SmartSight AI: A Survey of Context-Aware Obstacle Detection for Visually Impaired Users

Aditya Kokane[1], Shreya Bhujbal[2], Pranjal Gade[3], Om Jagtap[4] Prof. Anjali M. Bhunje

Department of Computer Engineering
SVPM's College of Engineering Malegaon BK, Pune, India
Email: adityakokane60@gmail.com[1], bhujbalshreya31@gmail.com[2], pranjal.gadebmt@gmail.com[3], omjagtap2805@gmail.com[4]

Abstract - Independent mobility remains one of the hardest challenges for people with visual impairment. Traditional aids such as the white cane and guide dog are reliable but limited in range, information richness, and availability. Advances in embedded computing, computer vision, and deep learning have enabled a new generation of electronic travel aids (ETAs) that detect obstacles automatically. However, most systems still answer only the question "is something in front of me?" rather than "what is it, how dangerous is it, and what should I do?" This paper surveys the literature on obstacle detection for visually impaired users, organized around sensing modality, perception algorithm, context awareness, and user feedback. We then introduce SmartSight AI, a conceptual context-aware obstacle detection system, and position it against existing approaches. We conclude with evaluation practices, open challenges, and directions for future research.

Key Words: assistive technology, visually impaired, obstacle detection, context awareness, deep learning, electronic travel aids.

1. INTRODUCTION

Vision impairment affects a very large and growing population. The World Health Organization estimates that at least 2.2 billion people worldwide live with some form of near or distance vision impairment [1], and population-based modeling projects that the number of people who are blind will rise substantially by mid-century as populations grow and age [2]. For these individuals, safe and independent navigation through streets, campuses, transit hubs, and indoor spaces is a daily challenge.

The white cane remains the most widely used mobility tool, but it detects only ground-level and near-field obstacles and gives no information about what an object is. Guide dogs provide excellent guidance, yet they are costly, require long training, and are not accessible to most users. Electronic travel aids (ETAs) aim to fill this gap by sensing the environment and conveying hazards through audio or haptic channels [3], [4].

Early ETAs relied on ultrasonic or infrared range sensors that report the distance to the nearest surface. Modern systems increasingly use cameras and neural networks to recognize objects, estimate depth, and understand scenes [5], [6]. Even so, a fundamental gap persists: detection is not the same as understanding. A parked bicycle at the edge of a sidewalk, a moving car at a crossing, and an open manhole are all "obstacles," but they demand very different responses. Context—where the user is, how they are moving, what the scene semantics imply, and what the user prefers—determines which alerts are useful and which are noise.

This paper makes three contributions:

- A structured review of obstacle detection approaches for visually impaired users, from range sensors to deep-learning and vision–language methods.
- A discussion of context awareness as a distinct layer, covering environmental, situational, and user context.
- A conceptual architecture, SmartSight AI, that positions context-aware risk prioritization as the central design principle, together with evaluation guidelines and open challenges.

The remainder of the paper is organized as follows. Section 2 reviews user needs. Section 3 surveys detection methods. Section 4 examines context awareness. Section 5 discusses feedback design. Section 6 compares representative systems. Section 7 presents SmartSight AI. Sections 8–10 cover evaluation, challenges, and future work, and Section 11 concludes.

2. BACKGROUND AND USER NEEDS

2.1 Mobility Challenges

Navigation for a person with visual impairment involves two intertwined tasks: wayfinding (planning and following a route) and obstacle avoidance (detecting and safely negotiating hazards along that route). This survey focuses on the second task while acknowledging that the two are coupled in practice. Typical hazards include static objects (poles, signs, parked vehicles), dynamic objects (pedestrians, cyclists, vehicles), overhead obstacles (branches, signboards), and terrain hazards (steps, curbs, potholes, drop-offs, wet floors).

2.2 Requirements from Users

Studies and reviews of assistive technology repeatedly highlight a consistent set of user requirements [3], [7]:

- Reliability and safety: false negatives can cause injury, so systems must complement rather than replace the cane or guide dog.
- Low cognitive load: audio channels are also used to hear traffic and conversation, so feedback must be concise and selectively delivered.
- Real-time response: alerts must arrive early enough for a walking user to react.
- Comfort, discretion, and cost: devices should be lightweight, socially unobtrusive, affordable, and have adequate battery life.
- Personalization: users differ in walking speed, residual vision, preferred feedback, and environments.

3. OBSTACLE DETECTION APPROACHES

3.1 Range-Sensor-Based Systems

The earliest ETAs use ultrasonic, infrared, or laser range finders mounted on canes, belts, or glasses. They are inexpensive, low-power, and computationally light, and they provide direct distance measurements [3], [4]. Their limitations are narrow fields of view, sensitivity to surface material and angle, and no ability to classify what they detect. Table I summarizes common sensing modalities.

Table -1: Sensing Modalities for Obstacle Detection

Modality	Strengths	Limitations
Ultrasonic / IR	Cheap, low power, direct range	Narrow beam, no object identity, noisy on soft or angled surfaces
Monocular RGB camera	Rich semantics, low cost, ubiquitous on phones	No metric depth without inference; affected by lighting
Stereo camera	Passive depth, good outdoors	Needs calibration; struggles on textureless surfaces
RGB-D / ToF	Dense depth, good indoors	Limited range and sunlight interference for many sensors
LiDAR	Accurate long-range geometry	Cost, power, and size; limited semantics on its own

3.2 Classical Vision-Based Methods

Before deep learning, camera-based systems used edge detection, optical flow, ground-plane estimation, and feature tracking to identify traversable regions and approaching objects. Simultaneous localization and mapping (SLAM) systems such as ORB-SLAM [8] provide camera pose and sparse maps that can support obstacle awareness. These methods are efficient but tend to be brittle under illumination changes, motion blur, and cluttered scenes.

3.3 Deep-Learning Object Detection

Convolutional and transformer-based detectors now dominate visual obstacle detection. Two-stage detectors such as Faster R-CNN [9] emphasize accuracy, while single-stage detectors such as SSD [10] and the YOLO family [11], [12] favor speed, which matters for wearable and smartphone deployment. Lightweight backbones such as MobileNets [13] enable inference on mobile and embedded hardware. Training data such as MS COCO [14] cover common object categories, but they under-represent assistive-relevant hazards such as open drains, low-hanging branches, uneven pavement, and street furniture in developing-region streetscapes.

3.4 Depth Estimation and RGB-D Perception

Knowing that an object exists is insufficient; the system must know how far away it is. RGB-D wearable systems [6], [15] combine color and depth to detect obstacles and free space. Stereo-plus-CNN pipelines have been implemented on embedded hardware for wearable aids [5]. Self-supervised and zero-shot monocular depth methods such as Monodepth2 [16] and MiDaS [17] make it possible to estimate relative depth from a single camera, which is attractive for smartphone-based aids, though absolute scale and reliability at close range remain concerns.

3.5 Semantic and Instance Segmentation

Segmentation labels every pixel (walkable ground, road, wall, vehicle, person), enabling systems to reason about free space and hazards rather than isolated bounding boxes. Urban-scene datasets such as Cityscapes [18] have driven progress, and promptable foundation models such as the Segment Anything Model [19] offer flexible mask generation. Real-time performance on wearable hardware is the main obstacle to deployment.

3.6 Smartphone and Wearable Platforms

Smartphones are attractive assistive platforms because they are already owned, socially unremarkable, and equipped with cameras, inertial sensors, GPS, speech, and haptics. Commercial applications and devices, such as scene-description apps and head-worn cameras that read text and recognize objects, show demand for camera-based assistance, though they are generally oriented toward recognition and description rather than continuous walking-time obstacle avoidance. Surveys of wearable assistive devices document the trade-offs among form factor, compute, and functionality [7], [20].

4. CONTEXT AWARENESS

We define context-aware obstacle detection as detection whose alerts are conditioned on information beyond the immediate sensor reading. Three categories of context are relevant.

4.1 Environmental Context

Environmental context includes the type of place (indoor corridor, sidewalk, crosswalk, market, transit platform), lighting and weather, and scene semantics such as the presence of roads, stairs, or crowds. The same object warrants different treatment in different environments; a moving vehicle is routine background on a distant road but critical at a crossing.

4.2 Situational and Motion Context

Situational context describes the user's current motion: walking speed, heading, whether they are stationary or turning, and the relative motion of nearby objects. Time-to-contact, rather than raw distance, is a better proxy for urgency. Inertial sensors and visual odometry can estimate ego-motion, while object tracking provides the trajectories of dynamic obstacles.

4.3 User Context

User context includes the person's preferences, residual vision, mobility tool (cane, dog, none), familiar routes, and tolerance for alerts. A system that learns that a user habitually avoids a certain hazard, or that prefers haptic cues in noisy areas, can reduce annoyance and alert fatigue.

4.4 Toward Multimodal and Vision–Language Understanding

Large vision–language models such as CLIP [21] and instruction-tuned multimodal models such as LLaVA [22] can describe scenes and answer questions in natural language, which opens the possibility of conversational, explanatory assistance. Their current drawbacks for safety-critical mobility are latency, hallucination risk, and compute cost. A practical design therefore uses fast, deterministic perception for time-critical hazards and reserves heavier models for optional, on-demand description.

5. FEEDBACK MODALITIES

How hazards are communicated matters as much as how they are detected. Speech is expressive but slow and masks environmental sounds. Non-speech audio cues and spatialized (3-D) audio convey direction and urgency compactly. Bone-conduction headphones leave the ears open to ambient sound.

Haptic feedback through belts, wristbands, or cane handles is private and fast but offers a limited vocabulary. Effective designs combine modalities according to urgency: for example, a haptic pulse and short tone for an imminent collision, and a brief spoken phrase for lower-priority information. Alert rate limiting and hysteresis are needed to avoid overwhelming the user.

6. COMPARATIVE ANALYSIS OF EXISTING SYSTEMS

6.1 Comparison of Representative Approaches

Table II compares representative approaches along the dimensions most relevant to this survey. The comparison is qualitative and based on the published descriptions of each work.

Table -2: Representative Approaches (Qualitative)

Approach	Sensing	Perception	Context use
Ultrasonic / IR canes and belts [3], [4]	Range sensors	Distance thresholding	Minimal; distance only
Wearable stereo + CNN [5]	Stereo camera	Depth plus learned recognition	Limited; focused on detection
RGB-D wearable navigation [6], [15]	RGB-D camera	Geometric obstacle and free-space detection	Limited; geometry-driven
Smartphone recognition apps and head-worn readers	Phone or wearable camera	Object, text, and scene recognition	Description-oriented; not continuous avoidance
SmartSight AI (proposed, conceptual)	RGB + IMU, optional depth	Detection, depth, segmentation	Environment, motion, and user context with risk ranking

Three observations emerge. First, most systems optimize detection accuracy or coverage but not the relevance of what is reported. Second, few explicitly model the user's motion and the time available to react. Third, personalization and alert management are rarely treated as first-class design goals.

6.2 Summary of Related Literature

Table III summarizes fifteen representative studies published between 2021 and 2026, most recent first, listing for each the problem addressed, the solution proposed, and the remaining gap or future scope. Reading across the table, four patterns emerge.

First, detector-centric systems built on YOLO variants and transformer-based detectors [25], [28], [32], [34] report gains in detection accuracy for outdoor obstacles, but they output detections without deciding which ones matter to the user. Second, SLAM and RGB-D systems [24], [35], [37] add localization and handling of dynamic objects, yet they typically rely on simple voice prompts and on sensor accuracy and compute resources; low-cost wearable and stick prototypes [33], [36] pair cameras with ultrasonic sensing on embedded boards, favoring portability and distance cues at the cost of limited compute and sensor range. Third, LLM- and VLM-based systems [23], [26], [27], [29]–[31] contribute scene context and natural-language interaction, but they face hallucination, latency, and dependence on network connectivity. Fourth, evaluation with visually impaired participants remains thin: several studies were validated only by the research team, with a single walking participant, or with very small samples [23], [25], [30], and only one of the surveyed works reports a usability study with a larger group of participants with visual impairment [27]. These gaps motivate the context-aware, on-device-first design proposed in Section 7.

Table -3: Summary of Related Literature

Sr. No	Paper Title	Author(s)	Year	Existing Problem Statement	Existing Problem Solution	Future Scope / Gap
1	Multimodal Navigation System for Visually Impaired Users Using Environmental Perception and Vision-Language Models [23]	H.-Y. Lin, Y.-H. Fan, C.-C. Chang	2026	Indoor aids lack semantic, explainable guidance, and standalone VLMs hallucinate and are weakly grounded.	Combines RTAB-Map localization, YOLO-World open-vocabulary detection, GPT-4o-mini, and OSMAG topological semantic maps for perception-grounded guidance. Precision improves by up to 2.29% on RePOPE; guidance took 3.3 to 7.2 s in two real-world cases.	Multi-sensor fusion and uncertainty estimation, lower latency via on-device or hybrid inference, automated semantic mapping, studies with visually impaired participants (none included), personalized guidance.
2	PISHYAR: A Socially Intelligent Smart Cane for Indoor Social Navigation and Multimodal Human-Robot Interaction for Visually Impaired People [24]	M. H. Joo, M. K. Jafari, A. Taheri	2026	Indoor navigation must account for safe path generation and social interaction, beyond obstacle detection alone.	Smart cane combining RGB-D sensing, YOLOv8 object detection, and D* Lite path planning, with voice-based multimodal human-robot interaction.	Performance depends on sensor accuracy and computational resources.
3	Enhancing Object Detection in Assistive Technology for the Visually Impaired: A DETR-Based Approach [25]	S. Ikram, I. S. Bajwa, S. Gyawali, A. Ikram, N. Alsubaie	2025	Earlier computer-vision navigation aids often fall short of the accuracy needed outdoors, and modern detectors needed a real-time comparison for outdoor obstacle detection.	Cap-mounted mini camera streams to a mobile app (TensorFlow Lite) that compares YOLOv8, Faster R-CNN, and DETR on 80 object classes, with text-to-speech audio alerts. DETR performs best: 98% precision, 99% median confidence, 40 ms per frame. Trained on 2,000 collected images plus augmentation; tested in two 30-minute road walks (day and night) by a visually impaired user.	Larger and more diverse datasets; real-world occlusion, lighting, and environmental dynamics; reliance on GPU-class hardware limits low-resource deployment; per-model hyperparameter tuning.
4	Enhancing the Travel Experience for People with Visual Impairments through Multimodal Interaction: NaviGPT, a Real-Time AI-Driven Mobile Navigation System [26]	H. Zhang, N. J. Falletta, J. Xie, R. Yu, S. Lee, S. M. Billah, J. M. Carroll	2025	Travel support for people with visual impairments needs richer interaction than tactile or simple audio cues.	Mobile system that integrates computer vision, speech interaction, and a large language model to understand surroundings and give conversational navigation guidance.	Depends on accurate visual perception and stable mobile computing, which may limit performance in complex environments.
5	AI-Based Wearable Vision Assistance System for the Visually Impaired: Integrating Real-Time Object Recognition and Contextual Understanding Using Large Vision-Language Models [27]	M. S. A. Baig, S. A. Gillani, S. M. Shah, M. Aljawarneh, A. A. Khan, M. H. Siddiqui	2024	Existing aids lack contextual understanding and personalization, such as recognizing the user's own contacts and objects.	Hat-mounted camera and Raspberry Pi 4 with MobileNet SSD and FaceNet, a user-updatable database, GPT-4o-mini descriptions on button press, and an ultrasonic buzzer below 20 cm. Tested with 50 visually impaired participants: 90% accuracy in good light, 80% in low light, SUS 85; about 1.5 s response and 5.5 s for detailed descriptions.	More capable hardware (e.g., Jetson), depth or thermal sensing for low light, on-device or open-source LLMs for real-time use, less network dependence, smartphone compatibility.

Sr. No	Paper Title	Author(s)	Year	Existing Problem Statement	Existing Problem Solution	Future Scope / Gap
6	YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance [28]	W. Wang, B. Jing, X. Yu, Y. Sun, L. Yang, C. Wang	2024	Outdoor detection suffers from occlusion, dense crowds, and small obstacles such as potholes and traffic cones.	YOLOv8-based detector with a Feature Weighting Block, Adaptive Bottleneck Block, and Enhanced Feature Attention Head. Reaches 30.02 mAP on a public road-obstacle dataset (6,276 images; car, person, traffic cone, pothole).	Improve performance in low light; explore wavelet convolution and StarNet to raise accuracy and inference speed.
7	A Large Vision Language Model Based Environment Perception System for Visually Impaired People [29]	Z. Chen, Z. Liu, K. Wang, K. Wang, S. Lian	2024	Large vision-language models can hallucinate objects when describing scenes for wearable assistive use.	Introduces external segmentation data to reduce LVLM errors and improve the accuracy of scene descriptions in a wearable assistive system.	VLM-based aids still show weak environmental grounding and insufficient fine-grained instructions.
8	Generating Contextually-Relevant Navigation Instructions for Blind and Low Vision People [30]	Z. Merchant, A. Anwar, E. Wang, S. Chattopadhyay, J. Thomason	2024	Existing BLV aids focus on low-level obstacle avoidance and object lists, and often ignore context and relevance to the user's goal.	Builds a dataset of 48 VizWiz images with goals; compares template, LLM (detection, depth, captions, OCR fed to GPT-4), and VLM (GPT-4 Vision) instructions in studies with 8 sighted and 4 BLV participants. VLM instructions rated most useful by BLV users.	Reducing hallucination, training on diverse data, giving more specific spatial directions, and tuning tone and verbosity to user preferences.
9	VisionGPT: LLM-Assisted Real-Time Anomaly Detection for Safe Visual Navigation [31]	H. Wang, J. Qin, A. Bastola, X. Chen, J. Suchanek, Z. Gong, A. Razi	2024	Safe visual navigation requires spotting unusual or hazardous situations, not just listing known object classes.	Pairs YOLO-World open-vocabulary detection with prompt-based LLM guidance to produce concise voice descriptions of unusual situations.	Grounding language output in the real environment and generating fine-grained, actionable instructions remain open.
10	An Effective Obstacle Detection System Using Deep Learning Advantages to Aid Blind and Visually Impaired Navigation [32]	A. B. Atitallah, Y. Said, M. A. B. Atitallah, M. Albekairi, K. Kaaniche, S. Boubaker	2024	Blind and visually impaired (BVI) users need a detector that recognizes and locates common indoor and outdoor obstacles in real time.	A modified YOLOv5 architecture that recognizes and locates a set of indoor and outdoor objects useful for BVI navigation aids.	Detection of small objects and objects in complex backgrounds remains difficult; alert relevance is not addressed.
11	Obstacle Detection and Distance Estimation for Visually Impaired People [33]	X. Leong, R. K. Ramasamy	2023	White canes and guide dogs are limited in range and cost; existing electronic aids are bulky, cannot estimate distance in real time, or require handheld phones.	Wearable smart glasses (3D-printed frame) with a Raspberry Pi, camera module, pretrained TensorFlow Lite CNN detector, and HC-SR04 ultrasonic sensor for distance, with user-selectable audio or vibration feedback. Checked with four pass/fail test cases (a person and a cup, in range and in the danger zone).	Cane or lower-body sensor for below-eye-level obstacles; custom-trained model on diverse data; stereo or time-of-flight depth; custom board with more compute; better ergonomics and integrated audio.
12	Outdoor Mobility Aid for People with Visual Impairment: Obstacle Detection and Responsive Framework for the Scene Perception during the Outdoor Mobility of People with Visual Impairment [34]	W. Khan, A. Hussain, B. M. Khan, K. Crockett	2023	Outdoor mobility requires real-time obstacle detection and scene perception in unstructured environments.	A smart data-driven mobility aid combining responsive devices, IoT-enabled smart-city infrastructure, deep learning, computer vision, and data processing for real-time scene perception.	Reliable detection of small objects and cluttered backgrounds remains a challenge; dependence on smart-city infrastructure limits deployment.
13	Indoor Navigation Assistance for Visually Impaired People via Dynamic SLAM and Panoptic Segmentation with an RGB-D Sensor [35]	W. Ou, J. Zhang, K. Peng, K. Yang, G. Jaworek, K. Müller, R. Stiefelhagen	2022	Indoor navigation aids must cope with moving objects and understand scene structure, which static-scene SLAM handles poorly.	Combines dynamic SLAM with panoptic segmentation on an RGB-D sensor to separate static structures from moving objects and gives real-time voice guidance.	Better adaptation to dynamic environments, stronger semantic understanding, and more extensive real-world evaluation with users.
14	Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification [36]	U. Masud, T. Saeed, H. M. Malaikah, F. U. Islam, G. Abbas	2022	Visually impaired users need a compact, adaptable obstacle-avoidance aid that keeps them safe among people in everyday settings.	Combines a camera with an ultrasonic sensor: the sensor detects nearby objects and sends beep alerts, and the captured data goes to a mobile app that classifies objects with the TensorFlow Object Detection algorithm. Reported as highly efficient and convenient.	Alert relevance and prioritization are not described; ultrasonic range and stick-based operation limit coverage (as summarized in [25]).
15	A Multi-Sensory Blind Guidance System Based on YOLO and ORB-SLAM [37]	C. Rui, Y. Liu, J. Shen, Z. Li, Z. Xie	2021	Blind guidance needs localization and object recognition together in one wearable device.	Combines ORB-SLAM with YOLO-based object detection in a wearable device that uses depth sensing and multimodal feedback (vibration and speech).	SLAM-based aids rely on conventional path planning and simple voice prompts; semantic reasoning and richer interaction remain limited.

7. SMARTSIGHT AI: A CONCEPTUAL FRAMEWORK

SmartSight AI is a conceptual system intended to illustrate how context awareness can be organized as a distinct layer between perception and feedback. It is presented here as a design proposal to structure the survey's findings, not as an evaluated implementation.

7.1 Architecture

Fig. 1 shows the layered pipeline.

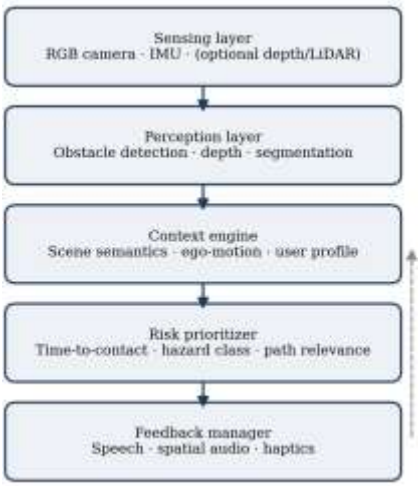


Fig -1: Layered architecture of the conceptual SmartSight AI system.

- Sensing layer: a head- or chest-mounted RGB camera and inertial sensors, with optional depth or LiDAR. A smartphone is the reference platform.
- Perception layer: real-time obstacle detection, monocular or sensor-based depth, and walkable-surface segmentation using lightweight models.
- Context engine: fuses scene semantics, ego-motion, and object tracks with a user profile to estimate what the situation is and how the user is moving.
- Risk prioritizer: ranks hazards by estimated time-to-contact, hazard class, and relevance to the user's forward path, suppressing low-value detections.
- Feedback manager: selects the modality, phrasing, and timing of alerts, with rate limiting and user-tunable verbosity.

7.2 Design Principles

- Relevance over recall alone: report what changes the user's next action.
- Fail-safe complementarity: support, never replace, the cane or guide dog, and degrade gracefully when sensors are unreliable.
- On-device first: run safety-critical perception locally to bound latency and protect privacy; use cloud or large models only for optional description.
- Adaptive feedback: adjust alerts to noise level, urgency, and user preference.
- Explicit uncertainty: communicate low confidence rather than silently guessing.

7.3 Positioning

Relative to the systems in Table II, SmartSight AI's distinguishing idea is not a new detector but a context-conditioned decision layer. It treats the choice of whether, when, and how to alert as a modeled problem informed by motion and user context. Realizing this idea raises research questions in latency budgeting, hazard-severity modeling, and personalization, discussed in Section 9.

8. EVALUATION METHODOLOGIES AND DATASETS

Evaluation of assistive navigation systems is challenging because technical accuracy does not guarantee usability. A rigorous evaluation should combine three levels.

- Algorithmic: precision, recall, and mean average precision for detection; depth error; segmentation IoU; and end-to-end latency and frame rate on target hardware.
- System-level: alert timeliness (lead time before contact), false-alarm rate per minute, missed-hazard rate, battery life, and thermal behavior.
- User-level: studies with visually impaired participants measuring task completion, collisions or near-misses, walking speed, cognitive load (for example NASA-TLX), and subjective trust and comfort.

General-purpose datasets such as MS COCO [14] and Cityscapes [18] are useful for pretraining but are captured from a sighted, often vehicle-centric viewpoint. There is a shortage of large, openly available datasets recorded from a pedestrian, body-worn perspective that include assistive-relevant hazards and diverse geographies. Testing with participants should follow ethical review and safety protocols, beginning in controlled environments.

9. CHALLENGES AND OPEN ISSUES

- Latency versus accuracy on constrained hardware: robust detection and depth on a phone or embedded board while managing heat and battery.
- Robustness: low light, glare, rain, motion blur, reflective and transparent surfaces, and crowded scenes.
- Hard-to-detect hazards: negative obstacles (drop-offs, holes), overhead obstacles, glass doors, and small ground-level objects.
- Dataset bias and domain shift: models trained on Western urban scenes may fail in other geographies and infrastructure conditions.
- Cognitive load and alert design: determining how much information users can absorb while walking.
- Trust and safety: calibration of user reliance, clear communication of limits, and accountability when failures occur.
- Privacy: continuous camera capture raises concerns for users and bystanders, favoring on-device processing and data minimization.
- Affordability and accessibility: solutions must be low-cost and usable in low-resource settings, where a large share of affected people live.

10. FUTURE DIRECTIONS

- Efficient multimodal models that provide reliable scene understanding under tight latency and energy budgets.
- Learned risk models that predict collision likelihood and severity from tracked motion and scene semantics.
- Personalization through on-device and federated learning, adapting to individual routes and preferences without exporting raw video.
- Integration of wayfinding (maps, GNSS, indoor positioning) with obstacle avoidance for unified navigation.
- Richer haptic and spatial-audio interfaces co-designed with users.
- Standardized, pedestrian-view benchmarks and evaluation protocols for assistive navigation, developed with disability communities.

11. CONCLUSION

Obstacle detection for visually impaired users has progressed from simple range sensing to deep-learning-based perception on wearable and mobile platforms. The remaining gap lies less in detecting objects than in deciding which detections matter, when, and how to communicate them. This survey organized the field by sensing, perception, context, and feedback, and proposed SmartSight AI as a conceptual framework that places context-aware risk prioritization at the center of system design. Advancing this vision will require robust on-device models, representative datasets, and, above all, sustained co-design and evaluation with visually impaired users.

REFERENCES

[1] World Health Organization, World Report on Vision. Geneva, Switzerland: WHO, 2019.

[2] R. R. A. Bourne et al., "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," Lancet Glob. Health, vol. 5, no. 9, pp. e888–e897, 2017.

[3] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Trans. Syst., Man, Cybern. C, vol. 40, no. 1, pp. 25–35, 2010.

[4] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," Sensors, vol. 17, no. 3, p. 565, 2017.

[5] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning," in Proc. IEEE Symp. Comput. Commun. (ISCC), 2016, pp. 208–213.

[6] Y. H. Lee and G. Medioni, "RGB-D camera based wearable navigation system for the visually impaired," Comput. Vis. Image Underst., vol. 149, pp. 3–20, 2016.

[7] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," Pattern Recognit. Lett., vol. 137, pp. 37–52, 2020.

[8] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," IEEE Trans. Robot., vol. 31, no. 5, pp. 1147–1163, 2015.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[10] W. Liu et al., "SSD: Single shot multibox detector," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 21–37.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779–788.

[12] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proc. IEEE/CVF CVPR, 2023, pp. 7464–7475.

[13] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.

[14] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in Proc. ECCV, 2014, pp. 740–755.

[15] A. Aladrén, G. López-Nicolás, L. Puig, and J. J. Guerrero, "Navigation assistance for the visually impaired using RGB-D sensor with range expansion," IEEE Syst. J., vol. 10, no. 3, pp. 922–932, 2016.

[16] C. Godard, O. Mac Aodha, M. Firman, and G. J. Brostow, "Digging into self-supervised monocular depth estimation," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 3828–3838.

[17] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 3, pp. 1623–1637, 2022.

[18] M. Cordts et al., "The Cityscapes dataset for semantic urban scene understanding," in Proc. IEEE CVPR, 2016, pp. 3213–3223.

[19] A. Kirillov et al., "Segment anything," in Proc. IEEE/CVF ICCV, 2023, pp. 4015–4026.

[20] M. M. Kuriakose, N. Shrestha, and F. E. Sandnes, "Tools and technologies for blind and visually impaired navigation support: A review," IETE Tech. Rev., vol. 39, no. 1, pp. 3–18, 2022.

[21] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proc. Int. Conf. Mach. Learn. (ICML), 2021, pp. 8748–8763.

[22] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual instruction tuning," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2023.

[23] H.-Y. Lin, Y.-H. Fan, and C.-C. Chang, "Multimodal navigation system for visually impaired users using environmental perception and vision-language models," Sensors, vol. 26, no. 10, Art. no. 3045, 2026.

[24] M. H. Joo, M. K. Jafari, and A. Taheri, "PISHYAR: A socially intelligent smart cane for indoor social navigation and multimodal human-robot interaction for visually impaired people," arXiv:2602.12597, 2026.

[25] S. Ikram, I. S. Bajwa, S. Gyawali, A. Ikram, and N. Alsubaie, "Enhancing object detection in assistive technology for the visually impaired: A DETR-based approach," IEEE Access, vol. 13, pp. 71647–71661, 2025.

[26] H. Zhang et al., "Enhancing the travel experience for people with visual impairments through multimodal interaction: NaviGPT, a real-time AI-driven mobile navigation system," in Proc. ACM Int. Conf. Supporting Group Work (GROUP), 2025, pp. 29–35.

[27] M. S. A. Baig, S. A. Gillani, S. M. Shah, M. Aljawarneh, A. A. Khan, and M. H. Siddiqui, "AI-based wearable vision assistance system for the visually impaired: Integrating real-time object recognition and contextual understanding using large vision-language models," arXiv:2412.20059, 2024.

[28] W. Wang, B. Jing, X. Yu, Y. Sun, L. Yang, and C. Wang, "YOLO-OD: Obstacle detection for visually impaired navigation assistance," Sensors, vol. 24, no. 23, Art. no. 7621, 2024.

[29] Z. Chen, Z. Liu, K. Wang, K. Wang, and S. Lian, "A large vision language model based environment perception system for visually impaired people," in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), 2024, pp. 221–228.

[30] Z. Merchant, A. Anwar, E. Wang, S. Chattopadhyay, and J. Thomason, "Generating contextually-relevant navigation instructions for blind and low vision people," arXiv:2407.08219, 2024.

[31] H. Wang et al., "VisionGPT: LLM-assisted real-time anomaly detection for safe visual navigation," arXiv:2403.12415, 2024.

[32] A. B. Atitallah, Y. Said, M. A. B. Atitallah, M. Albekairi, K. Kaaniche, and S. Boubaker, "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation," Ain Shams Eng. J., vol. 15, no. 2, Art. no. 102387, 2024.

[33] X. Leong and R. K. Ramasamy, "Obstacle detection and distance estimation for visually impaired people," IEEE Access, vol. 11, pp. 136609–136629, 2023.

[34] W. Khan, A. Hussain, B. M. Khan, and K. Crockett, "Outdoor mobility aid for people with visual impairment: Obstacle detection and responsive framework for the scene perception during the outdoor mobility of people with visual impairment," Expert Syst. Appl., vol. 228, Art. no. 120464, 2023.

[35] W. Ou et al., "Indoor navigation assistance for visually impaired people via dynamic SLAM and panoptic segmentation with an RGB-D sensor," in Computers Helping People with Special Needs (ICCHP), 2022, pp. 160–168.

[36] U. Masud, T. Saeed, H. M. Malaikah, F. U. Islam, and G. Abbas, "Smart assistive system for visually impaired people obstruction avoidance through object detection and classification," IEEE Access, vol. 10, pp. 13428–13441, 2022.

[37] C. Rui, Y. Liu, J. Shen, Z. Li, and Z. Xie, "A multi-sensory blind guidance system based on YOLO and ORB-SLAM," in Proc. IEEE Int. Conf. Progress Informat. Comput. (PIC), 2021, pp. 409–414.