huijin Park, and Jongwoo Lee. 2020. "Implementation of an Obstacle Recognition System for the Blind" *Applied Sciences* 10, no. 1: 282. https://doi.org/10.3390/app10010282
[26] A. Ajina, R. Lochan, M. Saha, R. B. K. Showghi and S. Harini, "Vision Beyond Sight: An AI-Assisted Navigation System in Indoor Environments for the Visually Impaired," *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*, Bengaluru, India, 2024, pp. 1-6, doi: 10.1109/ICETCS61022.2024.10543550.
[27] Gupta, C., Gill, N.S., Gulia, P. et al. A novel finetuned YOLOv6 transfer learning model for real-time object detection. J Real-Time Image Proc 20, 42 (2023). https://doi.org/10.1007/s11554-023-01299-3
[28] Wu, YiHeng, and JianXin Chen. "A Lightweight Real-Time System for Object Detection in Enterprise Information Systems for Frequency-Based Feature Separation."
IJSWIS vol.19, no.1 2023: pp.1-18. https://doi.org/10.4018/IJSWIS. 330015
[29] Zhiran Zhou, Ting Wu, Yixi Ye, Yu Zhang, Yuting He, and Yangguang Shi. 2023. "Dual Guidance Of Optical Flow And Decoupled Attention For Infrared Video Object Detection Network". In Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE '22). Association for Computing Machinery, New York, NY, USA, 191–195. https://doi.org/10.1145/3573428.3573461

[30] M. Christine, H. H. Nuha, M. Irsan, A. G. Putrada and S. B. Izhar Hisham, "Object Tracking in Surveillance System Using Extended Kalman Filter and ACF Detection," 2024 International Conference on Decision Aid Sciences and Applications (DASA), Manama, Bahrain, 2024, pp. 1-6, doi: 10.1109/DASA63652.2024.10836597.