Journal of Artificial Intelligence Practice (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2025.080308
ISSN 2371-8412 Vol. 8 Num. 3

Design and Implementation of Smart Guide Glasses for the Blind Based on Deep Perception and Bone Conduction Technology

Liu Yan, Luan Zengxu, Gao Chang, Zheng Ziye, Zheng Hui*

School of Computer Science and Technology, Beihua University, Jilin Province, Jilin City, China
**Corresponding Author*

Keywords: Deep perception; bone conduction; smart guide; obstacle recognition; wearable devices

Abstract: Visually impaired individuals face challenges in independent mobility, as traditional assistive devices have limitations in terms of single-dimensional environmental perception and delayed real-time interaction. Current mainstream guide devices rely on ultrasound or basic obstacle avoidance logic, whose technical architecture struggles to analyze the dynamic spatial depth of complex urban environments. Technological breakthroughs in miniaturized RGB-D cameras and ToF sensors have opened up new possibilities, while bone conduction technology provides an open auditory channel, laying the foundation for non-invasive navigation. This paper delves into the cross-modal collaboration mechanism between depth perception and bone conduction, achieving three major innovations in its technical framework: lightweight depth computing units perform millimeter-level scene modeling with the support of embedded visual processors; spatial sound field modeling technology drives bone conduction audio directional prompts; and edge computing architecture ensures the efficiency of multi-sensor spatiotemporal fusion. The system identifies the attributes of dynamic obstacles through semantic segmentation algorithms, utilizes infrared assistance to mitigate strong light interference, and establishes a redundant verification mechanism for rainy and foggy environments. The ultimate goal is to reconstruct the spatial cognitive paradigm of visually impaired individuals and provide centimeter-level environmental understanding capabilities.

1. Introduction

The complexity of urban navigation scenarios poses a significant challenge, with dynamic pedestrian flows, varying terrain, and sudden obstacles creating a multi-layered decision-making system. Traditional guide devices for the visually impaired rely on ultrasonic detection for basic obstacle avoidance, but their single-point sensing mode cannot establish a spatial topology model, resulting in weak adaptability to complex environments. Deep perception technology offers a revolutionary solution: a miniature RGB-D camera emits infrared structured light to obtain scene depth information, while a ToF sensor measures the time of flight of photons to generate a 3D point cloud. The two technologies work together to establish a precise spatial modeling foundation. The

bone conduction acoustic system design breaks through the physical limitations of traditional headphones. This technology transmits audio signals to the inner ear via temporal bone vibrations while preserving the environmental sound perception channel. Core technological breakthroughs are reflected in three aspects: a multi-sensor data fusion framework (vision + IMU + GPS) achieves spatiotemporal alignment, real-time path planning algorithms predict dynamic obstacle trajectories, and lightweight processing units perform semantic segmentation under low-power conditions. These technologies collectively form a closed-loop spatial cognition system, upgrading basic obstacle avoidance to active environmental understanding capabilities.

2. Core technological theories of the smart guide system

2.1. Principles of deep perception technology

The depth perception module integrates three complementary imaging mechanisms to build spatial resolution capabilities. The RGB-D camera emits a specific coded infrared structured light pattern onto the surface of an object. Its built-in dual sensors receive texture information and distorted light spots, respectively, and calculate depth distance values based on the degree of spot deformation. The ToF component continuously emits modulated infrared light waves and captures the phase shift of reflected signals, directly inferring the target object's 3D coordinates based on the time of flight of photons, while maintaining a stable sampling rate even in low-light scenarios. The stereo vision solution deploys high-frame-rate stereo cameras to synchronously capture environmental images. Based on feature point matching algorithms, it calculates left-right parallax to generate a parallax map. Combined with pre-calibrated camera intrinsic parameter matrices, it converts pixel displacement into actual physical distances. The three technologies are fused through a spatiotemporal alignment strategy to produce a dense point cloud. Structured light excels at precise modeling of static objects, ToF performs better in responding to fast-moving targets, while stereo vision provides redundant verification in outdoor high-light environments, collectively overcoming the physical limitations of a single sensor [1].

2.2. Bone conduction acoustic transmission mechanism

The piezoelectric ceramic transducer is closely fitted to the user's temporal bone mastoid region. Voltage changes at specific frequencies drive the ceramic unit to produce microscopic mechanical deformation, directly converting electrical signals into high-frequency vibrations perpendicular to the surface of the skull. Vibrational energy is transmitted along the skeletal network toward the cranial cavity (see Figure 1). When passing through the junction between compact bone and cancellous bone, part of the acoustic energy is converted into shear waves, whose propagation efficiency is significantly influenced by bone density and contact pressure. When vibrations reach the outer lymph of the cochlea, they induce traveling wave oscillations of the basilar membrane, and the deflection of hair cell cilia generates neural electrical impulses. The open-fit design maintains the connection between the ear canal and the air, allowing environmental natural sound waves and bone conduction vibrations to be processed in parallel within the cochlea. When the auditory cortex integrates the two types of signals, it automatically enhances high-frequency speech components, while the low-frequency masking effect caused by mechanical vibrations is actively suppressed by the central nervous system [2].

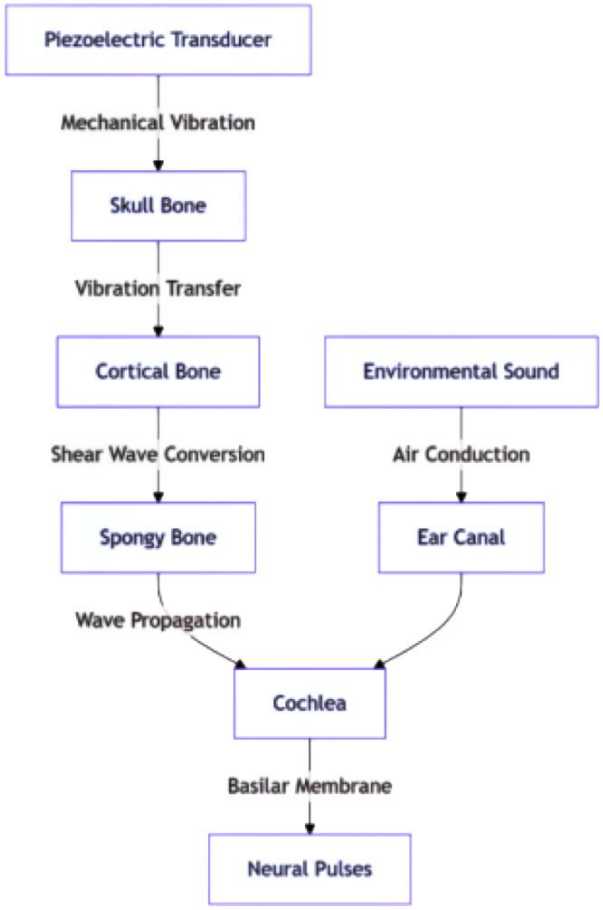

Figure 1. Anatomical diagram of bone conduction vibration path

2.3. Multi-sensor data fusion framework

The depth camera outputs a dense point cloud data stream marked with spatial coordinates of environmental feature points. The conversion of pixel coordinates to world coordinates depends on the six degrees of freedom pose changes provided in real time by the IMU. The three-axis accelerometer of the inertial measurement unit continuously samples the differential displacement of limb movements, while the angular velocity gyroscope captures the instantaneous angular acceleration of head turns. The raw data is converted into a three-dimensional pose matrix through quaternion calculation. When GPS signals are available in outdoor scenes, the latitude and longitude coordinates obtained by the receiver are mapped to Cartesian plane coordinates via UTM projection and aligned with the point cloud map reconstructed by vision through feature point matching. A Kalman filter establishes a state prediction model, where the motion vectors of visual feature points constrain the cumulative drift of the inertial trajectory. The absolute coordinates from satellite positioning periodically correct the scale errors of the entire topological map. The three data sources are fused into a unified spatio-temporal coordinate system under the synchronization mechanism triggered by hardware pulses.

2.4. Real-time path planning and obstacle avoidance algorithms

The environmental perception layer constructs a real-time 3D occupancy grid map to annotate the spatial coordinates of static obstacles. The dynamic target tracking module analyzes the

displacement of feature point clusters between consecutive point cloud frames to calculate the velocity vector of moving objects. The Kalman filter establishes a linear predictive model based on Newton's laws of motion, integrating pedestrian gait cycle characteristics and vehicle steering geometric constraints to generate a probability distribution map of future trajectories. The velocity-space projection algorithm maps the obstacle motion envelope onto the user's path plane, calculating the collision risk factor for each possible path within the collision time window. The navigation decision engine combines terrain semantic classification results with user walking preferences to generate smooth Bézier curve paths in areas where the risk coefficient is below the safety threshold. Trajectory re-planning is executed every 150 milliseconds to ensure motion continuity. When sudden obstacle acceleration changes trigger the path switching mechanism, the emergency avoidance plan is immediately activated.

3. Key technological breakthroughs in the design of smart guide glasses

3.1. Lightweight deep perception module design

The embedded vision processing chip integrates a customized sparse convolutional neural network accelerator architecture, whose parallel computing array breaks down the depth map generation process into three pipeline tasks: feature extraction layer, disparity optimization layer, and point cloud reconstruction layer. The depth sensor module integrates a miniature RGB-infrared dual-lens system and a VCSEL laser projection unit. The actively emitted 940nm wavelength structured light encoding pattern is diffused by diffractive optical elements to form a low-power area illumination. A near-threshold voltage operation circuit dynamically adjusts the supply voltage of the convolutional kernel array, and based on the analysis results of image texture complexity, it hierarchically disables the clock signals of inactive processing units. The point cloud data compression engine extracts planar geometric features from the environment to construct a sparse octree spatial index structure, retaining millimeter-level precision coordinate data only for obstacle contour regions. The motion prediction assistance module uses forward acceleration parameters from the inertial measurement unit to compensate for depth calculation delays, automatically switching to a low-resolution fast sampling mode in motion blur scenarios to maintain basic obstacle avoidance capabilities [3]. A thermal imaging sensor embedded in the nose pad of the glasses continuously monitors the temperature distribution of the processor. When the temperature in the temporal bone contact area exceeds the comfort threshold, the peak frequency of the computational core is dynamically constrained, and the thermal management strategy synchronously adjusts the duty cycle of the laser projector to reduce local heat accumulation effects.

3.2. Bone conduction audio directional prompt system

The piezoelectric ceramic transducer unit closely adheres to the skin surface of the temporal bone mastoid region to generate localized micro-vibrations. Its drive circuit modulates the pulse width of high-frequency carrier pulses based on target azimuth angle parameters. The vibrational energy is transmitted through the cranial bone medium, inducing resonance of the cochlear basilar membrane at specific frequency bands to form azimuth perception. The spatial sound field modeling engine analyzes the three-dimensional coordinate information of environmental obstacles obtained by the depth camera. Based on a personalized head-related transfer function database, it maps the spatial vector into a dual-channel vibration phase difference signal. The azimuth encoding algorithm dynamically adjusts the phase difference gradient change rate based on the relative motion speed of the obstacle to enhance directional recognition. The vibration parameter optimization module loads a finite element model of the skull transmission reconstructed from the

user's temporal bone CT scan. It compensates for individual attenuation curves in the 800 Hz to 2 kHz speech frequency band. The ear canal sound leakage monitoring unit uses a miniature pressure sensor to detect pressure fluctuations at the ear lobe position and assess vibration energy loss. The bone conduction vibration focusing algorithm optimizes the stress distribution pattern on the transducer contact surface, forming an efficient vibration coupling zone with a diameter of 12 mm in the mastoid region to enhance energy transfer efficiency. The tactile semantic encoder converts stair edge warnings into a three-pulse short vibration sequence, while pedestrian crossing guidance generates a gradually changing sine waveform. The multi-prompt conflict arbitration mechanism employs a time-division multiplexing strategy to coordinate the vibration output sequences of navigation instructions and obstacle alerts. When environmental light intensity sensor data triggers the privacy protection mode, the system automatically switches to a high-frequency micro-vibration encoding scheme. The dual-microphone beamforming array continuously monitors changes in the environmental sound field, dynamically adjusting the vibration signal-to-noise ratio to maintain speech clarity. The temporal bone contact pressure feedback unit ensures that vibration coupling stability is unaffected by head movements.

3.3. Development of multi-modal environmental perception algorithms

The multimodal data fusion processor simultaneously receives RGB image streams and depth point cloud data streams. Its dual-branch convolutional neural network architecture separately extracts visual texture features and point cloud geometric topology features. A cross-modal attention mechanism establishes pixel-level feature mapping to associate semantic information. The semantic segmentation engine analyzes the gradient distribution patterns of multi-scale feature maps. Based on an improved DeepLabv3+ architecture, it fuses hollow spatial pyramid pooling layers to capture contextual information. The edge optimization module enhances the boundary continuity of critical areas such as pedestrian crossing lines and stair edges. The dynamic obstacle recognition unit analyzes the consistency of motion in feature cluster displacement vectors between consecutive frames. It combines Kalman filter trajectory prediction with spatial matching of measured point clouds to reduce false detection rates. The static obstacle classifier distinguishes semi-transparent obstacles such as vegetation fences based on point cloud normal vector distribution characteristics and surface curvature parameters. The terrain classification decision engine integrates elevation distribution histograms with surface texture spectral features. The decision tree model divides passable areas based on slope change rates and local roughness coefficients. The real-time semantic map construction unit integrates segmentation masks and obstacle coordinates to generate a topological navigation base map. The vibration encoding protocol converter maps terrain semantic labels to bone conduction vibration waveform parameters. Paved road guidance generates low-frequency continuous waves, while gravel road warnings are converted into intermittent pulse sequences. The dynamic power consumption controller switches the neural network computation precision based on scene complexity levels. The multi-task scheduler prioritizes real-time perception refresh rates for the five-meter fan-shaped area ahead [4].

3.4. Low-latency interaction logic design

The millimeter-wave radar sensor array scans the fan-shaped area in front of the temple to capture hand movement trajectories. The micro-Doppler feature extraction algorithm analyzes the frequency shift characteristics caused by movement speed to generate gesture feature vectors. The convolutional neural network classifier maps the feature vectors to seven predefined control commands, including path re-planning requests and volume adjustment commands. The emergency braking monitoring module compares the rate of change in the user's gait acceleration with the

approaching speed of obstacles in real time. When the predicted collision time falls below the safety threshold, it automatically overrides the current command and triggers a three-level vibration alarm. Interaction arbitration logic assigns the highest priority to specific gestures. A double tap on the thigh immediately pauses all navigation outputs, and a two-second press on the temple area restores the device to factory settings. The end-to-end latency of the haptic feedback link is strictly constrained within a 50-millisecond time window. The interaction process is shown in Figure 2.

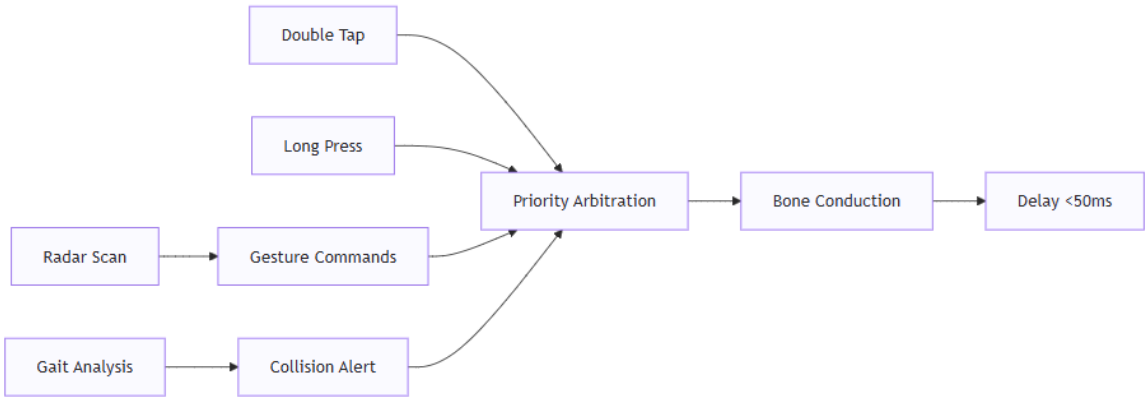

Figure 2. Interaction flowchart

4. Core challenges and solutions in system implementation

4.1. Deep perception failure in complex lighting scenarios

The multispectral imaging system simultaneously captures environmental data streams in the visible light and long-wave infrared bands. Its adaptive exposure control module calculates the illuminance histograms of each region in the image and dynamically generates multiple exposure sequences. The high dynamic range fusion algorithm aligns the gradient features of different exposure frames to preserve the contours of high-contrast targets such as window frames. The near-infrared VCSEL laser projector automatically increases the illumination power at the 940nm wavelength when ambient light intensity is insufficient. Diffractive optical elements modulate the laser phase to generate non-uniform speckle patterns, suppressing the light saturation effect in mirror-reflective regions. The dual-band infrared auxiliary channel processes depth information at 850nm and 1350nm wavelengths in parallel. The long-wave infrared sensor penetrates foggy media to capture thermal radiation differences and reconstruct the topological structure of obstructed obstacles. The multi-spectral feature fusion engine establishes a Bayesian probability model to assess the confidence weights of depth values across different wavelengths. The spatiotemporal calibration unit compensates for flight time differences between light sources of different wavelengths to eliminate multipath interference. The dynamic ambient light monitor automatically switches sensor gain parameters based on changes in solar elevation angle to maintain signal-to-noise ratio stability. The active anti-fog module's miniature thermoelectric cooler maintains a temperature gradient between the lens surface and the environment. A distributed temperature sensor network tracks the thermal conductivity coefficients of various parts of the frame in real time. The anti-condensation control algorithm dynamically optimizes the spatial distribution of heating power to avoid local overheating that could affect wearing comfort. In extreme backlighting scenarios, the depth restoration protocol prioritizes the integrity of point clouds in high-risk areas such as stair edges and suspended obstacles [5].

4.2. Balancing bone conduction voice clarity and privacy

The piezoelectric ceramic drive circuit of the bone conduction transducer employs an adaptive bias voltage regulation strategy, with its pulse width modulation waveform dynamically adjusting the carrier fundamental frequency based on the environmental noise spectrum characteristics. An array of environmental acoustic sensors continuously collects wind noise and traffic background sounds to construct a multi-channel noise suppression model. The personalized frequency response calibration module loads skull conduction simulation parameters generated from the user's temporal bone CT scan data to compensate for individual attenuation curves in the 800 Hz to 1.5 kHz speech critical frequency band. The ear canal seal detection unit uses a miniature pressure sensor to measure pressure changes at the ear canal opening to assess vibration leakage risks. The vibration energy focusing algorithm optimizes the stress distribution pattern on the transducer contact surface, creating an efficient vibration coupling zone with a diameter of less than 15 mm in the temporal bone mastoid region. This enhances the vibration energy flux density along the bone conduction sound wave transmission path while reducing the sound radiation efficiency of air conduction. The privacy protection mechanism monitors changes in ambient light intensity and automatically switches the vibration encoding protocol. In high-light scenarios, a high-frequency micro-vibration mode is activated to reduce audible noise on the skin surface. The dual-microphone feedback system detects sound leakage exceeding the threshold and immediately triggers phase reversal compensation of the drive current.

4.3. Real-time performance assurance and power consumption control

The hardware acceleration pipeline of edge computing chips deploys sparse convolutional neural network weight matrices, and its pulsed array processing architecture decomposes the depth map generation task into three levels of parallel computing units: feature extraction, disparity optimization, and point cloud compression. The dynamic voltage-frequency regulation module continuously monitors the task queue depth of each computational unit, dynamically switching the supply voltage of the computational cores based on the complexity of image textures. The convolution kernel activation mapping analyzer disables the clock signals of the multiplicative accumulators corresponding to inactive feature maps. The bone conduction vibration encoding task is offloaded to a dedicated digital signal processor for execution. Its finite state machine architecture pre-stores twelve standard navigation prompt vibration waveform templates, and during the waveform synthesis stage, it skips floating-point operations and directly calls pre-quantized integer sequences. The multi-sensor collaborative wake-up mechanism is based on the working cycle of the IMU motion state prediction environmental perception module. In stationary states, the depth camera frame rate automatically drops to 1 Hz to maintain basic obstacle avoidance capabilities. Temperature sensor data from the inner side of the eyeglass temples triggers dynamic adjustments to the heat dissipation strategy, automatically constraining the processor's peak frequency.

4.4. Enhancing robustness in extreme weather conditions

The multispectral imaging system simultaneously collects environmental data in the visible light and long-wave infrared bands. Its physical model is based on Mie scattering theory to construct a rain drop size distribution function, dynamically calculating light attenuation coefficients to compensate for depth measurement errors under different precipitation intensities. A millimeter-wave radar array emits frequency-modulated continuous waves to penetrate rain and fog media. The Doppler signal processing unit separates the reflection spectra of moving targets from

static rain droplet clutter. The point cloud spatiotemporal alignment engine fuses infrared thermal imaging and radar distance gate data to reconstruct the contours of obstacles obscured by fog. A sensor redundancy verification mechanism establishes an innovative residual detection logic for the Kalman filter. When the position estimation differences among visible light, infrared, and millimeter-wave data exceed the threshold, it automatically triggers a confidence-weighted voting decision. The active anti-fog module's miniature thermoelectric cooler maintains the lens surface temperature slightly above the environmental dew point. Its distributed temperature sensor network continuously monitors temperature gradient differences across the frame, while the anti-condensation algorithm dynamically adjusts heating power distribution patterns to prevent local overheating that could affect wearing comfort. In extreme environments, the multi-source data conflict resolution protocol prioritizes perception reliability in high-risk areas such as stair edges and suspended obstacles.

5. Conclusion

The complexity of urban navigation scenarios poses a significant challenge, with dynamic pedestrian flows, varying terrain, and sudden obstacles creating a multi-layered decision-making system. Traditional guide devices use ultrasonic detection for basic obstacle avoidance, but their single-point sensing mode cannot establish a spatial topology model, leading to poor adaptability in complex environments. The intelligent guide glasses system achieves millimeter-level spatial modeling accuracy, enabling semantic attribute recognition of dynamic obstacles (e.g., distinguishing between stationary vehicles and moving bicycles). The bone conduction module transmits spatial information through vibration frequency gradients (high-frequency vibrations indicate nearby obstacles, while low-frequency sound fields represent distant targets), establishing a visual-like cognitive mapping mechanism. A multi-sensor redundancy verification system maintains reliability in rainy or foggy environments, while an edge computing architecture ensures real-time decision-making capabilities at the 10-millisecond level. The core value of this technological framework lies in redefining the essence of human-machine interaction: when visually impaired users navigate around temporary obstacles based on bone conduction sound fields or perceive changes in road curvature through vibration cues, the smart device transitions from a tool to a spatial cognitive partner. The ultimate goal of technological evolution is to build a natural extension of perception, making environmental interaction as precise and fluid as instinct. Future research should focus on the natural integration of multimodal perception, enabling technology to become invisible and form an intangible spatial cognitive system.

Acknowledgements

This article is the outcome of the 2024 entrepreneurship project of Beihua University, titled "Chaser of Light" - Bone Conduction Braille Glasses Based on Deep Perception Technology. Project Number: 202410201007.

References

[1] Zhang Pingxin, Luo Tang, Du Haijun, et al. Design of Smart Blind-Assist Glasses Based on Augmented Reality Technology [J]. Henan Science and Technology, 2020, (22): 11-12.
[2] Huang Yulong. Smart Blind Glasses Based on Artificial Intelligence Technology [J]. Electronic World, 2017, (23): 173+175.
[3] Kang Yanan, Fan Changxi, Zhang Xueyi, et al. Application of Binocular Vision in Smart Guide Glasses [J]. Electronic World, 2018, (09): 183-184.
[4] Tang Zihang, Huang Xu, Tan Boming. Spring for the Blind: A Comprehensive Study on Intelligent Guide Glasses

Based on Machine Vision [J]. Science and Technology Economy Market, 2024, (09): 45-47.
[5] Yao Peiqi, Lu Yuming. Design and Exploration of Smart Guide Glasses Based on Inclusive Design Principles [J]. Design, 2024, 37(01): 8-11.