

Article

SnapStick: Merging AI and Accessibility to Enhance Navigation for Blind Users

Shehzaib Shafique ^{1,*}, Gian Luca Bailo ², Silvia Zanchi ¹, Mattia Barbieri ¹, Walter Setti ¹, Giulio Sciortino ², Carlos Beltran ², Alice De Luca ¹, Alessio Del Bue ²  and Monica Gori ¹

¹ Unit for Visually Impaired People (U-VIP), Italian Institute of Technology, 16152 Genova, Italy; silvia.zanchi@iit.it (S.Z.); mattia.barbieri@iit.it (M.B.); walter.setti@iit.it (W.S.); alice.deluca@iit.it (A.D.L.); monica.gori@iit.it (M.G.)

² Pattern Analysis and Computer Vision (PAVIS), Italian Institute of Technology, 16152 Genova, Italy; gian.bailo@iit.it (G.L.B.); giulio.sciortino@iit.it (G.S.); carlos.beltran@iit.it (C.B.); alessio.delbue@iit.it (A.D.B.)

* Correspondence: shehzaib.shafique@iit.it

Abstract

Navigational aids play a vital role in enhancing the mobility and independence of blind and visually impaired (VI) individuals. However, existing solutions often present challenges related to discomfort, complexity, and limited ability to provide detailed environmental awareness. To address these limitations, we introduce SnapStick, an innovative assistive technology designed to improve spatial perception and navigation. SnapStick integrates a Bluetooth-enabled smart cane, bone-conduction headphones, and a smartphone application powered by the Florence-2 Vision Language Model (VLM) to deliver real-time object recognition, text reading, bus route detection, and detailed scene descriptions. To assess the system's effectiveness and user experience, eleven blind participants evaluated SnapStick, and usability was measured using the System Usability Scale (SUS). In addition to the 94% accuracy, the device received an SUS score of 84.7%, indicating high user satisfaction, ease of use, and comfort. Participants reported that SnapStick significantly improved their ability to navigate, recognize objects, identify text, and detect landmarks with greater confidence. The system's ability to provide accurate and accessible auditory feedback proved essential for real-world applications, making it a practical and user-friendly solution. These findings highlight SnapStick's potential to serve as an effective assistive device for blind individuals, enhancing autonomy, safety, and navigation capabilities in daily life. Future work will explore further refinements to optimize user experience and adaptability across different environments.

Keywords: navigational aid; VLM; assistive technology; environment understanding; audio feedback; blind navigation



Academic Editor: Jeffrey W. Jutai

Received: 16 June 2025

Revised: 5 July 2025

Accepted: 9 July 2025

Published: 11 July 2025

Citation: Shafique, S.; Bailo, G.L.; Zanchi, S.; Barbieri, M.; Setti, W.; Sciortino, G.; Beltran, C.; De Luca, A.; Del Bue, A.; Gori, M. SnapStick: Merging AI and Accessibility to Enhance Navigation for Blind Users. *Technologies* **2025**, *13*, 297. <https://doi.org/10.3390/technologies13070297>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Blind and visually impaired (VI) people usually rely on aids like white canes, guide dogs, and volunteer or certified guide assistance to help them navigate [1–4]. Improved auditory and echolocation abilities are frequently developed as a result of early-onset vision loss, allowing for navigation utilizing landmarks and background noise [5]. Tactile pavements are essential for improving safety and spatial awareness in urban areas, especially in the vicinity of public transportation hubs [6]. The majority of VI people who are independent travelers have had specific training to teach them these skills, called orientation and mobility training [7,8].

1.1. Technology and Assistive Aids

Technology has drastically changed assistive aids, making them more accessible to people with disabilities. According to Bhowmick [9], assistive technologies are any equipment, services, or modifications that help people with impairments overcome obstacles to accessibility and lead independent, active lives. Wayfindr and Envision are examples of technologies that have become essential for navigation [10]. The potential of mobile devices with sophisticated sensors and processing power as platforms for mobile-based navigation aids—which are quickly becoming the norm in assistive technology—was highlighted by Csapo [11].

The human, technological, and physiological aspects of mobility, direction, and information availability have all been extensively studied in the context of assisted navigation for the blind and visually impaired. Dynamic interaction and adaptation for indoor and outdoor locations are currently features of many navigation systems. Many solutions, however, continue to be unduly complicated and fall short of consumers' expectations for usability and simplicity despite their capability [9,12,13]. Due in large part to discomfort, intrusive designs (such as those that block auditory input), the cognitive load they impose, the need for intensive training, and other factors, no one navigational aid has been widely adopted by the visually impaired community [14,15].

1.2. Electronic Travel Aids

Electronic Travel Aids (ETAs) take advantage of standard cameras (such as smartphone cameras), depth cameras, Radio Frequency Identification tags, Bluetooth beacons, ultrasonic sensors, and infrared sensors. Despite being small and producing high-quality photos, smartphone cameras alone are not very good at identifying objects that are farther away due to their inability to record depth information. On the other hand, Time-of-Flight technology is used by depth cameras, such as the Microsoft Kinect, to provide range information and precise obstacle placement; however, its performance is limited under bright sunlight. Likewise, although Light Detection and Ranging (LiDAR) devices provide accurate range analysis, their bulkiness makes them less appropriate for mobile navigation. To enhance visual navigation, smartphones have recently integrated depth sensors, combining portability, processing power, and compactness [16].

1.3. AI-Powered Systems

The effectiveness of commercial AI-powered systems, like Seeing AI (Microsoft), has been assessed through pilot studies. Kupferstein [17] presented initial results from diary research where participants completed a variety of tasks using Seeing AI. Even while AI applications might offer useful information that people would be reluctant to ask for from others, more study is required to determine the circumstances in which these tools can best serve users' requirements [17,18].

Although apps using AI have significant advantages, there are still a number of drawbacks. Existing research on the long-term usefulness of these devices is limited since it frequently relies on anecdotal evidence or brief observation periods. Furthermore, it can be difficult to hear auditory feedback in busy environments, and traditional earphones that cover the pinna make it difficult to hear background noise, which is necessary for safe navigation, while smartphone loudspeakers can be too weak to provide helpful feedback.

To overcome these obstacles, a comprehensive strategy that takes into account the preferences, struggles, and experiences of visually impaired users is required. Such a strategy must incorporate cutting-edge features catered to their particular navigational and informational requirements in dynamic, real-world settings. By creating an AI-powered

scene description application and assessing user satisfaction, interactions, and potential areas for development, this study seeks to close these gaps.

2. Methodology

In order to investigate the real-world uses of scene description technologies, we surveyed eleven visually impaired (VI) people in depth. Participants used bone-conduction headphones, a smartphone app, and a Bluetooth-enabled cane to interact with our unique navigation tool, called SnapStick. This technology gave users the ability to take pictures of their environment and instantly get thorough scene descriptions.

Participants used SnapStick to explore one indoor experimental room and some pictures outside the windows. They gave thorough feedback on their experiences after their investigations, emphasizing the system's usability, the scene descriptions' correctness and applicability, and any difficulties they ran into when utilizing the tool. We have gained a deeper understanding of the real world uses of AI-powered navigation aids thanks to this input, which has also helped us identify important areas for future development.

2.1. Participants

Eleven blind subjects (mean age: 43.02 ± 11.96 years, six females) were included in the study; they were all healthy. Participants were recruited from a large internal database of blind individuals maintained by the Italian Institute of Technology. Due to the rarity of such databases and institutional restrictions on external data access, recruitment from outside sources was not feasible. Participants were randomly selected based on availability, with specific inclusion criteria of visual impairment and no other motor or sensory disability. This approach helped minimize bias while ensuring ethical compliance and logistical feasibility.

Each subject gave written informed consent before the trial. The ethics committee of the local health authority (Comitato Etico, ASL 3, Genoa, Italy) approved the experimental procedures, which were carried out in compliance with the Declaration of Helsinki's ethical guidelines [19]. With an average session length of roughly 30 min, the length of each session was customized to the participants' particular abilities.

2.2. SnapStick

Several crucial elements are integrated into the SnapStick navigation aid system created for this study in order to properly support people with visual impairments. The main component is an easy-to-use smartphone application, which is enhanced with a Bluetooth-enabled cane for smooth communication. Furthermore, to ensure improved environmental awareness, audio descriptions are delivered using bone-conduction headphones without interfering with the user's ability to hear background noise.

The mobile application is powered by Florence-2, a Vision Language Model (VLM) at the core of the system that offers real-time descriptive support via an easy-to-use interface. A client-side application and a server service are the two main parts of the system, which cooperate to make it easier to take pictures, process them, and provide scene descriptions.

The Florence-2 model Vision-Language Model is locally hosted by the server service, allowing for quick picture processing without the lag that comes with distant servers. The Florence-2 is not executed directly on the smartphone. Instead, it runs on a separate private local server (e.g., a desktop computer on the same Wi-Fi network). The server can effectively accept images from the client application, process them using the VLM model, and produce thorough descriptions thanks to this local deployment, which reduces latency and network problems. The model has been improved to recognize objects, people,

and facial expressions; it can even read text from photographs and generate context-specific descriptions that are suited to the needs of visually impaired users.

The main interface for system interaction is the client application, which is installed on users' smartphones. It makes use of an accessible version of the open-source camera program OpenCamera. With a user interface (UI) that has big, easily accessible buttons, users may use the phone's camera to take pictures of their surroundings. The app allows users to take pictures hands-free by pairing with a cane's Bluetooth button for increased convenience.

After capturing an image, the server analyzes it and produces a description that is returned to the client application. Through bone-conduction headphones, the description is then read aloud so that people can hear it without losing awareness of their surroundings. The system's real-time feedback feature is essential for helping users navigate a variety of environments. The accuracy of picture descriptions was qualitatively assessed prior to users' experience and was deemed to be highly accurate.

In addition to scene descriptions, SnapStick supports text reading and bus route detection. These features are handled by the same Florence-2 VLM used for object recognition. For text reading, the model uses its OCR capabilities to extract and vocalize printed or digital text from signage, labels, or posters present in the captured image. Bus route detection is performed by analyzing the region of interest in front-facing images of buses, where the model identifies alphanumeric characters (e.g., bus numbers or destinations) and integrates contextual cues (e.g., "bus stop" and "bus front") to increase precision. Although real-time validation of the bus detection functionality was not conducted due to safety concerns for blind participants in active traffic environments, its performance was evaluated using a dataset of bus images. The feature is fully implemented and functions effectively within the same client-server pipeline.

The validation of auditory descriptions was conducted qualitatively. We compared generic descriptions provided by the user with those generated by the app for a sample of images divided into three categories: indoor context, people, and outdoor context. The evaluation was based on criteria of accuracy, completeness, clarity, relevance, consistency, detail, image depth, and colors. For images of people, the app can recognize facial expressions, such as smiles, sadness, seriousness, etc. The results showed that the app's descriptions are generally accurate and detailed, with a good representation of depth and colors, although in some cases, they could benefit from simplification to improve clarity and relevance. Figure 1 illustrates an office environment featuring various objects both on the table and within the room. The app successfully and accurately identified and described every item present in the scene. For example, it recognized a wrapper on the table, described the details of a poster on the wall, and even highlighted notes written on a whiteboard.

The client and server data transmission mechanism is dependable and effective. In order to ensure cross-platform compatibility, images are encoded into JSON format and sent via a RESTful API. Following image processing, the server provides descriptive text that includes important visual details that are pertinent to users who are blind or visually impaired. Users are then given timely and useful information by having this material read back to them via audio feedback.

In order to guarantee reliable performance, the system also integrates a number of optimization techniques, such as accurate message exchange protocols to reduce data loss, real-time interactions to offer immediate feedback, and effective network latency control for quick image processing and transmission. These adjustments are necessary to guarantee that the SnapStick system provides dependable and efficient support in a variety of settings, making it an invaluable navigation tool for those with visual impairments. Figure 2 illustrates the usage scenario of the Snapstick system through a sketch.



Figure 1. An indoor environment depicting an office, subject to an application validation test. The following text is the precise output generated by the application: *The image depicts an office workspace with a computer monitor displaying code and a chat application. To the left, a black tower computer with blue LED lighting stands on a wooden desk. In front of the computer, there is a black and yellow mechanical keyboard. On the desk, there's also a water bottle and a snack wrapper. In the background, there is a whiteboard with mathematical equations and notes, and a poster on the wall with information about machine learning and generative models.*



Figure 2. A sketch representation of the SnapStick system. The blind participant takes a picture with their smartphone by pressing the Bluetooth button attached to a commercial walking cane (see the left squared panel). The app elaborates the picture and provides an auditory description of it. The participant can safely listen to the description thanks to the bone-conducting headphones (see the right squared panel).

2.3. Data Collection

Eleven blind individuals in our study explored the experimental room using the SnapStick navigation device. Each participant received the system, which included bone-conduction headphones, a Bluetooth-enabled cane, and a smartphone application, and navigated their environment, which potentially included recognizing outdoor scenery through a window.

In order to evaluate their experience with the SnapStick technology, participants filled out two surveys in Italian after the excursion. The System Usability Scale (SUS), a well-known standardized instrument for evaluating the usability of systems, goods, or services, was used for the initial survey. We also used a specially created survey meant to assess the navigational aid's overall usefulness, comfort, and level of satisfaction. These surveys helped pinpoint areas in need of more development and enhancement by offering insightful information about the system's functionality and user experience.

The custom-made questionnaire included several primary questions, which were administered in the participants' native language (Italian). Below, the corresponding English translations are provided:

1. How satisfied are you with the overall performance of the navigational aid (mobile app, cane, and bone-conduction headphones)? Answer choices included: Very satisfied, Satisfied, Neutral, Dissatisfied, and Very dissatisfied.
2. How comfortable did you feel using the navigational aid? Answer choices included Very Comfortable, Comfortable, Neutral, Uncomfortable, and Very Uncomfortable.
3. How confident did you feel navigating the environment using this aid? Answer choices included Very Confident, Confident, Neutral, Unconfident, and Very Unconfident.
4. How clear was the feedback provided through the bone-conduction headphones? Answer choices included Very Clear, Clear, Neither Clear nor Unclear, Unclear, and Very Unclear.

5. Did you encounter any problems or flaws while using the navigational aid? Answer choices included No or Yes.
If yes, please describe the issue.

2.4. Data Privacy and Security

SnapStick is designed with a privacy-first approach, particularly in compliance with the General Data Protection Regulation (GDPR). All image processing is performed locally on a private server, and no data is uploaded to third-party or cloud services. To further protect user data, the following measures are in place:

- All images captured through the app are temporarily stored in volatile memory and are automatically deleted immediately after processing.
- No images are saved to the device's storage or transmitted externally.
- Communication between the client app and local server uses secure protocols (e.g., HTTPS or local loopback with TLS).
- No personally identifiable information (PII) is recorded, and no logging of visual content takes place.

These measures ensure that user data remains confidential, is processed only transiently, and never leaves the user's control. Future updates will include options for end-to-end encryption, user-adjustable data-retention settings, and enhanced consent management to reinforce data security and trust.

3. Results

Eleven blind individuals evaluated the SnapStick navigational device across a range of real-world environments. User feedback was gathered through the System Usability Scale (SUS) and a custom questionnaire assessing usability, comfort, confidence, and audio clarity.

SnapStick achieved a mean SUS score of 84.7 (see Table 1), which corresponds to an "A+" usability rating and places the system in the 96th–100th percentile, based on standard interpretation guidelines established by Sauro and Lewis [20]. This score indicates a high level of user satisfaction and system usability. Notably, one participant (Subject S9) reported a lower SUS score of 35. Although this may seem inconsistent with some of their individual answers (e.g., Q1 = 4, Q5 = 3), the score was accurately calculated according to the standard SUS methodology and reflects the participant's original responses. Due to this outlier, the overall mean SUS score did not come close to the maximum score of 100, despite very high ratings from the majority of participants.

Responses to the custom questionnaire (see Table 2) further supported SUS findings: 54.5% of participants reported being "Very satisfied", and 36.4% were "Satisfied," yielding an overall satisfaction rate of 90.9%.

In terms of comfort, 90.9% of users rated the device as "Very comfortable" or "Comfortable," with only one neutral rating. Confidence in using SnapStick was also high, with 81.8% of participants reporting feeling "Confident" or "Extremely confident." While 18.2% of participants neither felt confident nor unconfident, no significant difficulties or hesitations were reported.

Crucially, all participants rated the auditory feedback provided via bone-conduction headphones as either "Clear" or "Very Clear," underscoring the effectiveness of the audio output. Moreover, no technical faults or usability issues were observed during the trial, reflecting a seamless and reliable user experience.

Table 1. Individual SUS score for each statement (Q1–Q10) and total score. Note: SUS statements—Q1: I think that I would like to use this system frequently; Q2: I found the system unnecessarily complex; Q3: I thought the system was easy to use; Q4: I think that I would need support to use this system; Q5: I found the various functions well integrated; Q6: I thought there was too much inconsistency; Q7: I would imagine that most people would learn to use this system very quickly; Q8: I found the system cumbersome to use; Q9: I felt very confident using the system; Q10: I needed to learn a lot before getting started.

Subject	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total Score
S1	4	1	5	3	5	1	5	1	5	1	92.5
S2	4	1	5	4	5	3	5	1	5	1	85
S3	5	3	3	2	4	1	4	2	4	1	77.5
S4	1	3	5	1	4	1	4	1	4	1	77.5
S5	5	1	5	1	5	2	5	1	5	1	97.5
S6	5	1	5	1	4	1	5	1	5	1	97.5
S7	5	1	5	3	5	1	5	1	5	1	95
S8	5	1	4	4	4	1	4	2	3	1	77.5
S9	4	4	2	5	3	3	4	4	2	5	35
S10	5	1	5	2	5	1	5	1	5	1	97.5
S11	5	1	5	1	5	1	5	1	5	1	100

Table 2. Individual answers for each question of the customized questionnaire (Q1–Q5).

Subject	Q1	Q2	Q3	Q4	Q5
S1	Very Satisfied	Very Comfortable	Very Confident	Very Clear	No
S2	Very Satisfied	Comfortable	Confident	Very Clear	No
S3	Satisfied	Comfortable	Unconfident	Very Clear	No
S4	Neutral	Comfortable	Confident	Very Clear	No
S5	Very Satisfied	Very Comfortable	Very Confident	Very Clear	No
S6	Very Satisfied	Comfortable	Very Confident	Very Clear	No
S7	Very Satisfied	Comfortable	Very Confident	Very Clear	No
S8	Satisfied	Comfortable	Confident	Clear	No
S9	Satisfied	Neutral	Unconfident	Clear	No
S10	Satisfied	Very Comfortable	Very Confident	Clear	No
S11	Very Satisfied	Very Comfortable	Very Confident	Clear	No

System Performance Evaluation

To complement the subjective usability evaluation, we conducted a set of quantitative benchmarks on system performance using a mid-range Android smartphone, specifically the Xiaomi Redmi Note 8 Pro (CPU: Octa-core (2×2.05 GHz Cortex-A76 & 6×2.0 GHz Cortex-A55), 4 GB RAM). The following metrics were recorded:

- **Recognition Accuracy:** We manually labeled and compared app-generated descriptions for 40 test images (including indoor scenes, images with texts, outdoor scenes, and images containing people). Descriptions were evaluated for correctness by 2 in-

dependent human raters using predefined criteria. SnapStick achieved a recognition accuracy of 94%.

- **Inference Latency:** The average end-to-end time (from image capture to audio feedback delivery) was 1.7 ± 0.2 s per image in our local server setup. Latency remained below 2 s in 96% of cases.
- **Power Consumption:** For a 30-min navigation session, the app consumed approximately 9.8% of battery on a fully charged 4500 mAh phone, corresponding to an estimated 198 mW power draw. Most of this came from camera use and VLM inference.

4. Discussion

The development of assistive technologies for blind individuals has gained increasing attention in recent years, particularly in the domain of spatial navigation. Despite the proliferation of navigational aids (see [21] for a comprehensive review), widespread adoption among blind users remains limited. This is largely due to persistent challenges such as discomfort, inaccuracy, intrusive designs, and impractical user interfaces. For instance, many existing devices block the ears to deliver auditory feedback, which poses a safety concern and is generally disliked by blind users [14,15]. In a previous study by Shafique [22], blind participants highlighted the limitations of current aids, including the bulkiness and oversensitivity of vibrating canes, vague directional feedback, and cumbersome GPS-based applications that suffer from signal loss, high battery consumption, or dependence on constant internet connectivity. Additionally, many sound-based applications overwhelm users with complex audio cues, and mobile apps often struggle with object recognition. Cost has also been a significant barrier to adoption.

In response to these challenges, we developed SnapStick, an integrated navigational aid composed of three primary components: (1) a smartphone application powered by a Vision-Language Model (VLM), (2) bone-conduction headphones, and (3) a cane equipped with a Bluetooth shutter button.

Understanding the legal requirement for canes in many countries (e.g., Italy), the system is designed to complement, rather than replace, the cane. The Bluetooth button allows users to take photos without needing to interact with the smartphone directly, enabling hands-free and safe operation while complying with legal norms. The use of bone-conduction headphones—delivering sound without covering the ears—ensures users remain aware of environmental sounds, addressing a common concern about traditional headphones [21].

At the core of SnapStick is the Florence-2 Vision Language Model, implemented in the mobile app to provide real-time, detailed scene descriptions, text reading, human identification, and recognition of bus numbers and routes. Importantly, the application runs on a private local server, ensuring user data remains secure and does not rely on external cloud services. This design supports autonomy and privacy while maintaining low latency for fast, accurate feedback that enhances both safety and navigation efficiency.

SnapStick directly addresses nearly all concerns voiced by blind users in earlier studies. It avoids the need for continuous handling, eliminates bulky or intrusive feedback systems, and is user-friendly. By allowing users to keep their phone in a pocket and enabling selective access to feedback (e.g., swiping the screen to stop descriptions), the system avoids overwhelming the user with unnecessary information.

In addition to subjective feedback, system performance was quantitatively assessed on a mid-range Android device (Xiaomi Redmi Note 8 Pro). SnapStick achieved a recognition accuracy of 94% across diverse image categories, validated by two independent raters. The system demonstrated efficient real-time processing, with a mean inference latency of 1.7 ± 0.2 s, remaining below 2 s in 96% of cases. Power consumption during a 30-min usage

session was approximately 9.8% of battery on a 4500 mAh phone, confirming the system's practicality for everyday use without significant energy cost. These results highlight SnapStick's technical robustness and suitability for mobile, real-time applications.

The usability and user satisfaction of SnapStick were assessed using the System Usability Scale (SUS), yielding a high average score of 84.7. This places the device within the 96th–100th percentile of Sauro's industry usability benchmarks, earning an A+ rating. Given that the average SUS score across technologies is 68, and scores above 80 are typically associated with excellent user satisfaction, these results suggest that users found SnapStick to be intuitive, efficient, and highly usable for key tasks such as reading text, interpreting scenes, and navigating independently.

User feedback during validation further reinforced these findings. A substantial 90.9% of participants (10 out of 11) reported being either "Very satisfied" or "Satisfied" with the system. Comfort was positively rated by 90.9% of users, while 81.8% reported feeling confident while using the device. Impressively, 100% of participants rated the clarity of the audio feedback—delivered through bone-conduction headphones—as either "Clear" or "Very Clear." Notably, there were no reports of technical or usability issues, indicating a dependable and seamless user experience.

Comparative feedback from participants highlighted SnapStick's clear advantage over existing tools. Among the eleven participants, eight had prior experience using Microsoft's SeeingAI. All eight reported that SnapStick offered superior comfort, ease of use, and accuracy of environmental descriptions. Participants consistently emphasized that SnapStick's outputs were more detailed and contextually relevant compared to SeeingAI.

These qualitative impressions align with findings from prior work. For instance, Gonzalez Penuela's scene description app received a satisfaction score of only 2.76 out of 5 (55.2%) [23], and SeeingAI achieved a user satisfaction rate of just 43% [17]. Lin's navigational aid scored higher at 3.87 out of 5 (77.5%) [3], while Rao's system did not report overall satisfaction, but it was noted that 85.1% of users found the system "comfortable" or "very comfortable" [24].

To further assess SnapStick's performance, we conducted a comparative analysis using a Function–Performance–Cost (FPC) framework. In our evaluation, SnapStick achieved a recognition accuracy of 94%, demonstrating reliable scene understanding across diverse settings. While SeeingAI's mAP or accuracy metrics are not publicly disclosed, comparable systems provide partial benchmarks: Lin's app achieved an mAP of 55%, Gonzalez's system achieved 65% accuracy, and Rao's framework achieved a higher mAP of 84%, though overall accuracy was not provided.

These comparisons collectively suggest that SnapStick offers a compelling balance of technical accuracy, real-time responsiveness, and user-centered design. Table 3 summarizes this comparison across key metrics. Notably, SnapStick operates independently of GPS, addressing a common concern among blind users who often report discomfort with GPS-based navigation systems in prior studies. This feature further enhances SnapStick's practicality and real-world applicability.

Despite these encouraging outcomes, SnapStick has some limitations. One important limitation of the current system is the need for a locally hosted server to run the Florence-2 vision–language model. While this setup minimizes latency and enhances user privacy, it may present a barrier to adoption among users who lack access to compatible hardware or the technical expertise required for setup and maintenance. This limitation could particularly impact users from low-income or underserved communities. Additionally, although the current validation study produced strong results, it was conducted with a small sample size. Future studies should aim to include a more diverse participant pool and assess performance across varied real-world environments.

Table 3. Comparison of SnapStick with existing assistive technologies.

Feature	SnapStick	Seeing AI	Lin's App	Gonzalez Penuela App	Rao System
Functionality	Scene description, text reading, facial expression, person describing, bus route recognition	Reading text, recognizing products and people, describing scenes, identifying currency	Scene description and rough distance calculation	Scene description	Scene description
Audio Delivery	Bone-conduction (open-ear)	Standard headphones or phone speaker	Standard headphones or phone speaker	Standard headphones or phone speaker	Bone-conduction transducer
Hands-Free Operation	Yes (via cane button)	No	No	No	Yes (voice-activated)
Offline Capability	Partial (local server)	No (cloud-based)	Partial (local server)	No (cloud-based)	No (cloud-based)
Qualitative Performance	SUS: 84.7 (A+)/90.9% user satisfaction	43% satisfaction	77.5% satisfaction	55.2% satisfaction	Not mentioned
Quantitative performance	94% Accuracy	Not mentioned	55% mAP	65% Accuracy	84% mAP
Cost	Not publicly available	Free	Not publicly available	Not publicly available	Not discussed
Privacy	Local processing	Cloud processing	Local processing	Cloud processing	Cloud processing

Looking ahead, future development will focus on improving accessibility, autonomy, and real-time functionality. We plan to enable offline support for key features such as text reading, facial expression recognition, and bus route detection by compressing the Florence-2 model through techniques like quantization and distillation, allowing for on-device inference on mid-range smartphones. To support users without technical expertise, we also aim to implement a hybrid edge–cloud system, where essential tasks are processed locally and complex operations are offloaded to the cloud when connectivity is available. To improve spatial awareness and navigation safety, we will integrate depth sensing and real-time obstacle tracking, along with vibrotactile feedback via wearable or cane-based actuators. These additions will offer intuitive, non-auditory cues in crowded or dynamic environments. Furthermore, we plan to implement adaptive learning algorithms that personalize system responses based on individual user preferences, ultimately enhancing usability and promoting greater independence for visually impaired users.

5. Conclusions

Many navigational aids have been developed as a result of advancements in assistive technology, with the goal of enhancing spatial navigation for blind people. Nevertheless, despite this increased interest, the majority of commercially available devices have not been widely adopted because of a number of drawbacks, including discomfort, inaccuracy, and

and poor design. Because blind people rely on auditory cues for navigating, existing solutions frequently require them to wear headphones that obscure external sounds. Additionally, despite being readily accessible, vibrating canes are frequently criticized for their large size, discomfort, sensitivity to impediments, and ambiguous path direction, all of which restrict their usefulness. Although useful, GPS-based apps are less dependable due to signal loss, as well as their high battery usage and reliance on internet connectivity. Additionally, a lot of sound-based apps provide users with a lot of inputs, and object detection apps can be difficult to operate and make accurate, which aggravates users. Device adoption is also significantly hampered by the associated high expenses.

SnapStick is a significant development in assistive technology, providing a complete solution that encourages greater autonomy and inclusion for blind people by fusing cutting-edge algorithms with purpose-driven design. This work emphasizes how crucial it is to create useful, approachable technologies that enable users to live more autonomous and active lives.

Author Contributions: S.S.: conceptualization, software, data curation, formal analysis, investigation, methodology, visualization, writing—original draft. G.L.B.: software. S.Z.: writing—review and editing. M.B.: writing—review and editing. W.S.: conceptualization. G.S.: software. C.B.: project administration. A.D.L.: project administration. A.D.B.: project administration, supervision, writing—review and editing. M.G.: project administration, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project “RAISE - Robotics and AI for Socio-economic Empowerment” (ECS00000035). All authors, except author 3 and 5, are part of the RAISE Innovation Ecosystem.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Comitato Etico, ASL 3, Genoa, Italy (Project Code: 2/2020 - DB id 10213) on 5 February 2024.

Informed Consent Statement: The participants provided their written informed consent to participate in this study.

Data Availability Statement: Data will be made available on special request to Shehzaib Shafique.

Acknowledgments: We would like to express our sincere gratitude to the researchers of the Unit for Visually Impaired People for their invaluable contributions to this study and assistance in the preparation of the manuscript. The authors also express special thanks to Jessica Bertolasi, Margherita Sturlese, Marta Guarischi, Gloria Calafatello, Maria Casado Palacios, Martina Riberto, and Serena Basta for their support.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

VLM Vision Language Model
SUS System Usability Scale

References

1. Riazi, A.; Riazi, F.; Yoosfi, R.; Bahmehi, F. Outdoor difficulties experienced by a group of visually impaired Iranian people. *J. Curr. Ophthalmol.* **2016**, *28*, 85–90. [[CrossRef](#)] [[PubMed](#)]
2. Manduchi, R.; Kurniawan, S.; Bagherinia, H. Blind guidance using mobile computer vision: A usability study. In Proceedings of the 12th international ACM SIGACCESS Conference on Computers and Accessibility, Orlando, FL, USA, 25–27 October 2010; pp. 241–242.

3. Lin, B.S.; Lee, C.C.; Chiang, P.Y. Simple smartphone-based guiding system for visually impaired people. *Sensors* **2017**, *17*, 1371. [[CrossRef](#)] [[PubMed](#)]
4. Zhao, Y.; Kupferstein, E.; Tal, D.; Azenkot, S. “It Looks Beautiful but Scary” How Low Vision People Navigate Stairs and Other Surface Level Changes. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 307–320.
5. Thaler, L.; Goodale, M.A. Echolocation in humans: An overview. *Wiley Interdiscip. Rev. Cogn. Sci.* **2016**, *7*, 382–393. [[CrossRef](#)] [[PubMed](#)]
6. Srikulwong, M.; O’Neill, E. Tactile representation of landmark types for pedestrian navigation: User survey and experimental evaluation. In Proceedings of the Workshop on Using Audio and Haptics for Delivering Spatial Information via Mobile Devices at MobileHCI 2010, Lisbon, Portugal, 7–10 September 2010; pp. 18–21.
7. Holbrook, M.C.; Koenig, A.J. *Foundations of Education: Instructional Strategies for Teaching Children and Youths with Visual Impairments*; American Foundation for the Blind: New York, NY, USA, 2000; Volume 2.
8. Sánchez, J.; Espinoza, M.; de Borja Campos, M.; Merabet, L.B. Enhancing orientation and mobility skills in learners who are blind through video gaming. In Proceedings of the 9th ACM Conference on Creativity & Cognition, Sydney, Australia, 17–20 June 2013; pp. 353–356.
9. Bhowmick, A.; Hazarika, S.M. An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends. *J. Multimodal User Interfaces* **2017**, *11*, 149–172. [[CrossRef](#)]
10. Khenkar, S.; Alsulaiman, H.; Ismail, S.; Fairaq, A.; Jarraya, S.K.; Ben-Abdallah, H. ENVISION: Assisted navigation of visually impaired smartphone users. *Procedia Comput. Sci.* **2016**, *100*, 128–135. [[CrossRef](#)]
11. Csapó, Á.; Wersényi, G.; Nagy, H.; Stockman, T. A survey of assistive technologies and applications for blind users on mobile platforms: A review and foundation for research. *J. Multimodal User Interfaces* **2015**, *9*, 275–286. [[CrossRef](#)]
12. Ran, L.; Helal, S.; Moore, S. Drishti: An integrated indoor/outdoor blind navigation system and service. In Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications, Orlando, FL, USA, 14–17 March 2004; pp. 23–30.
13. Cuturi, L.F.; Aggias-Vella, E.; Campus, C.; Parmiggiani, A.; Gori, M. From science to technology: Orientation and mobility in blind children and adults. *Neurosci. Biobehav. Rev.* **2016**, *71*, 240–251. [[CrossRef](#)] [[PubMed](#)]
14. Ferrand, S.; Alouges, F.; Aussal, M. An electronic travel aid device to help blind people playing sport. *IEEE Instrum. Meas. Mag.* **2020**, *23*, 14–21. [[CrossRef](#)]
15. Bujacz, M.; Skulimowski, P.; Strumillo, P. Sonification of 3D scenes using personalized spatial audio to aid visually impaired persons. In Proceedings of the International Community for Auditory Display 2011, Budapest, Hungary, 20–23 June 2011.
16. Filipe, V.; Fernandes, F.; Fernandes, H.; Sousa, A.; Paredes, H.; Barroso, J. Blind navigation support system based on Microsoft Kinect. *Procedia Comput. Sci.* **2012**, *14*, 94–101. [[CrossRef](#)]
17. Kupferstein, E.; Zhao, Y.; Azenkot, S.; Rojnirun, H. Understanding the use of artificial intelligence based visual aids for people with visual impairments. *Investig. Ophthalmol. Vis. Sci.* **2020**, *61*, 932.
18. Granquist, C.; Sun, S.Y.; Montezuma, S.R.; Tran, T.M.; Gage, R.; Legge, G.E. Evaluation and comparison of artificial intelligence vision aids: Orcam myeye 1 and seeing ai. *J. Vis. Impair. Blind.* **2021**, *115*, 277–285. [[CrossRef](#)]
19. Association, W.M. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* **2013**, *310*, 2191–2194.
20. Sauro, J.; Lewis, J.R. *Quantifying the User Experience: Practical Statistics for User Research*; Morgan Kaufmann: San Francisco, CA, USA, 2016.
21. Kuriakose, B.; Shrestha, R.; Sandnes, F.E. Tools and technologies for blind and visually impaired navigation support: A review. *IETE Tech. Rev.* **2022**, *39*, 3–18. [[CrossRef](#)]
22. Shafique, S.; Setti, W.; Campus, C.; Zanchi, S.; Del Bue, A.; Gori, M. How path integration abilities of blind people change in different exploration conditions. *Front. Neurosci.* **2024**, *18*, 1375225. [[CrossRef](#)] [[PubMed](#)]
23. Gonzalez Penuela, R.E.; Collins, J.; Bennett, C.; Azenkot, S. Investigating use cases of ai-powered scene description applications for blind and low vision people. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–21.
24. Rao, S.U.; Ranganath, S.; Ashwin, T.; Reddy, G.R.M. A Google glass based real-time scene analysis for the visually impaired. *IEEE Access* **2021**, *9*, 166351–166369.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.