Advances in Operation Research and Production Management Vol.4 Issue 3 EWA Publishing
Available Online: 20 November 2025 DOI: 10.54254/3029-0880/2025.29836

A collaborative navigation algorithm for smart wheelchairs and AI glasses based on multimodal data fusion

Zheng Yang[1*], Jingfeng Yang[1], Xiyu Huang[1], Han Kong[1], Jiacheng Yan[1], Baoshan Wang[1], Rongcan Li[1], Huayang Cao[1]

[1]Geely University of China, Chengdu, China

*Corresponding Author. Email: 267120970@qq.com

Abstract. In response to the growing demand for assistive devices under the backdrop of population aging, this paper proposes an innovative collaborative navigation algorithm integrating smart wheelchairs and AI glasses based on multimodal data fusion. The algorithm optimizes a closed-loop interaction among the "environment–user–device" triad. By integrating the wheelchair's autonomous navigation capabilities with the AI glasses' advanced environmental perception and real-time interaction functions, it significantly enhances system safety, autonomy, and user experience. The proposed approach employs a hybrid model combining Deep Belief Networks (DBN) and Stacked Autoencoders (SAE) to model and fuse multimodal data. Further integration with Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) enables improved system performance. Experimental comparisons in complex dynamic environments demonstrate that this method outperforms traditional Kalman filter–based fusion techniques. User survey results indicate a high level of acceptance and willingness to adopt this system among the target population. To promote its broader application, future research will focus on algorithm optimization, outdoor performance testing, and the enhancement of data security and privacy protection mechanisms.

Keywords: smart wheelchair, AI glasses, multimodal data fusion, collaborative navigation, user interaction

1. Introduction

The process of global population aging is accelerating rapidly. According to data from the United Nations, by 2050, individuals aged 65 and above will account for 16% of the world's population, up from 9% in 2017. Furthermore, statistics from the World Health Organization indicate that approximately one billion people worldwide live with disabilities. Many of them rely on others for mobility, facing significant restrictions on their independence. With advances in science and technology, their expectations for intelligent assistive devices are shifting from basic functional support to fully autonomous, all-scenario assistance. Establishing a closed-loop system that optimizes the interaction among the "environment–user–device" can effectively enhance their independence and quality of life.

Existing smart wheelchairs have achieved basic navigation capabilities, such as LiDAR-based obstacle avoidance and preset route planning. However, they still face limitations in perceiving complex dynamic environments and enabling natural human–machine interaction. Traditional smart wheelchairs rely primarily on LiDAR and ultrasonic sensors, which struggle to detect low-level obstacles or moving targets—leading to an accident rate of up to 12% in complex indoor and outdoor environments [1]. Current mainstream control methods, including brain–computer interfaces and gesture recognition, also have notable shortcomings. These approaches require users to maintain a high level of sustained concentration during operation, thereby increasing their cognitive and physical load. Moreover, existing environmental perception technologies remain inadequate in certain situations—for instance, failing to provide timely alerts such as "detour ahead," which disrupts the continuity of user experience and can even lead to safety risks [2]. By contrast, smart glasses such as Meta Ray-Ban and Envision Glasses exhibit superior performance in environmental perception and real-time interaction. Their visual enhancement functions, supported by ultra–wide-angle cameras and deep learning algorithms, enable real-time recognition of traffic signs, stair gradients, and other environmental details. Experimental data show an impressive recognition accuracy of up to 93%, demonstrating their robust capability in processing complex environmental information [3]. Both technologies share a core foundation in multi-source sensing and intelligent algorithms. Therefore, integrating smart wheelchairs with AI glasses could yield a highly efficient and safe mobility system, forming a closed-loop process that unifies "environmental perception–decision control–user feedback."

Such a system would not only empower individuals with disabilities and the elderly to participate more fully in social life but also improve their quality of life and promote greater social inclusivity.

At present, the collaborative application of AI glasses and mobile assistive devices remains in the early research stage, with several technical bottlenecks yet to be overcome. Data fusion mechanisms and system reliability frameworks are still immature. Research at the Massachusetts Institute of Technology indicates that in path planning, spatiotemporal calibration errors between visual sensors and LiDAR may lead to conflict issues [4]. Moreover, in the collaborative use of AI glasses and mobile devices, there exists a risk of privacy breaches if user data are maliciously exploited by third parties. Therefore, while promoting technological innovation and practical applications, special attention must also be paid to ensuring data security and protecting user privacy [5].

2. Research background

2.1. Intelligent wheelchair

Research on intelligent wheelchairs can be traced back to the early 1980s, when the focus was primarily on applying mobile robotics technology to assistive devices. Connell and Viola from the IBM T.J. Watson Research Center in the United Kingdom were among the first to integrate sensor measurement and control technologies into a standard electric wheelchair, creating what is widely recognized as the world's first intelligent wheelchair [6]. Consequently, early intelligent wheelchairs primarily relied on simple sensors for obstacle detection and avoidance. While these technologies performed reliably in static environments, their adaptability to dynamic environments remained limited. With the progressive development of embedded systems and microprocessor technologies, the functionality of intelligent wheelchair control systems has become increasingly powerful [7]. For instance, wheelchairs based on FPGA + ARM architectures can perform low-power edge computing, efficiently executing complex navigation and control tasks [8]. Bandara proposed a gesture classification model based on skeletal angle compensation, which extracts finger joint angles and palm velocity features using a Leap Motion sensor and employs an Artificial Neural Network (ANN) for dynamic gesture recognition, achieving an accuracy rate of 95%. This significantly enhanced the naturalness and fluency of user interaction [9]. Zhang et al. developed a Brain-Computer Interface (BCI)-based intelligent wheelchair that uses Electroencephalogram (EEG) signals for precise control, greatly improving the mobility autonomy of patients with severe motor impairments. In addition, recent designs of intelligent wheelchairs have increasingly incorporated ergonomic principles to enhance user comfort and safety [10].

The evolution of intelligent wheelchairs highlights the profound impact of technology on daily life. From manually operated models to modern intelligent systems, this transformation integrates advancements from multiple disciplines, including mechanical engineering, electronics, and biomedical technology. Whereas early intelligent wheelchairs emphasized electric drive systems and basic navigation functions, contemporary designs focus more on user experience and functional diversity. With ongoing progress in artificial intelligence algorithms, future intelligent wheelchairs are expected to interpret and respond to user intentions with greater precision, ultimately providing more personalized and adaptive assistive services.

2.2. AI glasses

The voice recognition technology in AI glasses enables users to issue verbal commands to control the wheelchair's movements, such as moving forward, backward, or turning, significantly simplifying the operation process and improving usability. In addition, speech synthesis technology can provide real-time voice prompts when the wheelchair encounters obstacles or needs to plan a path, thereby enhancing the user experience. Currently available AI glasses typically use cameras and depth sensors to detect road signs, pedestrians, and obstacles, allowing real-time monitoring of the surrounding environment. By overlaying virtual navigation paths within the user's field of view, they provide Augmented Reality (AR) guidance, helping users understand the road environment more intuitively. For visually impaired users, AI glasses leverage image recognition to identify surrounding objects and scenes and deliver real-time feedback via audio or vibration, effectively serving as a visual substitute and facilitating safer mobility.

The AI glass voice recognition system receives spoken commands and translates them into wheelchair movements, simplifying operation and enhancing user convenience. Simultaneously, speech synthesis provides timely guidance when obstacles are detected or path planning is required, delivering a more user-friendly experience [11]. Commercial AI glasses generally combine camera and depth sensor data to recognize road signs, pedestrians, and obstacles, while AR overlays strengthen navigation cues and improve situational awareness [12]. For visually impaired users, AI glasses convert visual information into auditory or haptic signals using image recognition, enabling safer and more efficient navigation [13].

By integrating voiceprint recognition and semantic parsing technologies, AI glasses can further improve operational convenience. However, a key challenge lies in seamlessly integrating these voice commands with the wheelchair's embedded control system, ensuring both system safety and low-latency response, thereby enabling natural and intuitive multimodal user interaction. In complex environments, improving the accuracy of voice command recognition and the timeliness of responses is crucial. To address this, a dynamic noise reduction algorithm tailored to individualized voice models has been developed. Furthermore, by integrating lidar data from the wheelchair with visual sensor inputs from the AI glasses, the system provides critical technical support for collaborative navigation. Optimizing both algorithmic and hardware configurations ensures rapid and accurate command execution, laying a solid foundation for the widespread adoption of intelligent assistive devices and enhancing overall user experience.

2.3. Multimodal data fusion algorithm

Multimodal data fusion algorithms are increasingly prevalent in intelligent navigation systems, encompassing various techniques such as sensor fusion, feature-level fusion, and decision-level integration. By combining data from different types of sensors—such as vision, Inertial Measurement Units (IMU), and GPS—these algorithms significantly enhance the accuracy and stability of navigation systems. For instance, Kalman filtering is widely used for precise localization and attitude estimation. By integrating visual and IMU data, the temporal drift in estimation errors can be reduced to under ten minutes. Within an IMU, fusing data from gyroscopes and accelerometers through Kalman filtering effectively eliminates noise and bias, yielding system states that closely approximate the true values [14]. Moreover, multimodal fusion enables more accurate environmental perception, as the system can identify obstacles through the combined interpretation of features across different data modalities. By further integrating decision outcomes from various sensor channels, the reliability and safety of the navigation system are substantially improved.

3. Feasibility analysis

3.1. Analysis of algorithm design

Although substantial progress has been made in both smart wheelchair systems and AI glasses, existing research on multimodal data fusion and collaborative navigation still faces several limitations. For instance, the adaptability of multimodal fusion algorithms in dynamic environments needs improvement, data synchronization requires higher precision, user intent recognition must ensure continuity, and issues related to communication latency and power consumption within system integration remain to be further optimized [15,16]. Research on collaborative navigation algorithms for smart wheelchairs and AI glasses based on multimodal data fusion is therefore of critical importance in addressing these challenges. By deeply integrating visual information from AI glasses with wheelchair sensor data, the proposed SensorData-based system achieves enhanced environmental perception and more precise navigation, improving both user safety and autonomy. Specifically, the multimodal data fusion algorithm integrates multiple sensor inputs—such as visual data, IMU, and GPS—thereby enhancing the accuracy and robustness of the navigation system [17]. At the same time, optimization of user intent recognition allows for a more natural and intuitive mode of interaction.

The original motivation for this study stems from a commitment to addressing the real needs of people with disabilities and mobility impairments, guided by a human-centered design philosophy. Through in-depth research and systematic analysis of the practical difficulties faced by target users, the study strictly adheres to established product development standards, encompassing key stages such as structural and functional design, computer modeling, and simulation experiments. Feasibility and performance were comprehensively evaluated and validated through digital simulation methods. As illustrated in Figure 1, the complete development workflow has been clearly visualized using three-dimensional modeling.

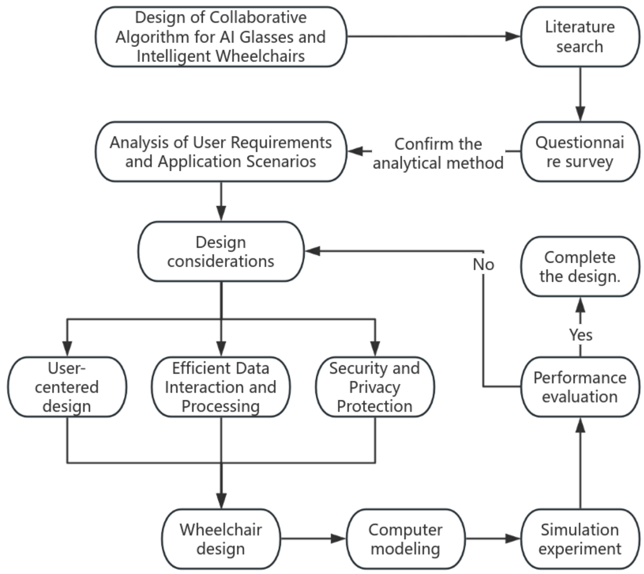


Figure 1. Design flowchart

Compared with conventional intelligent wheelchairs, traditional designs—although simpler in operation—still impose functional barriers for certain users, as they require a basic level of physical mobility. The proposed innovation breaks through these physical limitations by establishing a collaborative system between AI glasses and smart wheelchairs, using audio-frequency signals as the medium of interaction. This integrated application not only optimizes human–machine interaction but also creates an accessible mode of control for users with severe motor impairments. Consequently, both the precision and ease of wheelchair control are greatly improved, empowering users with enhanced autonomy and dignity in daily life.

3.2. Analysis of the questionnaire survey

To gain a deeper understanding of the needs and acceptance levels of the target user group regarding the collaborative navigation system of smart wheelchairs and AI glasses, a structured questionnaire survey was designed and implemented. The survey aimed to assess users' attitudes toward intelligent wheelchair navigation technologies and their practical demands, in light of the growing importance of assistive devices for individuals with limited mobility as well as the global trends of population aging and increasing disability prevalence. The survey respondents primarily consisted of elderly individuals, persons with disabilities, and others with potential needs for intelligent assistive devices. The questionnaire was designed comprehensively, covering aspects such as users' habits in using assistive devices, acceptance of emerging technologies, emphasis on system safety, and expectations regarding functional features. To ensure that the survey accurately reflected user requirements and expectations, particular attention was paid to several key parameters: operational convenience of the smart wheelchair and AI glasses, accuracy of environmental perception, personalization needs, and awareness of data security and privacy protection. The feedback collected from the questionnaire provides strong empirical support for algorithmic design, helping to more precisely identify and address users' pain points and practical challenges during the use of intelligent assistive devices.

Based on an analysis of user behavioral characteristics and application scenarios, a specialized questionnaire was developed to capture the practical requirements of people with disabilities for intelligent wheelchair design. A total of 500 paper questionnaires were distributed to the target population, and 421 valid responses were collected. The survey data are presented in Table 1.

Table 1. Survey questionnaire data table

Question	Options	Percentage
Experience using a wheelchair or assistive mobility device	Yes	71.73%
	No	28.27%

Whether the current wheelchair has navigation functions	Yes	21.14% (all individuals with wheelchair usage experience)
	No	78.86% (all individuals with wheelchair usage experience)
Main problems encountered when using wheelchair navigation (multiple choices)	Inaccurate navigation	61.04% (all individuals with wheelchair usage experience)
	Outdated map updates	34.44% (all individuals with wheelchair usage experience)
	Complicated operation	50.36% (all individuals with wheelchair usage experience)
	Lack of real-time traffic information	30.17% (all individuals with wheelchair usage experience)
Awareness or experience of using AI glasses	Yes	16.15%
	No	83.85%
Desired functions of AI glasses when integrated with wheelchairs (multiple choices)	Real-time navigation guidance	84.32%
	Environmental obstacle detection and alerts	74.35%
	Voice assistant interaction	64.85%
	Call functionality	29.45%
Willingness to try a collaborative navigation system for smart wheelchairs and AI glasses based on multimodal data fusion	Very willing	30.88%
	Willing	40.62%
	Uncertain	19.24%
	Unwilling	9.26%

The survey results show that most respondents possess a certain level of understanding and experience with existing smart wheelchair and AI glasses technologies. In general, the acceptance rate for the concept of collaborative navigation between smart wheelchairs and AI glasses is relatively high. Respondents believe that such an innovative system could significantly enhance their mobility convenience and safety. Particularly in complex environments, users expressed strong demand for functions that can automatically identify obstacles, plan safe routes, and provide real-time navigation prompts.

By analyzing the 421 valid questionnaires and conducting interviews with several individuals with disabilities, four major dimensions were summarized: user experience with smart wheelchairs, perceptions of AI glasses, effectiveness of collaborative navigation, and degree of technological acceptance. A further correlation analysis was performed to examine the relationships among these variables, and the specific results are presented in Figure 2.

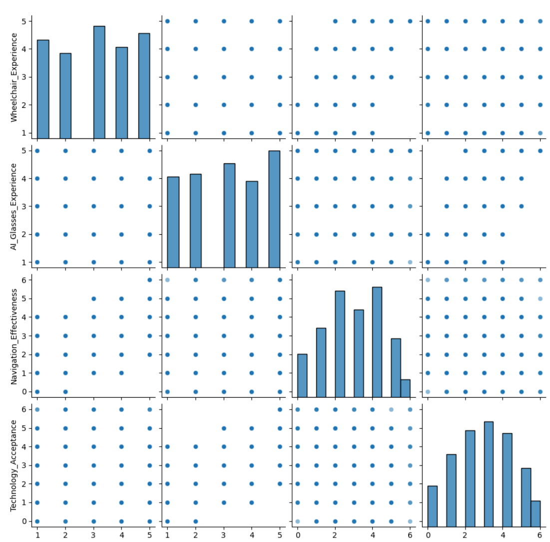


Figure 2. Correlation analysis bar chart

The correlation analysis reveals a significant positive relationship between user satisfaction with smart wheelchairs and their acceptance of AI glasses, indicating that a positive wheelchair experience enhances user trust and willingness to adopt AI glasses. Moreover, user satisfaction with the performance of collaborative navigation shows a strong positive correlation with overall technological acceptance. The calculated correlation coefficients demonstrate a clear linear relationship between the practical effectiveness of the collaborative navigation system and users' readiness to embrace the technology. Additionally, the survey highlights users' diverse expectations and preferences regarding personalized functions of both smart wheelchairs and AI glasses. For instance, some respondents emphasized the need for enhanced environmental perception capabilities to provide more accurate navigation and obstacle alerts, while others preferred improved voice assistant and communication features to facilitate convenience and social interaction.

From the correlation analysis bar chart, it can be observed that a cross-enhancement effect exists between the perception dimensions of the two devices. The correlation coefficient (r) between wheelchair experience and perception of AI glasses is 0.144, and between the wheelchair and collaborative navigation is 0.132, suggesting that the combination of physical interaction and digital interface helps form a perceptual closed loop. Furthermore, AI glasses act as the core hub of technological acceptance, with a correlation coefficient (r) of 0.209 between interaction design and user trust in the intelligent assistive ecosystem. By quantifying the coupling mechanism between the two devices, the algorithm design achieves three key breakthroughs:

Anchoring AI glasses as the fusion hub—Utilizing eye-tracking technology and augmented reality interfaces to reduce cognitive load in human–machine interaction.

Developing a dynamic physical co-compensation algorithm—In response to the weak correlation between wheelchair and navigation performance ($r = 0.132$), this algorithm converts mechanical motion signals into spatial perception cues for the user.

Constructing a dual-channel feedback system—To address low inter-device correlation, cognitive mapping was enhanced on the glasses side, while somatosensory integration was strengthened on the wheelchair side. This integration creates a synergistic assistive tool that combines digital guidance with physical response, establishing a new paradigm for cross-modal intelligent rehabilitation devices.

4. Algorithm establishment

To achieve collaborative navigation between the intelligent wheelchair and AI glasses, this study proposes an innovative multimodal data fusion algorithm designed to enhance the real-time performance, accuracy, and robustness of the navigation system. By integrating diverse sensor data—including visual information, IMU readings, and GPS positioning—from both devices, the system achieves more precise environmental perception and navigation control. To enrich environmental understanding, deep learning techniques are employed to extract multimodal data features, which are then combined through feature-level fusion to form an integrated environmental representation. At the decision-level fusion stage, the algorithm comprehensively considers information from different sensors to produce reliable navigation decisions.

4.1. Multimodal joint distribution modeling (based on DBN)

In multimodal joint distribution modeling, the correlations among multiple data sources must be considered. Assuming that the spatial data of the wheelchair ($\mathrm{x}_{\mathrm{space}}$) and the visual data of the AI glasses ($\mathrm{x}_{\mathrm{vision}}$) are independently encoded into hidden representations h_1 and h_2 , the joint Restricted Boltzmann Machine (RBM) is used for further learning. The summation operator represents the accumulation over all possible hidden and visible states:

$$P\left(x_{space}, x_{vision}, h_1, h_2|\theta\right) = \sum_{h_1}\sum_{h_2} P\left(x_{space}|h_1\right) P\left(h_1|h_2\right) P\left(h_2|\theta\right) \tag{1}$$

Where, h_1 and h_2 denote the latent representations of spatial and visual data, respectively.

4.2. Feature fusion (based on SAE)

As shown in Figure 3, the Stacked Autoencoder (SAE) is used to achieve multimodal feature fusion. When combining multiple input modalities, weighted summation is typically applied to the encoded hidden features of each data source. Let x_{space} and x_{vision} denote the spatial data from the wheelchair and visual data from the AI glasses, respectively. The fused representation can be expressed as:

$$\hat{y} = f\left(\sum_{i=1}^{N_{space}} W_{space,i} \cdot x_{space,i} + \sum_{j=1}^{N_{vision}} W_{vision,j} \cdot x_{vision,j} + b\right) \tag{2}$$

where N_{space} and N_{vision} represent the feature dimensions of spatial and visual data, respectively, and bbb denotes the bias term.

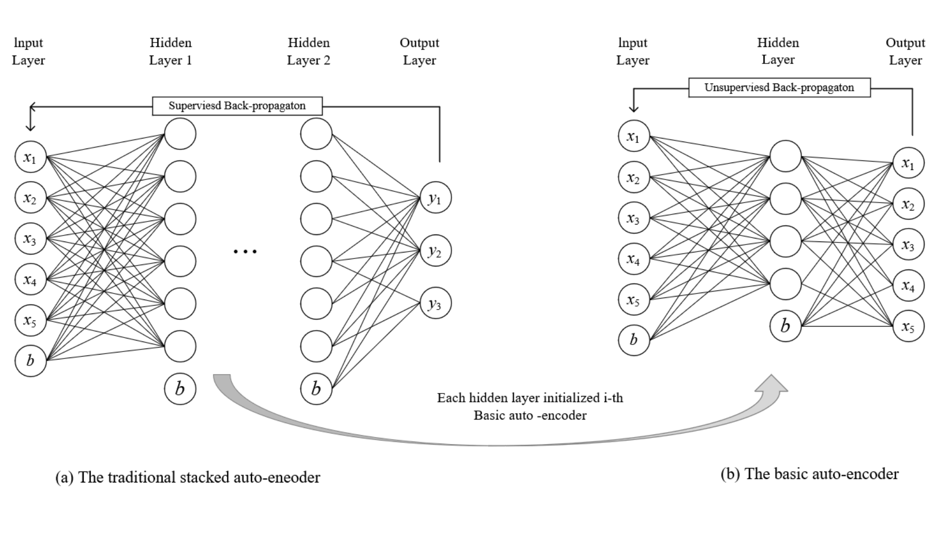


Figure 3. Structure of the stacked autoencoder

4.3. Convolutional feature extraction (based on CNN)

In the Convolutional Neural Network (CNN), convolution is performed by taking the element-wise product between the input and the kernel, followed by summation. This operation is applied to each local region of the input to generate feature maps, which are then processed by pooling layers to enhance feature invariance:

$$F_o = f\left(\sum_{i=1}^{M}\sum_{j=1}^{N} (F_i{*}K)_{ij} + b\right) \tag{3}$$

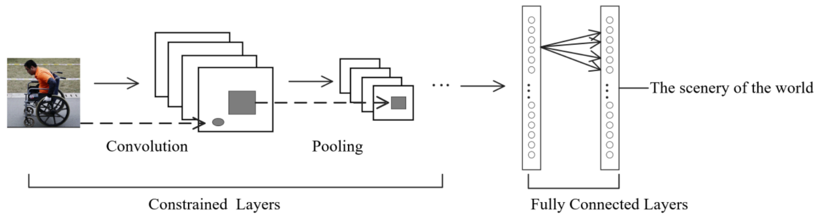
4.4. Temporal decision modeling (based on RNN)

In the Recurrent Neural Network (RNN), sequential data are processed so that both the current input and the previous hidden state jointly determine the new hidden state and output. The update equations of RNN usually requires accumulating the calculations of all time steps, as illustrated in Figure 4.

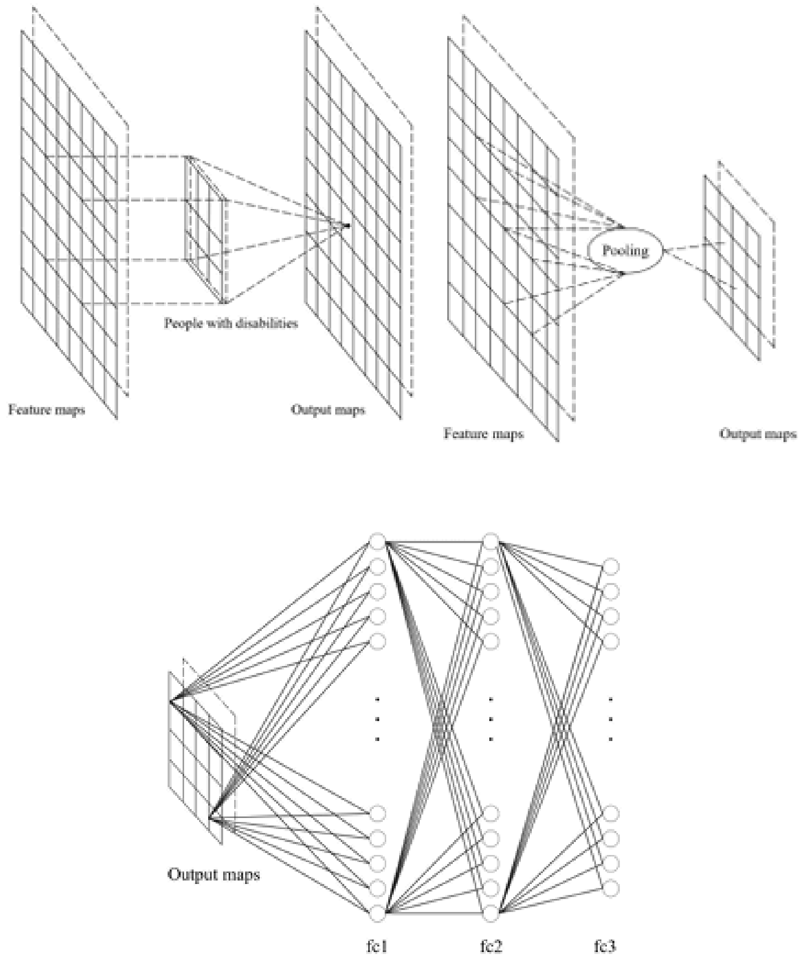
$$s_t = g\left(\sum_{i=1}^{N_{space}} W_{space,j} \cdot x_{space,i,t} + \sum_{j=1}^{N_{vision}} W_{vision,j} \cdot x_{vision,j,t} + W_{cmd} \cdot x_{cmd,t} + W_s \cdot s_{t-1} + b\right) \tag{4}$$

$$o_t = f\left(W_o \cdot s_t + b_o\right) \tag{5}$$

Here, $\mathrm{N}_{\mathrm{space}}$ and $\mathrm{N}_{\mathrm{vision}}$ denote the feature dimensions of the spatial and visual inputs, respectively, and the summations represent the accumulation across input features at each time step.



(a) CNN architecture



(b) Convolution layer (c) Pooling layer (d) Fully connected layer

Figure 4. Example of a convolutional neural network

4.5. Multimodal navigation combined with optimization

To minimize prediction errors—including positional errors, image recognition errors, and voice command errors—a joint optimization framework is established, as illustrated in Figure 5. The loss function is defined as:

$$L = \lambda_1 \cdot \sum_{i=1}^{N_{space}} \| \widehat{x}_{space,i} - x_{space,i} \|^2 + \lambda_2 \cdot \sum_{i=1}^{N_{vision}} \| \widehat{I}_i - I_i \|^2 + \lambda_3 \cdot \sum_{i=1}^{N_{cmd}} \| \hat{y}_{cmd,i} - y_{cmd,i} \|^2 \tag{6}$$

where N_{space} , N_{vision} , and N_{cmd} represent the number of samples for spatial data, visual data, and voice command data, respectively. By including the summation notation, we can more accurately represent the accumulation of multiple data points, neurons, or time steps, which is a common operation in multimodal data fusion tasks.

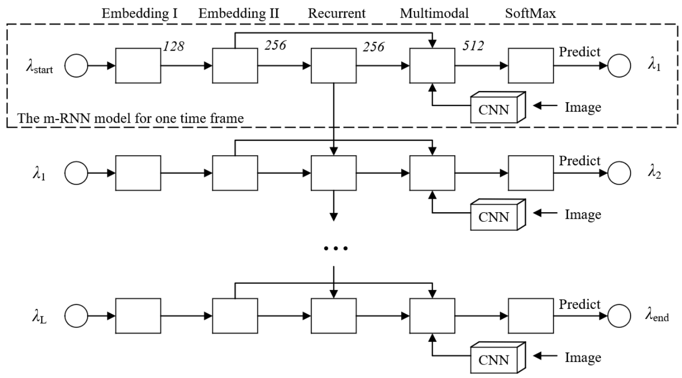


Figure 5. Structure diagram of the multimodal optimization model

During feature integration, the SAE framework extracts and refines deep-level representations to effectively combine multimodal information. This ensures that the complementarity and consistency among modalities are preserved, providing an accurate depiction of both environmental states and user intent. Moreover, an attention mechanism is introduced to dynamically adjust the weights of different modalities, allowing the system to adapt to environmental fluctuations and user needs. This unified strategy enables precise multimodal integration, laying a solid data foundation for intelligent decision-making and coordinated control in assistive navigation systems.

5. Experimental results and analysis

To further validate the effectiveness of the proposed multimodal data fusion algorithm based on the Recurrent Neural Network (RNN) and Stacked Autoencoder (SAE) frameworks, a series of comparative experiments were conducted against the traditional Kalman filter–based multimodal fusion method. The experimental results clearly demonstrate that the proposed algorithm achieves superior performance across multiple key metrics, including accuracy, Mean Squared Error (MSE), and processing time. For instance, in multi-sensor SLAM fusion, integrating heterogeneous sensors provides additional constraints that prevent SLAM degradation—consistent with our algorithm's improved robustness in localization and environmental perception. Particularly under complex and dynamic navigation scenarios, the proposed fusion model exhibits outstanding precision and stability, underscoring its strong potential and practical applicability in intelligent assistive mobility systems [18].

5.1. Design and data

In this study, we propose a multimodal data fusion algorithm based on Stacked Autoencoders (SAE) and Recursive Neural Networks (RNN). The algorithm processes sequential data using RNNs, which reduces data complexity while preserving core information, and employs autoencoders for feature extraction and dimensionality reduction, thereby achieving effective multimodal data integration. Compared with our proposed algorithm, the traditional Kalman filter-based multimodal fusion method—a classical signal processing technique—has been widely applied in multisensor data fusion scenarios.

When evaluating model or algorithm performance, key metrics are emphasized. Accuracy, which reflects the precision of environmental perception and navigation control, serves as a primary indicator of the consistency between predicted and observed values. Comparative experiments show that our proposed algorithm significantly outperforms the Kalman filter approach in terms of accuracy, with predictions closely matching actual observations, even in complex dynamic environments.

Another commonly used metric for evaluating the difference between predicted and actual values is the Mean Squared Error (MSE), calculated as the average of the squared differences between predictions and true values. Generally, a smaller MSE indicates lower prediction error and higher prediction accuracy. For example, in the area of data analysis:

$$MES = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{7}$$

where n is the total number of samples, $\mathrm{y_i}$ and $\widehat{\mathrm{y}}_\mathrm{i}$ denote the true and predicted values of the i-th sample, respectively, and $\left(\mathrm{y_i} - \widehat{\mathrm{y}}_\mathrm{i}\right)^2$ represents the squared prediction error for that sample.

We utilized publicly available multimodal datasets, including the MMSCI dataset in the scientific domain and the MultiOOD dataset for anomaly detection, to obtain a rich set of information. These datasets are not limited to Inertial Measurement Unit (IMU) and LiDAR sensor data from intelligent wheelchairs but also include visual data captured by cameras in AI glasses. By integrating multiple types of sensor and visual inputs, these datasets provide comprehensive, multidimensional data support for algorithm development. The datasets were carefully divided into training and testing sets. The training set allows the model to learn underlying patterns and relationships, providing essential data for machine learning. The testing set ensures the algorithm's accuracy and reliability in real-world applications, enabling an objective evaluation of model performance after training.

5.2. Experimental results

The proposed RNN+SAE algorithm achieved higher performance than the Kalman filtering method in terms of accuracy, recall rate, and noise robustness. By leveraging SAE for deep feature extraction and RNN for accurate temporal modeling, the algorithm demonstrates improved adaptability in dynamic and noisy environments.

Table 2. Experimental data comparison

Algorithm	Accuracy rate (%)	Recall rate (%)	Mean Squared Error (MSE)	Accuracy in noisy environment (%)
RNN + SAE	93.5	92.8	0.018	91.2
Kalman filtering	86.2	85.4	0.042	83.5

As shown in Table 2, the proposed algorithm achieves an MSE of 0.018, significantly lower than the 0.042 obtained by the Kalman filter, indicating superior precision and stability in prediction and decision-making processes. Even under noisy conditions, the algorithm maintains high accuracy and robust performance. This improvement stems from the synergistic combination of SAE and RNN, which effectively suppresses noise interference while preserving the integrity of multimodal feature representations.

6. Conclusion

This study innovatively proposes a collaborative navigation system integrating an intelligent wheelchair with AI glasses, achieving efficient and safe assisted navigation through multimodal data fusion technology. Considering the existing challenges in dynamic environment perception, user interaction, and data fusion mechanisms for current intelligent wheelchairs and AI glasses, this paper develops a multimodal data fusion-based collaborative navigation algorithm to address these limitations.

The proposed algorithm employs a Stacked Autoencoder (SAE) for feature fusion, utilizes Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for feature extraction and temporal sequence modeling, and adopts a Deep Belief Network (DBN) to model the joint distribution of multimodal data. This design significantly enhances environmental perception accuracy, robustness of navigation decisions, and system response speed. In addition, the system achieves joint optimization of the multimodal navigation model, thereby providing more precise navigation assistance and effectively reducing prediction errors. User survey results indicate that the system demonstrates broad application potential in real-world scenarios. The target user groups showed high acceptance and a strong willingness to use the system. Through the optimization of data fusion mechanisms and interactive design, the proposed system introduces a novel concept and technical pathway for the development of intelligent assistive devices.

Future research should focus on large-scale outdoor testing, algorithmic performance enhancement, and the improvement of data security and privacy protection mechanisms to ensure the system's stability and reliability and to foster user trust. With these advancements, the collaborative navigation system can provide safer and more convenient mobility solutions for people with limited mobility, thereby contributing to technological innovation and promoting a more inclusive society.

References

[1] Deligani, R. J., Hosni, S. I., & Borgheai, S. B. (2021). Multimodal EEG and fNIRS brain signal integration for enhanced neural decoding: A systematic review and bibliometric analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, *29*, 43–43. https: //doi.org/10.1109/TNSRE.2021.3089975

[2] Taborri, J., Keogh, J., & Kos, A. (2020). Sport biomechanics applications using inertial, force, and EMG sensors: A literature review. *Applied Bionics and Biomechanics,* 2020, Article 2041549. https: //doi.org/10.1155/2020/2041549

[3] Waisberg, E. (2024). Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. *Eye*, 38, 1036–1038. https: //doi.org/10.1016/S0140-6736(24)00111-2

[4] Chignoli, M., Kim, D., & Stanger-Jones, E. (2020). The MIT humanoid robot: Design, motion planning, and control for acrobatic behaviors. 2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids) (pp. 1–8). IEEE. https: //doi.org/10.1109/HUMANOIDS47582.2021.9555784

[5] Fan, X. (2024). The communication pattern and industrial transformation in the era of mobile internet. *China Media Technology*, 3, 24–27.

[6] Tao, L., Kui, Y., & Haibing, Z. (2008). Research status and development trend of Intelligent Wheelchair. *Robot Technology and Application*, 2, 1-5.

[7] Han, S. (2008). Research on software adaptability oriented to mobility in ubiquitous computing [Doctoral dissertation, Shanghai Jiao Tong University].

[8] Griffin, R. J. (2019). Footstep planning for autonomous walking over rough terrain. 2019 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids) (pp. 9–16). IEEE. https: //doi.org/10.1109/HUMANOIDS47582.2021.9555795

[9] Bandara, H. M. R. T. (2020). An intelligent gesture classification model for domestic wheelchair navigation with gesture variance compensation. *Applied Bionics and Biomechanics*, 2020, Article 9160528. https: //doi.org/10.1155/2020/9160528

[10] Zhang, X., Li, J., & Zhang, R. (2024). A brain-controlled and user-centered intelligent wheelchair. *Sensors*, 24(10), 3201. https: //doi.org/10.3390/s24103201

[11] Wensing, P. M., & Orin, D. E. (2016). Improved computation of the humanoid centroidal dynamics and application for whole-body control. *International Journal of Humanoid Robotics*, 13(1), 1550039. https: //doi.org/10.1142/S0219843615500398

[12] Featherstone, R. (2014). Rigid body dynamics algorithms. Springer.

[13] Lee, S. H., & Goswami, A. (2007). Reaction mass pendulum (RMP): An explicit model for centroidal angular momentum of humanoid robots. Proceedings of the 2007 IEEE International Conference on Robotics and Automation (pp. 4667–4672). IEEE. https: //doi.org/10.1109/ROBOT.2007.364191

[14] Gao, J., Li, P., & Chen, Z. (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5), 829–864. https: //doi.org/10.1162/neco_a_01273

[15] Katz, B., Di Carlo, J., & Kim, S. (2019). Mini cheetah: A platform for pushing the limits of dynamic quadruped control. 2019 International Conference on Robotics and Automation (ICRA) (pp. 6295–6301). IEEE. https: //doi.org/10.1109/ICRA.2019.8793865

[16] Kim, D., Jorgensen, S. J., & Stone, P. (2016). Dynamic behaviors on the NAO robot with closed-loop whole body operational space control. 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids) (pp. 1121–1128). IEEE. https: //doi.org/10.1109/HUMANOIDS.2016.7803393

[17] Sun, Z. (2022). Research on point cloud mapping and relocalization of indoor mobile robots [Master's thesis, Southwest Jiaotong University].

[18] Luo, K. (2022). A multi-robot mapping construction algorithm based on ORB-SLAM3. *Experimental Technology and Management*, 39(6), 82–91.