

面向视障人群的双目视觉与超声波融合智能避障系统

李金花 钟宇轩 王艳飞*

(广州理工学院, 广东 广州 510000)

摘要: 针对当前视障辅助设备在透明障碍识别、低矮物体探测以及动态目标感知方面存在的不足, 本文设计并实现了一种融合双目视觉与超声波信息的避障系统。系统以树莓派 4B 和 STM32 为核心, 搭建主从式嵌入式架构; 在视觉端引入改进的半全局匹配算法以支持中远距离语义感知, 同时结合五向布置的超声波阵列, 实现 0.1 - 8.0 m 范围内的近场高精度探测。通过扩展卡尔曼滤波与占据栅格地图对多源感知数据进行时间与空间上的协同融合, 并借助骨传导语音与振动手杖输出低干扰提示信息。实验结果表明, 在复杂室内外环境中, 该系统对多类型障碍物的平均识别成功率达到 98.5%, 预警延时约为 210 ms, 整体性能优于单一传感方案。系统无需依赖网络环境, 且成本可控, 为普及型无障碍智能辅助设备提供了一种可行的实现思路。

关键词: 视障辅助; 双目视觉; 超声波; 多模态融合; 避障系统; 无障碍技术

DOI: <https://doi.org/10.65196/gxqzj18>

第一章 引言

视觉功能受损会显著削弱个体的独立生活与出行能力。世界卫生组织 (WHO) 2023 年报告指出, 全球约有 2.8 亿人存在中重度视力问题, 其中约 4300 万人完全失明; 我国视障人群规模已超过 1700 万, 其日常通行过程中常需面对玻璃门、水坑、台阶及移动车辆等多种潜在风险。传统白手杖难以识别悬空或透明障碍, 而现有电子辅助设备多依赖单一传感方式: 视觉方案虽然能够提供一定语义信息, 但易受光照变化影响且近距离存在盲区; 超声波传感器响应速度快、成本较低, 却难以区分障碍物类型, 环境理解能力有限。

近年来, 多传感器融合被认为是提升系统稳定性与环境适应能力的重要方向。王远新等人^[6]将 CNN 与激光雷达技术结合应用于导盲眼镜设计, 虽在感知能力上有所提升, 但较高成本制约了其推广应用; Meshram 等^[2]通过融合摄像头与超声波实现物体识别, 但在透明障碍及地面断差检测方面仍存在不足。同时, 一些可穿戴辅助设备更强调轻量化设计^[1], 却未能在感知可靠性与交互直观性之间取得理想平衡。

针对以上问题, 本文设计了一套面向视障用户的低成本智能避障方案。系统以双目视觉与五向超声波阵列作为主要感知手段, 在改进的半全局匹配 (SGM) 算法和扩展卡尔曼滤波算法下实现多源信息融合, 并结合局部占据栅格地图完成安全决策。交互层采用骨传导语音与振动手杖输出方向提示, 以降低干扰。实验结果表明, 在复杂室内外环境中系统对障碍物的平均识别成功率达到 98.5%, 端到端预警延迟约 210 ms, 无需依赖网络即可运行, 具有较好的实用性与推广价值。

第二章 系统设计与实现

2.1 系统需求分析

面向视障群体的实用型避障系统需要在功能需求、运行性能与可靠性之间取得合理平衡。本文从实际使用场景出发, 要求系统能够以不低于 5 Hz 的频率, 对 0.2 m - 8.0 m 范围内前方及侧前方的静态与动态障碍物进行实时感知, 并具备对行人、车辆、杆状物以及地面断差等目标的基础分类能力, 同时输出明确的通行方向提示; 端到端延迟需控制在 300 ms 以内, 在日常光照变化条件下保持稳定工作, 单次充电续航时间不少于 6 h, 头戴部分重量限制在 500 g 以内。此外,

作者简介: 李金花 (2003-), 女, 本科, 研究方向为人工智能、自动化。

钟宇轩 (2005-), 男, 本科, 研究方向为人工智能、自动化。

通讯作者: 王艳飞 (1997-), 助教, 研究生。

系统还需具备一定的容错能力，在局部模块失效时仍可提供基本安全提示，整体误报率与漏报率分别控制在 2%和 5%以内，以增强用户信任。

2.2 系统总体架构

为实现上述需求，本文设计了如图 1 所示的四层系统架构，自左而右分别为感知层、处理层、决策层与交互层。

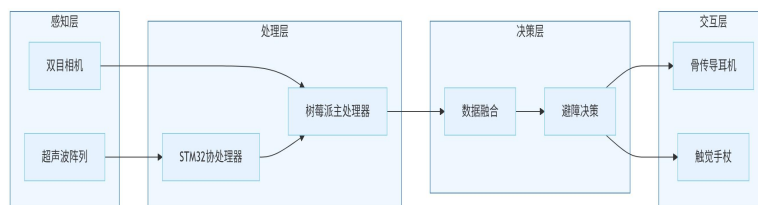


图 1 系统总体架构图

2.3 硬件与传感器选择

在兼顾感知性能与佩戴舒适度的前提下，系统采用眼镜式可穿戴结构，并参考“多功能智能导盲眼镜”设计理念^[1]，强调日常隐蔽性与用户接受度。树莓派主机、电池及电源管理模块置于随身腰包中，智能手杖则在原有结构上加装振动单元和蓝牙接收模块。精确的多传感器标定是数据融合的前提：

1. 双目相机标定：采用张正友平面标定方法，借助高精度棋盘格标定板，通过 OpenCV 采集多组图像，求解出左右相机的内参矩阵 K_l, K_r 、畸变系数 D_l, D_r 以及描述两者相对位置姿态的外参矩阵（旋转矩阵 R 和平移向量 T ），随后完成立体校正，使图像行方向严格对齐，从而将匹配过程简化为一维搜索。

2. 超声波传感器标定：在标准温湿度条件下，于已知距离范围（0.1 - 4.0 m）内采集测量数据，建立测距值与实际距离之间的对应关系，用于修正非线性误差与温漂影响。

3. 手眼标定：由于超声波与双目相机固定安装，需要确定二者的空间变换关系 T_u^c 。通过在公共视野内布置靶标，同时获取其在相机与超声坐标系中的位置，利用最小二乘法求解刚体变换，实现多传感器空间对齐。

第三章 系统核心算法

系统软件算法的核心任务是高效、鲁棒地将双目视觉的丰富信息与超声波测距的可靠信息融合起来，形成统一、准确的环境态势感知。

3.1 双目视觉深度恢复

双目视觉深度恢复的核心原理是基于三角测量法。如图 2 所示，空间点 P 在左右两个摄像机 C_1 和 C_2 的成像平面上分别投影为点 p_1 和 p_2 。若仅使用单个摄像机（如 C_1 ），则点 P 可能位于光心 O_1 与像点 p_1 连线上的任意位置，其深度信息无法唯一确定。当使用两个摄像机同时观测时，点 P 可通过两条投影射线 O_1p_1 与 O_2p_2 的交点唯一确定，从而恢复其三维坐标。

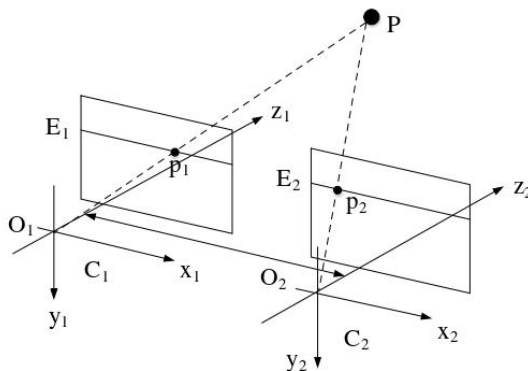


图2 双目视觉立体成像模型

为简化计算，需对左右图像进行立体校正，使其重投影至共面的图像平面上，确保匹配点仅存在于同一水平扫描线上。对于校正后的图像对 $I_L(x, y)$ 和 $I_R(x, y)$ ，其对应点的视差 d 定义为：

$$d = x_L - x_R$$

其中 x_L 和 x_R 分别为匹配点在左右图像中的水平坐标。视差 d 与深度 Z 成反比关系。

为精确计算稠密视差图 D ，本文采用改进的半全局立体匹配算法。该算法在传统 SGM 框架基础上，融合了 Census 变换与梯度信息作为匹配代价，增强了在弱纹理和视差不连续区域的鲁棒性，同时保持了较高的计算效率，适用于嵌入式平台。

对于视差图中一点 $p(u, v)$ ，其视差值为 d_p ，则该点在相机坐标系下的三维坐标 $P_c = [X_c, Y_c, Z_c]^T$ 可通过以下公式计算：

$$Z_c = \frac{f \cdot B}{d_p}, \quad X_c = \frac{(u - c_x) \cdot Z_c}{f}, \quad Y_c = \frac{(v - c_y) \cdot Z_c}{f}$$

其中， f 为相机焦距（像素单位）， B 为基线长度（物理单位）， (c_x, c_y) 为主点坐标。

通过上述过程，可将二维图像序列转换为包含空间距离信息的三维点云，为后续的多传感器融合提供几何基础。

3.2 基于扩展卡尔曼滤波的动态目标跟踪

对于场景中的动态障碍物（如行人），简单的单帧检测无法提供其运动趋势。扩展卡尔曼滤波器是一种适用于非线性系统的递归状态估计器，非常适合融合双目和超声波对同一运动目标的不完全观测，以得到更平滑、更准确的轨迹估计。

1. 状态模型：定义状态向量为 $\mathbf{x}_k = [p_x, p_y, v_x, v_y]^T$ ，包含目标在二维水平面（地面坐标系）的位置和速度。采用匀速运动模型，状态转移矩阵为 \mathbf{F}_k ，过程噪声协方差为 \mathbf{Q}_k 。

2. 观测模型：

双目视觉观测：通过目标检测 YOLOv8n 模型获得目标在图像中的边界框，结合其所在区域的视差，可计算出目标中心点的三维坐标 ${}^cP = [{}^cX, {}^cY, {}^cZ]^T$ ，再转换到地面坐标系得到观测向量 $\mathbf{z}_k^{st} = [p_x^{st}, p_y^{st}]^T$ ，观测噪声协方差为 \mathbf{R}_k^{st} 。

超声波观测：当目标进入超声波波束内，可获得传感器到目标的径向距离 r 。结合传感器自身在地面坐标系下的已知位置和朝向，可得到一个位于该距离圆弧上的位置约束，可作为另一个观测 \mathbf{z}_k^{us} ，噪声协方差为 \mathbf{R}_k^{us} 。

3. EKF 融合过程：遵循“预测&更新”框架。在预测步骤，利用状态模型估计下一时刻状态。在更新步骤，当双目或超声波任一观测到来时，便利用其观测模型（在非线性情况下计算雅可比矩阵 \mathbf{H}_k ）和对应的噪声协方差，对预测状态进行修正。通过交替或顺序处理两种观测，实现了对动态目标状态的持续、融合估计，其估计误差协方差 \mathbf{P}_k 小于任一单一传感器。

3.3 基于占据栅格地图的静态环境建模

为统一表示静态环境和未被跟踪的动态障碍物，系统维护一个二维的占据栅格地图。该地图将环境离散化为大小固定的网格（5cm×5cm），每个网格 m_i 包含一个其被占据的概率值 $P(m_i|z_{1:t})$ 。

系统采用二进制贝叶斯滤波来递归更新每个栅格的占据概率。使用对数几率形式可以避免概率值接近 0 或 1 时的数值不稳定问题。定义栅格 m_i 在时刻 t 的对数几率为：

$$l_{t,i} = \log \frac{P(m_i|z_{1:t})}{1 - P(m_i|z_{1:t})}$$

其递归更新公式为：

$$l_{t,i} = l_{t-1,i} + \log \frac{P(m_i|z_t)}{1 - P(m_i|z_t)} - l_0$$

其中， l_0 是栅格的先验对数几率（设为 0，先验占据概率为 0.5）。 $P(m_i|z_t)$ 是逆观测模型，它定义了当传感器获得观测 z_t 时，栅格 m_i 被占据的概率。对于双目视觉生成的三维点云，将其投影到地面并离散到对应栅格，这些栅格获得一个高的“被占据”观测概率。对于超声波返回的一个距离值，认为该距离内的射线经过的栅格是空闲的（低占据概率），而恰好位于该距离处的栅格是被占据的（高占据概率）。

通过持续融合来自双目和超声波的观测数据，地图逐渐收敛，准确地反映了环境中静态障碍物（如墙壁、桌椅）和临时静态障碍物（如停放的自行车）的位置。

3.4 避障决策与多模态交互生成

决策模块基于实时更新的占据栅格地图和动态目标轨迹，采用一种轻量化的向量场直方图（Vector Field Histogram, VFH）算法变种进行实时避障决策。该算法将用户当前位置为中心的前方半圆区域划分为若干扇形方向单元，并为每个单元计算一个综合的“危险度”值。此危险度由该扇形区域内所有栅格的占据概率加权其与用户的距离得出，从而同时反映了静态障碍物的密度分布与动态目标的接近趋势。

算法的核心是构建一个如图 3(a) 所示的极坐标直方图，其横轴（ α ）代表以机器人为中心的障碍物方位角（ -180° 至 180° ），纵轴（ p ）代表该方向存在障碍物的概率密度。

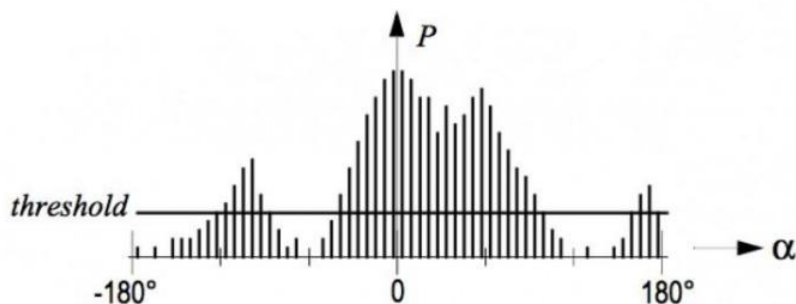


图 3(a) VFH 极坐标直方图

系统首先识别出所有宽度足以通行的“空隙”（即连续的低危险度扇形区域），然后通过一个代价函数（综合考虑目标方向、转向成本与通过宽度）从中选择最优的前进方向。图 3(b) 直观展示了在实际环境栅格地图中，以用户为中心的扇形探测区域与计算出的安全前进方向。

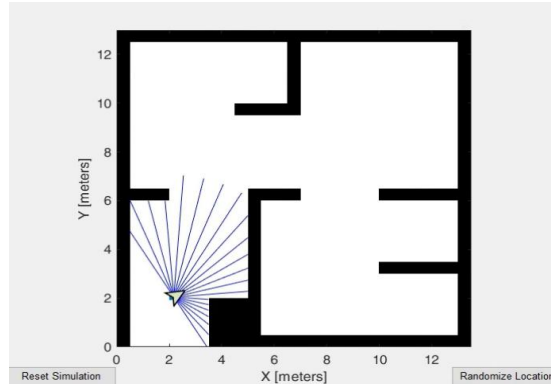


图 3(b) 基于占据栅格地图的扇形方向单元划分与安全方向决策示意图

决策结果通过双模态交互通道直觉地传达给用户，听觉通道通过骨传导耳机播报结构化的自然语言提示 T_{audio} ，其内容由决策输出与感知结果共同生成函数 $T_{audio} = f(\theta_{rec}, \Delta\theta, O_{type}, d, \phi)$ 决定，例如生成“左前方一米有行人，请右转”等提示。

触觉通道通过智能手杖手柄的振动电机阵列，振动模式 M_{vib} 由建议转向角 $\Delta\theta$ 映射：

$$M_{vib} = g(\Delta\theta) = \begin{cases} \text{持续轻微振动} & \text{if } |\Delta\theta| \leq 15^\circ \text{ (直行/微调)} \\ \text{左侧交替强振动} & \text{if } \Delta\theta > 15^\circ \text{ (提示右转)} \\ \text{右侧交替强振动} & \text{if } \Delta\theta < -15^\circ \text{ (提示左转)} \\ \text{急促双脉冲} & \text{if } \min(d) < 0.5 \text{ m (紧急停止)} \end{cases}$$

该双模态反馈策略借鉴了人因工程的“多通道互补”原则^[12]，利用语音传递丰富的语义信息，同时通过触觉提供即时、隐蔽的方向指引，两者互补确保了信息在复杂环境下的可靠传达，显著提升了交互的鲁棒性与用户的接受度。

第四章 实验与评估

为全面评估本系统的性能，设计了涵盖客观指标测试与主观用户体验调研的综合实验方案。

4.1 实验环境与设置

测试在三种具有代表性的场景中进行，以模拟视障人士日常出行的主要环境：

1. 室内结构化场景：大学教学楼的直线与L型走廊，包含墙壁、门窗、消防栓等。
2. 室内非结构化场景：摆有桌椅、盆栽、箱子的实验室环境，通道狭窄且存在不规则障碍。
3. 室外半结构化场景：校园人行道，包含行人、自行车、路灯杆、路缘石（台阶）以及树木阴影造成的光照变化。

在每条测试路径上，预先设置多种类型的障碍物，包括静态障碍（箱子、玻璃门、低台阶）和动态障碍（以正常速度行走的实验人员）。对比系统设置为：S-Vision（仅使用双目视觉）、S-Ultrasonic（仅使用超声波）、Proposed（本文提出的融合系统）。

4.2 客观性能评估结果与分析

评估主要关注检测成功率、误报率、漏报率和系统延迟四个核心指标，结果总结于表 1。

表 1 不同系统在各场景下的客观性能对比

场景	系统	检测成功率 (%)	误报率 (%)	漏报率 (%)	平均延迟 (ms)
室内结构化	S-Vision	96.2	1.5	2.3	175
	S-Ultrasonic	88.7	4.8	11.3	45
	Proposed	99.1	1.0	0.9	195
室内非结构化	S-Vision	89.5	2.1	10.5	182
	S-Ultrasonic	82.3	5.2	17.7	48
	Proposed	97.8	1.8	2.2	215

室外半结构化	S-Vision	85.1	1.8	14.9	185
	S-Ultrasonic	75.6	7.5	24.4	52
	Proposed	98.5	2.5	1.5	220
综合平均	S-Vision	90.3	1.8	9.7	180
	S-Ultrasonic	82.2	5.8	17.8	48
	Proposed	98.5	1.8	1.5	210

结果分析:融合系统在各类场景下均表现最优,综合检测率达 98.5%,显著优于纯视觉(85.1%)和纯超声波(75.6%)系统,有效克服了单一传感器在光照变化或远距/小目标感知上的局限。在可靠性方面,系统漏报率降至 1.5%(最低),误报率与纯视觉相当,而纯超声波因环境干扰误报最高。尽管融合引入额外计算,端到端延迟为 210 ms,仍远低于 300 ms 的实时性阈值,满足安全避障的及时性要求。

4.3 主观用户体验评估

邀请 10 名视力正常的志愿者(蒙眼进行模拟体验)和 2 名视障人士参与初步可用性测试。采用 5 分制李克特量表(1=非常不同意,5=非常同意)对系统进行评价,结果如表 2 所示。

表 2 主观用户体验评分(平均分)

评估维度	S-Vision	S-Ultrasonic	Proposed
我能信任该系统提供的障碍物信息	3.8	2.5	4.6
系统给出的提示(语音/振动)清晰易懂	4.0	3.0	4.8
使用该系统让我在行走时感到更安全	3.5	3.8	4.5
我愿意在日常生活中使用该系统	3.8	3.1	4.7

结果分析:融合系统在感知信心、环境理解、交互自然性与信任度上均优于单模态方案。用户认为其“能提示障碍物并识别类别”,信息更可靠;“骨传导语音+方向振动”交互自然、低干扰,触觉在嘈杂环境中有效补充听觉。纯超声波因无语义、误报多而信任度最低,纯视觉在弱光或透明障碍下鲁棒性差。结果与 Meshram 等^[2]和 Lv 等^[12]的发现一致,初步验证了系统实用性(n=2)。

第五章 结论

本文围绕视障人群独立出行的实际需求,构建并实现了一套融合双目视觉与超声波信息的实时辅助避障系统。系统以树莓派 4B 为核心平台,搭建可穿戴式硬件结构,通过改进的半全局立体匹配算法与多方向超声波阵列协同工作,形成覆盖远距与近距的环境感知机制。在此基础上,引入扩展卡尔曼滤波与占据栅格地图对多源数据进行统一处理,实现空间信息的综合融合,并借助骨传导耳机与振动手杖向用户提供直观的多模态提示。

实验结果表明,在室内外复杂场景条件下,系统对多类型障碍物的平均检测成功率达到 98.5%,平均预警延迟约为 210 ms,相比单一传感器方案在稳定性与可靠性方面均有明显提升,同时在实时性和使用体验上也表现出较好的综合性能。

在此基础上,后续研究将从多个方面进一步展开:首先,引入惯性测量单元(IMU)以构建视觉惯性里程计,提高系统在快速运动状态下的稳定性;其次,探索轻量化多任务神经网络模型,推动感知与决策的一体化优化;再次,通过更大规模的视障用户实地测试,进一步完善人因工程与交互设计;最后,推进系统的小型化与产品化,结合云端辅助与公共无障碍环境,提升实际应用价值。

参考文献:

- [1] 胡桐. 多功能智能导盲眼镜[J]. 农村青少年科学探究, 2021(22): 24.
- [2] MESHARAM V V, PATIL K, MESHARAM V A, et al. An astute assistive device for mobility and object recognition for visually impaired people[J]. IEEE Transactions on Human-Machine Systems, 2019, 49(5): 449-460.
- [3] WORLD HEALTH ORGANIZATION. World report on vision[R]. Geneva: WHO, 2023.
- [4] 刘隆辉. 多传感器多特征融合的移动机器人视觉 SLAM 系统研究[D]. 东莞: 东莞理工学院, 2024.

- [5] 高强, 陆科帆, 吉月辉, 等. 多传感器融合 SLAM 研究综述[J]. 现代雷达, 2024, 46(8): 29 - 39.
- [6] 王远新, 闫凯航, 邓炬鑫, 等. 基于 CNN 和激光雷达融合的智能语音避障导盲眼镜的研究与设计[J]. 电子制作, 2024, 32(12): 45 - 47+44.
- [7] GARCIA-MACIAS J A, RAMOS A G, HASIMOTO-BELTRAN R, et al. Uasisi: A modular and adaptable wearable system to assist the visually impaired[J]. Procedia Computer Science, 2019, 151: 425 - 430.
- [8] LV Z, LI J, LI H, et al. Blind travel prediction based on obstacle avoidance in indoor scene[J]. Wireless Communications and Mobile Computing, 2021, 2021: Article ID 5536386.

Intelligent Obstacle Avoidance System Based on Binocular Vision and Ultrasonic Fusion for Visually Impaired People

LI Jinhua, ZHONG Yunxuan, WANG Yanfei*

(Guangzhou Institute of Science and Technology, Guangzhou, Guangdong 510000, China)

Abstract: Aiming at the shortcomings of current visual impairment assistance devices in transparent obstacle recognition, low-lying object detection, and dynamic target perception, this paper designs and implements an obstacle avoidance system fusing binocular vision and ultrasonic information. The system adopts a master-slave embedded architecture with Raspberry Pi 4B and STM32 as the core. An improved semi-global matching algorithm is introduced at the vision end to support medium and long-distance semantic perception. Meanwhile, combined with a five-direction ultrasonic array, high-precision near-field detection within the range of 0.1–8.0 m is realized. Multi-source perception data are collaboratively fused in time and space through extended Kalman filter and occupancy grid map, and low-interference prompt information is output via bone conduction voice and vibration cane. Experimental results show that in complex indoor and outdoor environments, the average recognition success rate of the system for various types of obstacles reaches 98.5%, and the early warning delay is about 210 ms. The overall performance is superior to single-sensor schemes. The system does not rely on network and has controllable cost, providing a feasible implementation for universal accessible intelligent assistance equipment.

Keywords: Visually impaired assistance; Binocular vision; Ultrasonic sensing; Multi-modal fusion; Obstacle avoidance system; Accessibility technology