Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

WVAB: Temporal Risk-Gated Multimodal Guidance and Trusted Edge Streaming for Offline Assistive Vision

Md Shahanur Islam Shagor

smshagor.ru@gmail.com

Voronezh State University of Forestry and Technologies https://orcid.org/0009-0003-7730-3202

Research Article

Keywords: assistive technology, blind and low-vision users, computer vision, edge AI, human-machine interaction, multimodal guidance, authenticated streaming, temporal stability

Posted Date: September 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11073218/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

WVAB: Temporal Risk-Gated Multimodal Guidance and Trusted Edge Streaming for Offline Assistive Vision

MD SHAHANUR ISLAM SHAGOR

Independent Researcher, Department of Computer Science and Microelectronics Engineering,
Voronezh State University of Forestry and Technologies, Voronezh, Russia

Corresponding author: smshagor.dev@gmail.com

ABSTRACT—Assistive vision systems for blind and low-vision users must manage not only recognition accuracy but also temporal instability, information overload, and the trustworthiness of remotely acquired observations. This article investigates WVAB, an offline-first assistive-vision architecture that combines local object detection, risk-gated multimodal guidance, and authenticated edge-camera transport. The principal research question is whether persistent focus with urgency-based preemption can suppress unstable guidance without masking a newly critical hazard. A controlled temporal-scene study evaluates the guidance policy against instantaneous risk ranking under measurement jitter, an approaching-hazard transition, and a same-band overtake stress case. Across 5,000 near-tie sequences, instantaneous risk ranking produced 47.02 focus switches per 10-s sequence (95% CI: 46.88-47.16), whereas the WVAB persistence policy produced none; announcements fell from 6.00 to 3.00 per sequence. In 5,000 approaching-hazard sequences, instantaneous risk ranking changed focus at 2.76 s on average, while WVAB preempted at 4.73 s when the new target entered the immediate-risk band, yielding a mean responsiveness cost of 1.97 s. A same-band overtake experiment exposed an important limitation: the persistent policy did not relinquish a continuously observed target when the competitor became larger but remained in the same proximity band. Separately, protocol-level AES-GCM fault injection detected all 10,000 tested header, ciphertext, and nonce mutations, while persisted counters rejected all 1,000 replay attempts after reload. These results characterize a measurable stability-responsiveness trade-off rather than claiming universal navigation safety, and they identify a concrete direction for margin-aware preemption in future human-centered evaluation.

INDEX TERMS—*assistive technology, blind and low-vision users, computer vision, edge AI, human-machine interaction, multimodal guidance, authenticated streaming, temporal stability.*

I. INTRODUCTION

Mobile and wearable vision systems can now detect common objects, estimate scene structure, and deliver speech or haptic cues at the edge. However, the engineering problem faced by an assistive interface begins after a detector returns its boxes. If two detections exchange rank from frame to frame, the spoken target can oscillate. If stabilization is too aggressive, a newly dangerous object can be suppressed. If a remote camera is used, duplicated or modified datagrams can be mistaken for current evidence unless freshness and authenticity are enforced. These are temporal and systems problems rather than conventional object-detection problems.

Prior research has established the feasibility of vision-based indoor navigation [1], time-of-flight and haptic local guidance [2], crosswalk-scene understanding [3], edge vision-language assistance [4], and real-time detector-based mobile aids [5]. Wearable and navigation surveys show that practical assistance depends on sensing modality, feedback design, usability, and environmental assumptions rather than detector accuracy alone [6]-[10]. Consequently, this study does not claim novelty for object detection, speech output, or selecting a single relevant object.

The research contribution is a temporal control policy around perception. WVAB retains a focused target while it remains continuously observed, but permits a target in a strictly nearer proximity band to preempt that focus. This mechanism is coupled to a trust boundary for remote camera data, where authenticated encryption and persistent replay state are used before observations are admitted to perception. The central question is not whether the policy is always superior; it is how much instability it removes, what responsiveness cost it introduces, and under which conditions its banded preemption rule becomes too conservative.

The article makes four contributions. First, it formulates the WVAB guidance logic as a reproducible temporal-selection policy rather than a sequence of interface heuristics. Second, it provides controlled comparative experiments that isolate near-tie jitter, cross-band urgency, and same-band overtaking. Third, it evaluates the authenticated transport boundary through protocol-level fault injection and replay persistence. Fourth, it reports both benefits and failure modes, producing a concrete design requirement for future margin-aware preemption and human-subject evaluation.

II. RELATED WORK

A. Assistive Navigation and Multimodal Feedback

Vision-based assistive navigation spans semantic indoor mapping, obstacle avoidance, depth sensing, haptic interfaces, and spatial audio. Li et al. integrated localization, obstacle avoidance, semantic maps, speech, and SmartCane feedback [1]. Katzschmann et al. demonstrated a hands-free time-of-flight and vibrotactile device with blind participants [2]. Tian et al. coupled dynamic crosswalk understanding to audio guidance [3]. These studies, together with the reviews in [6]-[10], establish multimodal navigation as a mature research direction whose unresolved questions increasingly concern interaction timing, uncertainty, and user-centered deployment.

B. Object Detection, Tracking, and Information Selection

Detector-centered systems have become increasingly practical on edge hardware. DEEP-SEE combined object detection, tracking, recognition, and danger-aware acoustic warnings in a wearable navigation framework [11]. Ye and Qian addressed 3-D structural-object recognition for robotic blind navigation [12], and Jafri et al. used visual and infrared/depth-related sensing on Project Tango for indoor obstacle detection [13]. StereoPilot investigated spatial audio for target localization [14]. These works demonstrate that semantic perception, geometry, danger estimation, and multimodal output are established topics.

The novelty boundary is particularly important for focused speech. The Companion smartphone system explicitly observed that narrating all detections is impractical and selected the closest or most relevant object before text-to-speech output [15]. A human-in-the-loop navigation framework by Xiao et al. likewise emphasized contextual interpretation and communication rather than perception in isolation [16]. Therefore, one-object-at-a-time narration is not novel by itself. The present study instead evaluates temporal persistence, categorical urgency preemption, and the resulting stability-responsiveness trade-off.

C. Secure Remote Sensing

UDP offers low-overhead datagram transport but provides neither delivery guarantees nor cryptographic authenticity [17]. IETF guidance places reliability and congestion responsibilities on the application when UDP is used [18]. WVAB therefore treats remote-camera packets as untrusted until authenticated. AES-GCM is used as the authenticated-encryption primitive according to the construction standardized in NIST SP 800-38D [19]. The cryptographic primitive is not proposed as a research novelty; the research question concerns how authenticated metadata and persisted freshness state constrain what data may reach the assistive pipeline.

The literature therefore separates into three relevant layers: perception and navigation, information selection and feedback, and transport/trust. WVAB does not attempt to replace the strongest methods in the first layer; the experiments isolate the second and third layers so their behavior can be measured independently of detector accuracy.

III. PROPOSED SYSTEM AND TEMPORAL GUIDANCE POLICY

A. Runtime Architecture

WVAB is an offline-first assistive-vision architecture with a phone-camera path and optional remote-camera paths. The mobile path performs local object detection over the Microsoft COCO label space [20], converts detections into qualitative proximity and direction, selects one focus, and produces text-to-speech and haptic output. Remote camera data must pass authentication, ordering, and freshness checks before entering the same perception-to-guidance path. Navigation-oriented modules may expose STOP, DEGRADED, or GUIDANCE_AVAILABLE states, but the experiments in this article concentrate on temporal selection and the remote trust boundary rather than path-planning accuracy.

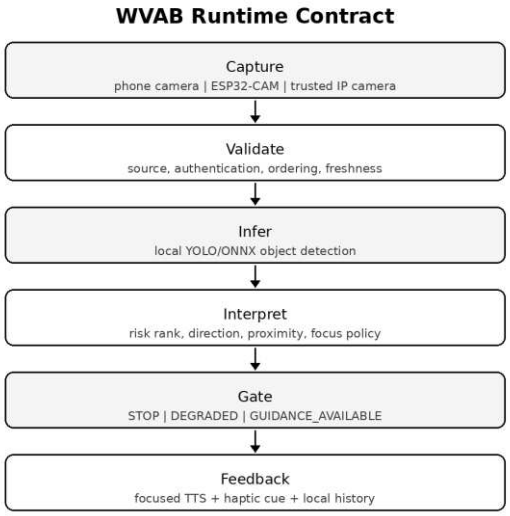


Fig. 1. WVAB capture-to-feedback architecture. Remote observations pass a trust boundary before inference; temporal risk selection operates after detection and before user feedback.

B. Qualitative Proximity and Risk Order

For a normalized detection box B=(x1,y1,x2,y2), let h_B be its normalized height and c_x its horizontal center. The implementation uses four qualitative proximity bands: immediate for h_B>0.60, close for 0.40<h_B<=0.60, medium for 0.20<h_B<=0.40, and far otherwise. Direction is left for c_x<0.38, center for 0.38<=c_x<=0.62, and right for c_x>0.62. These thresholds are interface heuristics, not calibrated collision distances.

Within a frame, detections are ranked lexicographically by proximity band, descending box height, corridor-centrality rank, descending box area, descending confidence, and label tie-break. Let R(d) denote this ordered tuple. The instantaneous target is

$$d^*(t) = \arg\min_d R(d,t) \quad (1)$$

where smaller tuple values indicate higher priority after sign reversal for descending terms. This form is intentionally deterministic and avoids tuning a weighted scalar score whose units mix geometry and confidence.

C. Persistence and Preemption

The temporal state stores the currently focused label and direction. While that label remains visible, the corresponding detection is reacquired and retained. A competing primary target preempts the held focus only when it belongs to a strictly nearer proximity band. Thus, the effective switch condition is

$$switch(t) = 1 \text{ if } band(d^*(t)) < band(d_hold(t)) \quad (2)$$

with equality preventing a switch even when the competitor becomes substantially larger inside the same band. A nominal two-second interval operates as a reacquisition tolerance after the focused label disappears; because the last-seen time is refreshed while the target remains continuously visible, the implementation is better described as persistent focus with a two-second reacquisition window, not a hard two-second maximum hold. This distinction is central to the same-band experiment below.

D. Remote Trust Boundary

Remote video datagrams use a fixed 14-byte network-order header containing session identifier, frame identifier, chunk count, chunk index, and payload length. The header is supplied to AES-GCM as additional authenticated data, so reassembly metadata is cryptographically bound to the ciphertext. Authentication state carries a monotonically increasing counter and next-frame identifier; accepted counter state is persisted so that a receiver restart does not intentionally reopen an old replay window.

IV. EXPERIMENTAL DESIGN

A. Research Questions

RQ1 asks how much persistent focus suppresses oscillation when two same-band detections are nearly tied and measurement noise changes their instantaneous order. RQ2 asks how much response delay is introduced when a new target approaches through successive proximity bands. RQ3 asks whether the categorical preemption rule can miss a meaningful ordering change that occurs entirely within one band. RQ4 asks whether the protocol construction detects metadata/payload tampering and whether persisted counters reject replay after state reload.

B. Temporal-Scene Experiments

All temporal experiments use deterministic policy logic transcribed from the WVAB implementation and a fixed pseudorandom seed (20260915). The simulations are deliberately low-dimensional: they perturb the detector outputs consumed by the policy rather than attempting to model camera optics, pedestrian biomechanics, or user behavior. This isolates the temporal controller from unmeasured perception accuracy.

TABLE II
CONTROLLED TEMPORAL EXPERIMENTS

Experiment	Design	Primary metrics
E1: near-tie jitter	5,000 sequences; 10 s; 10 Hz; two close-band objects; box-height noise sigma=0.025	focus switches; announcements
E2: approaching target	5,000 sequences; 8 s; 10 Hz; held close target plus competitor growing from medium to immediate	first selection time; preemption lag
E3: same-band overtake	5,000 sequences; 8 s; 10 Hz; held close target plus competitor growing within close band only	switch probability; switch time

E1 compares instantaneous risk ranking with the WVAB persistence-plus-preemption policy. Both objects remain within the close band; therefore, any frame-to-frame rank reversal is caused by near-tie jitter rather than a categorical urgency change. The announcement cooldown is four seconds, matching the implementation.

In E2, object A remains close at approximately h_B=0.49. Object B begins in the medium range and grows linearly with a randomized slope until it becomes immediate. Gaussian height noise with sigma=0.012 is added. We record the first frame in which the instantaneous policy selects B and the first frame in which WVAB preempts A. In E3, B grows enough to overtake A in instantaneous risk but is clipped below h_B=0.60, deliberately preventing a band transition.

C. Protocol Fault Injection

A protocol-level executable model reproduces the 14-byte header, 96-bit nonce use, and AES-GCM additional-authenticated-data semantics. Ten thousand random payloads are generated. For each payload, the unmodified packet is decrypted, then one random bit is changed independently in the authenticated header, ciphertext/tag, and nonce. A separate persistent-counter experiment records 1,000 sender sessions, reloads the serialized state, attempts an exact counter replay, and then submits counter+1 as fresh state. These experiments evaluate the stated protocol contract; they are not a radio-network throughput benchmark or a formal proof of AES-GCM.

D. Statistical Reporting

For continuous temporal metrics, means are reported with two-sided 95% confidence intervals based on the Student t distribution across independent sequences. Deterministic zero-event results are reported as observed counts rather than interpreted as universal probabilities. The study does not use participant data and therefore makes no claim about usability, mobility, collision reduction, or clinical benefit.

V. RESULTS

A. Near-Tie Stability

Under same-band near-tie jitter, instantaneous risk ranking changed the focused object 47.02 times per 10-s sequence on average (95% CI: 46.88-47.16). The WVAB persistence policy produced 0 focus switches in all 5,000 sequences because no competitor entered a nearer proximity band. With both targets positioned so that the directional cue remained stable, the four-second cooldown yielded 6.00 announcements per sequence for instantaneous ranking versus 3.00 for WVAB. Thus, under this controlled perturbation, persistence eliminated internal target oscillation and halved announcement count.

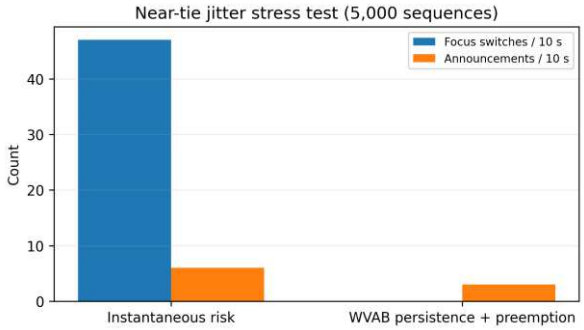


Fig. 2. Near-tie jitter experiment. Persistent focus suppresses frame-to-frame target oscillation and reduces repeated announcements in the controlled same-band scenario.

B. Cross-Band Urgency and Response Cost

In the approaching-hazard experiment, instantaneous risk ranking selected the growing competitor at 2.760 s on average (95% CI: 2.744-2.776 s). WVAB retained the current target until the competitor entered the immediate band, selecting it at 4.728 s (95% CI: 4.706-4.751 s). The mean lag relative to instantaneous ranking was 1.968 s (95% CI: 1.957-1.979 s). Once the observed competitor crossed the immediate-band threshold, preemption occurred in the same simulated frame in all 5,000 sequences.

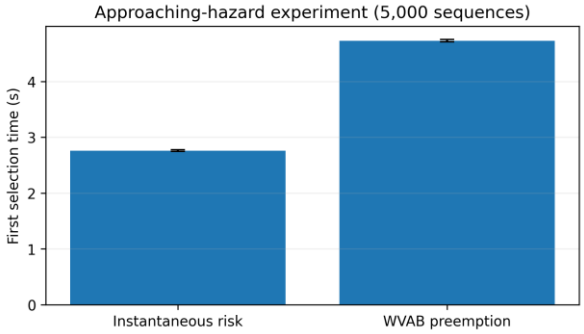


Fig. 3. First selection time in the approaching-hazard experiment. WVAB trades earlier continuous re-ranking for categorical urgency-triggered preemption.

C. Same-Band Overtake: A Limitation Revealed by the Experiment

E3 exposes the cost of categorical preemption. After initialization on object A, the competitor B became larger while remaining inside the close band. Instantaneous risk ranking selected B in every sequence, with mean first-switch time 0.867 s (95% CI: 0.857-0.878 s; median 0.9 s). WVAB did not switch in any of the 5,000 sequences while A remained continuously visible. This is not a detector failure; it follows directly from the policy condition in (2). The result shows that persistence can become over-conservative when meaningful risk changes occur inside one coarse proximity band.

D. Protocol Fault-Injection Results

All 10,000 unmodified AES-GCM packets decrypted successfully. A one-bit mutation in the authenticated header was rejected in 10,000/10,000 cases; a one-bit mutation in ciphertext/tag was rejected in 10,000/10,000; and a one-bit nonce mutation was rejected in 10,000/10,000. In the persisted-counter experiment, all 1,000 accepted counters were rejected when replayed after state reload, while all 1,000 counter+1 updates were accepted as fresh. These results demonstrate consistency with the intended authenticated-data and monotonic-counter contract under the tested fault model; they do not imply protection against key compromise, endpoint compromise, denial of service, or implementation flaws outside the model.

TABLE III
QUANTITATIVE RESULTS

Experiment	Instantaneous / baseline	WVAB / protected path
E1 focus switches / 10 s	47.02 [46.88, 47.16]	0 / 5,000 sequences
E1 announcements / 10 s	6.00	3.00
E2 first selection time	2.760 s [2.744, 2.776]	4.728 s [4.706, 4.751]
E2 lag vs. instant	0	1.968 s [1.957, 1.979]
E3 same-band switch	5,000 / 5,000	0 / 5,000
Header / ciphertext / nonce tamper	not applicable	10,000 / 10,000 rejected for each fault class
Post-reload exact counter replay	not applicable	1,000 / 1,000 rejected

VI. DISCUSSION

A. Stability Is Measurable, but It Is Not Free

The results clarify why temporal guidance should be evaluated separately from detector accuracy. A frame-wise ranking rule can be correct on every frame yet still produce a rapidly oscillating interface when two observations are near-tied. WVAB removes that oscillation by persisting a focus, but the same mechanism deliberately ignores many within-band re-orderings. The policy therefore occupies a stability-responsiveness trade-off rather than dominating instantaneous ranking on all objectives.

B. Preemption Granularity Is the Critical Design Choice

The strongest positive result is categorical urgency handling: a transition into a strictly nearer band immediately interrupts the persistent focus. The strongest negative result is equally clear: a large change inside the same band does not. This suggests that the next version should preserve categorical preemption while adding a margin-aware same-band rule, for example a switch when the competitor exceeds the held target by a validated box-height or time-to-collision proxy margin. Such a rule should be optimized on recorded walking sequences and then evaluated with blind and low-vision participants rather than selected from simulation alone.

C. Relationship to Prior Art

The controlled study does not position WVAB as the first system to prioritize danger or select one spoken object. DEEP-SEE already incorporated danger-aware acoustic warnings [11], and Companion selected a closest or relevant object before TTS [15]. The present contribution is narrower: explicit temporal persistence, urgency preemption, and a quantitative characterization of the conditions under which the policy stabilizes or delays guidance. This framing avoids novelty claims that are contradicted by prior assistive-navigation literature.

D. Security Evidence and Physical Safety

The protocol experiment shows that authenticated metadata, ciphertext, nonce use, and persisted counters behave as expected under the chosen fault model. This evidence should not be conflated with physical safety. A cryptographically authentic frame may still be visually misleading, stale before capture time is known, or processed by an inaccurate detector. Conversely, a conservative STOP/DEGRADED state can reduce functionality without proving absence of hazards. Security, perception, and mobility safety remain distinct layers of evidence.

VII. LIMITATIONS

The temporal evaluation is synthetic and intentionally isolates the guidance controller from detector accuracy. It does not reproduce camera blur, occlusion, lighting, network delay, head motion, speech duration variability, or human walking behavior. Box-height noise is a proxy for detection jitter, not a calibrated empirical distribution from a specific phone. The numerical results therefore characterize policy behavior under stated perturbations rather than real-world clinical or mobility performance.

The present policy tracks persistent focus primarily by class label and direction. Multiple simultaneous instances of the same class can therefore create identity ambiguity, which should be addressed with temporal tracking or instance identifiers. The same-band experiment also

shows that coarse categorical preemption can retain a target too long. These limitations should be resolved before interpreting the system as a mobility aid.

The cryptographic experiments use a protocol-level executable model matching the documented packet format and AES-GCM AAD semantics. They do not replace independent penetration testing, formal verification, secret-provisioning analysis, or RF loss/denial-of-service experiments. No blind or low-vision participants were recruited for this study; human effectiveness, workload, collision avoidance, and usability remain unmeasured.

VIII. DATA AND CODE AVAILABILITY

The WVAB implementation and reproducibility materials are publicly available at https://github.com/smshagor-dev/wireless-Vision-Aid-for-the-blind. The implementation state used to transcribe the evaluated temporal policy is identified by commit 43d8b59f2708e36f8f229f9cf09dd26fdee97021. Repository information is provided only for reproducibility; it is not used as a scientific reference in the bibliography. The controlled simulation parameters, fixed random seed, and aggregate results reported in this article are sufficient to recreate the policy-level experiments independently.

IX. CONCLUSION

This study evaluates WVAB as a temporal and systems problem rather than as another object-detection application. Persistent focus strongly suppresses near-tie oscillation and reduces repeated announcements, while urgency-based preemption responds immediately once a competitor crosses into a strictly nearer proximity band. The same experiments also reveal the mechanism's principal weakness: meaningful within-band risk changes can be masked indefinitely while the held class remains continuously visible. Protocol-level fault injection supports the intended authenticated-data and persisted-replay contract under the tested adversarial conditions. The resulting contribution is therefore a measured design trade-off, not a claim of certified navigation safety. The next research step is to replace categorical-only preemption with a margin-aware rule, evaluate it on recorded device trajectories, and then conduct an ethics-approved study with blind and low-vision users to determine whether the algorithmic stability observed here translates into lower cognitive load and safer real-world guidance.

ACKNOWLEDGMENT AND AI-ASSISTANCE DISCLOSURE

OpenAI ChatGPT was used for language drafting, structural editing, literature cross-check planning, and document formatting. The author remains responsible for the scientific framing, reported results, source-code interpretation, and bibliographic accuracy. The controlled experiments described in this article were executed computationally; no participant observation or hardware measurement was generated by the language model.

REFERENCES

[1] B. Li, J. P. Muñoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Arditi, and M. Yousuf, "Vision-based mobile indoor assistive navigation aid for blind people," IEEE Trans. Mobile Comput., vol. 18, no. 3, pp. 702-714, Mar. 2019, doi: 10.1109/TMC.2018.2842751.

[2] R. K. Katzschmann, B. Araki, and D. Rus, "Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 3, pp. 583-593, Mar. 2018, doi: 10.1109/TNSRE.2018.2800665.

[3] S. Tian, M. Zheng, W. Zou, X. Li, and L. Zhang, "Dynamic crosswalk scene understanding for the visually impaired," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 1478-1486, 2021, doi: 10.1109/TNSRE.2021.3096379.

[4] B. Arystanbekov, H. A. Varol, and A. Yazici, "Edge-based vision-language assistive system for the visually impaired: A quantized VLM approach," IEEE Trans. Neural Syst. Rehabil. Eng., early access, May 25, 2026, doi: 10.1109/TNSRE.2026.3696447.

[5] 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, doi: 10.1109/ACCESS.2025.3558370.

[6] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 1, pp. 25-35, Jan. 2010, doi: 10.1109/TSMCC.2009.2021255.

[7] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," Sensors, vol. 17, no. 3, Art. no. 565, Mar. 2017, doi: 10.3390/s17030565.

[8] M. D. Messaoudi, B.-A. J. Menelas, and H. Mcheick, "Review of navigation assistive tools and technologies for the visually impaired," Sensors, vol. 22, no. 20, Art. no. 7888, Oct. 2022, doi: 10.3390/s22207888.

[9] M. Hersh, "Wearable travel aids for blind and partially sighted people: A review with a focus on design issues," Sensors, vol. 22, no. 14, Art. no. 5454, Jul. 2022, doi: 10.3390/s22145454.

[10] M. M. Valipoor and A. de Antonio, "Recent trends in computer vision-driven scene understanding for VI/blind users: A systematic mapping," Universal Access Inf. Soc., vol. 22, pp. 983-1005, 2023, doi: 10.1007/s10209-022-00868-w.

[11] R. Tapu, B. Mocanu, and T. Zaharia, "DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance," Sensors, vol. 17, no. 11, Art. no. 2473, 2017, doi: 10.3390/s17112473.

[12] C. Ye and X. Qian, "3-D object recognition of a robotic navigation aid for the visually impaired," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 2, pp. 441-450, Feb. 2018, doi: 10.1109/TNSRE.2017.2748419.

[13] R. Jafri, R. L. Campos, S. A. Ali, and H. R. Arabnia, "Visual and infrared sensor data-based obstacle detection for the visually impaired using the Google Project Tango tablet development kit and the Unity engine," IEEE Access, vol. 6, pp. 443-454, 2018, doi: 10.1109/ACCESS.2017.2766579.

[14] X. Hu, A. Song, Z. Wei, and H. Zeng, "StereoPilot: A wearable target location system for blind and visually impaired using spatial audio rendering," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 1621-1630, 2022, doi: 10.1109/TNSRE.2022.3182661.

[15] K. U. Panchal, D. C. Khara, T. J. Gari, and V. Chavan, "Companion: Easy navigation app for visually impaired persons," in Proc. 2021 Int. Conf. Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1-6, doi: 10.1109/CONIT51480.2021.9498496.

[16] J. Xiao, S. L. Joseph, X. Zhang, B. Li, X. Li, and J. Zhang, "An assistive navigation framework for the visually impaired," IEEE Trans. Human-Mach. Syst., vol. 45, no. 5, pp. 635-640, Oct. 2015, doi: 10.1109/THMS.2014.2382570.

[17] J. Postel, "User Datagram Protocol," RFC 768, Aug. 1980, doi: 10.17487/RFC768.

[18] L. Eggert, G. Fairhurst, and G. Shepherd, "UDP usage guidelines," RFC 8085, Mar. 2017, doi: 10.17487/RFC8085.

[19] M. Dworkin, "Recommendation for block cipher modes of operation: Galois/Counter Mode (GCM) and GMAC," NIST Special Publication 800-38D, Nov. 2007, doi: 10.6028/NIST.SP.800-38D.

[20] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in Computer Vision - ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740-755, doi: 10.1007/978-3-319-10602-1_48.

AUTHOR BIOGRAPHY

MD SHAHANUR ISLAM SHAGOR is an independent researcher and software/AI engineer working on edge artificial intelligence, assistive computer vision, autonomous systems, trustworthy distributed systems, and security for cyber-physical platforms. His research interests include offline inference, multimodal human-machine interfaces, reproducible systems evaluation, and safety-bounded intelligent systems.