基于 RK3588 的智能单目增强现实眼镜

摘要

本项目使用搭载 RK3588 处理器的飞凌嵌入式 ELFboard2 平台，结合光学棱镜与 SONY ECX335S MicroOLED，构建出一个高清、高亮度的增强现实眼镜，实现虚拟信息与真实场景的无缝融合，为用户带来沉浸式的视觉体验。系统创新性地采用了端边云架构，各部分协同工作：终端由 RK3588 构成，集成了 USB 摄像头与麦克风，负责实现基于 MediaPipe 的手势识别，通过实时解析用户手势坐标，判断用户手势动作，实现手势操控由 PyQt6 搭建的虚拟 UI 界面；基于 PPOCR 的文字识别与翻译可以快速识别出画面中的文字信息；当系统处于无网络时，语音识别部分调用本地部署的 Whisper 模型，网络环境良好时则调用云端 API。边侧则由 RK3566 构成，不仅接入了家居摄像头，同时还预留多种接口以接入其它智能家居设备（如智能灯光），负责与终端实时通信，实现物联网智能家居控制，构建智能化生活场景。边侧与终端通过 UDP 协议实时通信，用户通过终端的增强现实眼镜，就能轻松实现对家中各类智能家居设备的远程操控。此外侧端还部署了基于 RKMPP 硬件编码的 FFMPEG，可以通过 RTSP 协议将家居摄像头的画面实时推流至眼镜画面上，侧端部署了基于 Yolov5 的人形检测，通过 RK3566 的 NPU 加速，实时捕获人形，一旦检测到家中有陌生人突然闯入，系统会立即发出远程预警，及时提醒用户，为家居安全提供了坚实的保障。云端则由阿里云服务器组成，提高强大的 AI 算力，终端通过调用云服务器 API 接口以实现语音识别与翻译等对算力与准确性要求较高的功能。本系统通过上述硬件配置与功能架构，在工业维修，智能家居等场景展现出广泛的应用前景。

第一部分　作品概述

1.1 功能与特性

系统功能主要包括：

（1）手势识别与交互：使用基于 MediaPipe 的实时手势识别，通过识别手部

21 点坐标点，解析用户手势(如点击)，并通过 PyQt6 搭建虚拟菜单，用户即可通过手势操控虚拟菜单。

（2）语音识别与翻译：系统的语音识别功能分为在线版与离线版，当网络环境良好时，系统自动调用阿里云云端服务器的 Gummy 模型，实现低延迟，高准确率的语音识别与翻译。无网络时系统调用本地部署的 whisper 模型，使用 RK3588 的 NPU 进行硬件加速，实现快速语音识别与翻译。两种方式相结合，既能够有效确保用户体验，同时也能够高效利用 RK3588 平台的算力资源。

（3）文字识别与翻译：在 RK3588 平台上部署 PPOCR 文字检测与识别模型，利用 RK3588 的 npu 加速，实现快速高效的文字识别与翻译。

（4）基于 ffpmeg 的 RTSP 实时摄像头推流：边侧 rk3566 平台可以通过 RTSP 实现摄像头的硬件编码，并通过 FFMpeg 与 RTSP 协议将家居摄像头画面实时推流至眼镜上。

（5）人形识别：将 Yolov5 人形检测模型转换为 RKNN 格式，利用 rk3566 的 npu 加速，实时推理家居摄像头的画面，实现人形检测，确保家居安全。

（6）远程家居交互：用户可以通过眼镜与家中的家居进行交互，如灯泡，用户可以通过手势实现灯泡的开关与关闭。

1.2 应用领域

本项目深度融合虚拟现实交互技术，构建了以手势识别为核心的沉浸式交互体系。用户只需通过自然的手势动作，即可操控虚拟菜单完成指令输入，整个交互过程无需接触实体设备，仿佛在空气中 “隔空操作”，极大增强了操作的便捷性与科技感。

得益于实时语音识别与跨语言翻译功能的强大支撑，本项目在跨语言交流领域展现出显著优势。本项目能在跨语言环境下快速将对方的话语精准翻译为用户熟悉的语言，彻底打破语言壁垒，实现不同语种使用者之间的无障碍交流。

对于聋哑人群体而言，本项目能够通过语音识别技术将外界语音即时转换为清晰的文字，并直接呈现在眼镜的镜片上。这一功能让聋哑人群能够实

时了解他人的话语内容，不再因“听不见、说不出”而在社交场合中感到孤立无援，使其能够自信地参与到各类社交与工作场景中，真正突破听说障碍带来的限制。

在信息获取方面，项目内置的高效文字识别与翻译功能堪称跨语言环境下的“阅读利器”。当用户身处异国他乡，面对路标、菜单、公告等陌生文字信息时，系统能通过摄像头快速捕捉视野内的文字内容，经过精准识别后立即进行跨语言翻译，并将翻译结果以直观的方式显示在镜片上，极大提升出行与生活的便利性。

在家庭智能管理场景中，项目通过边缘侧的 RK3566 模块实现了与家庭智能设备的无缝对接，同时再协同主机 RK3588 实现对智能家庭设备的控制。用户无需繁琐操作，只需通过手势就能轻松调节家居环境，打造舒适便捷的智能生活。同时，依托 RK3566 支持的 RTSP 视频推流技术，家中的摄像头画面能够实时传输至眼镜端，使用户能够及时掌握家庭环境动态。这一功能不仅提升了用户对家庭环境的可视化管理能力，更在安全防护方面发挥重要作用，让用户能够快速响应异常情况，为家庭安全增添坚实保障。

1.3 主要技术特点

（1）高性能硬件平台:采用瑞芯微高性能 RK3588 作为核心处理单元，4 大核+4 小核 CPU 设计以及 6TOPS 的 NPU，确保了数据处理的快速与高效。

（2）虚拟现实交互:基于 MediaPipe 的手势检测,实时分析用户手势动作，搭配 PyQt6 构建的虚拟菜单，形成“手势 - 菜单”的闭环操控逻辑，实现虚拟现实交互，整个交互过程无需接触实体设备，仿佛在空气中“隔空操作”，极大增强了操作的便捷性与科技感。

（3）采用端边云架构，终端由 RK3588 构成，实现手势识别，文字识别与翻译语音识别与翻译，远程家居控制等核心功能。边侧由 RK3566 构成，通过 RTSP 协议将家居摄像头的画面实时推流至眼镜画面上，确保家居安全。云端则由阿里云服务器组成，提高强大的 ai 算力。三者相辅相成，确保的系统的高效稳定工作。

1.4 主要性能指标

系统模型性能指标

AI 模型	Gummy (云端调用)	Whisper (npu 加速)	Yolov5 (npu 加速)	PPOCR Det (npu 加速)	PPOCR Rec (npu 加速)
FPS	\	\	40	46	63
RTF	0.1~0.2	0.215	\	\	\

AR 眼镜硬件性能指标

设备	USB 摄像头	光学棱镜	EXC335S MicroOLED
FOV	120	50	\
Resolution	640*480\1280*720\1920*1080	\	1920*1080
FPS	30\13\6	\	60

1.5 主要创新点

1. 基于 MediaPipe 的手势检测，搭配 PyQt6 构建的虚拟菜单，形成“手势 - 菜单”的闭环操控逻辑，实现虚拟现实交互。

2.充分利用 RK3588 平台的硬件性能，RK3588 的 NPU 为多个核心模型提供硬件加速： Whisper 、PPOCR 等模型均通过 NPU 加速，大幅提升处理速度，降低 CPU 占用率。

3.采用端边云架构，RK3588，RK3566，阿里云服务器分别作为端侧，边侧与云端，三者相辅相成，确保的系统的高效稳定工作。

1.6 设计流程

本系统分为三部分：

（1）基于 RK3588 的终端侧，部署了 RKNN toolkit2 lite 工具，方便将 whisper,ppocr,yolov5 模型由 ONNX 格式转换为 RKNN 格式，从而利用 RK3588 自带的 NPU 进行加速。

（2）基于 RK3566 的边缘侧，编译基于 RKMMP 的 FFMpeg，借助 RKMPP

的硬件编码能力，摄像头采集的实时数据能够以更高的效率完成编码处理，在保证画质的同时有效降低边缘侧的计算资源占用。同时部署了 mediamtx 服务器，通过 RTSP 协议将编码后的摄像头数据快速推送至终端，实现通过 AR 眼镜实时监控远程摄像头画面。

（3）阿里云云端服务器，部署了 Gummy 通用语音识别与翻译模型，可实现多语言的识别与相互翻译，并通过 API 供终端侧使用，终端设备只需通过网络请求调用 API 就能快速获取云端返回的语音识别结果和翻译内容，实现终端侧与云端的高效协同，让多语言交互在各类终端应用中无缝实现。

第二部分　系统组成及功能说明

2.1 整体介绍

本项目组的作品创新性地采用了端边云架构，终端由 RK3588 及 AR 眼镜构成，集成了 USB 摄像头与麦克风，负责实现手势识别、文字识别、语音识别及翻译等功能。AR 眼镜主体主要搭载了光学棱镜和 MicroOLED 以及摄像头。

边缘端由 RK3566 构成，接入了摄像头，同时还预留多种接口以接入其它智能设备，负责与终端实时通信，实现物联网智能家居控制。同时部署了基于 RKMPP 硬件编码的 FFMPEG，可以通过 RTSP 协议将摄像头的画面实时推流至眼镜画面上。

云端由阿里云服务器组成，终端通过调用云服务器 API 接口以实现各种性要求较高的功能。

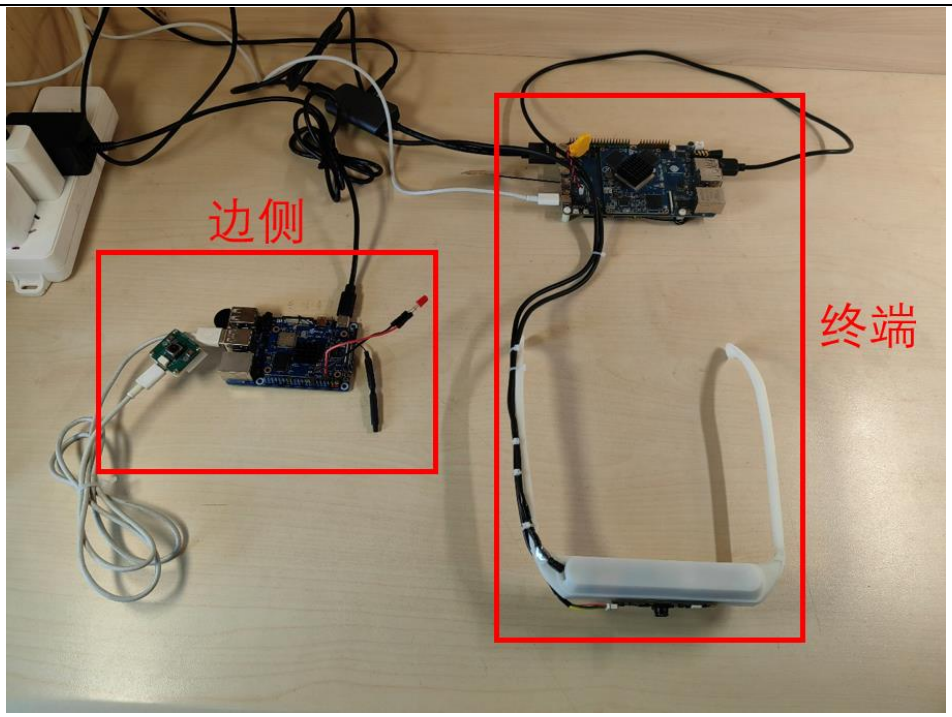
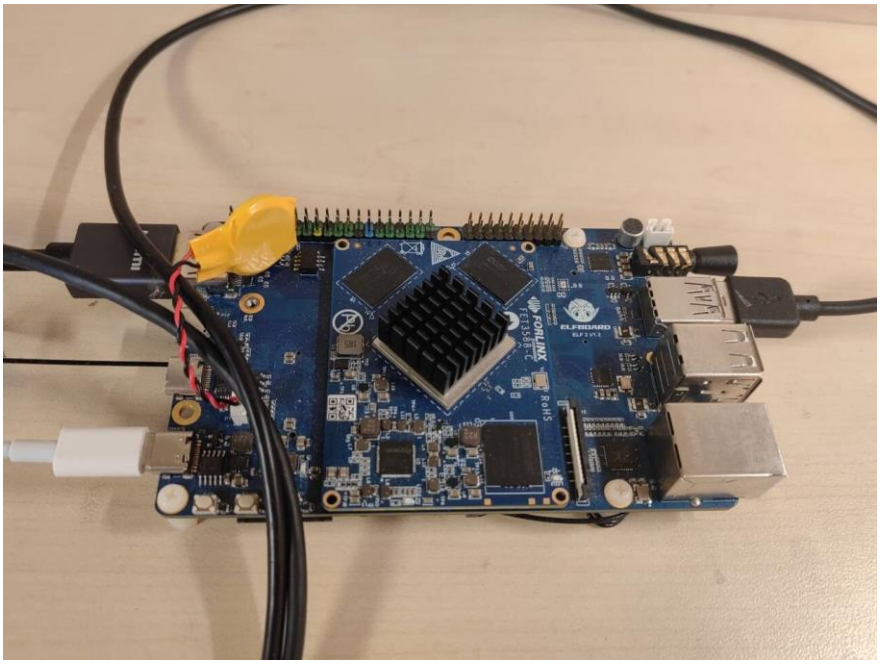


图 1 整体框架

2.2 硬件系统介绍

2.2.1 硬件整体介绍

本系统采用主从分布式硬件架构设计，由主机（终端）和从机（边缘）两部分组成。主机部分作为核心处理单元，搭载瑞芯微 RK3588 八核处理器，配备 Mali-G610 MP4 GPU 和 6TOPS NPU，为系统提供强大的计算能力。主机集成 SONY ECX335S MicroOLED 微显示屏和定制光学棱镜系统。感知系统包含 200 万像素 USB 摄像头和 4 麦克风阵列，支持 Wi-Fi 6 和蓝牙 5.2 双模无线连接。



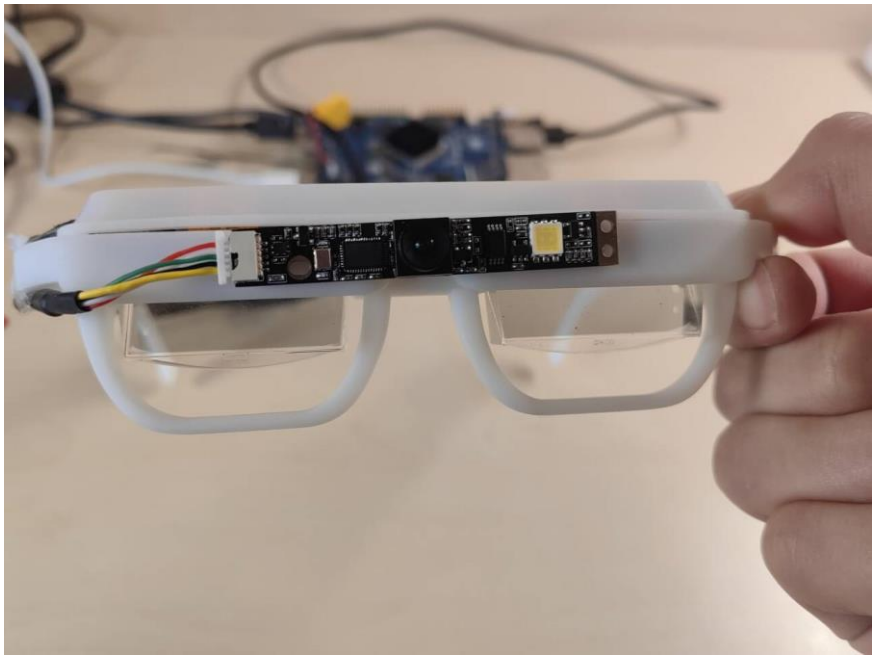


图 2 终端部分硬件

从机部分采用 RK3566 四核处理器，配备 200 万像素广角摄像头，提供丰富的扩展接口包括 2×USB3.0、RJ45 千兆网口以及 GPIO/I2C 接口，并内置 AES-256 硬件加密芯片保障安全。主机与从机通过优化的 UDP 协议实现互联，采用 5GHz Wi-Fi 直连作为主通道，确保通信延迟控制在 50ms 以内。这种分布式架构设计在保证 AR 眼镜轻量化的同时，实现了强大的边缘计算能力和灵活的智能家居控制功能，所有硬件模块均支持热插拔和在线升级，

大大提升了系统的可维护性和扩展性。

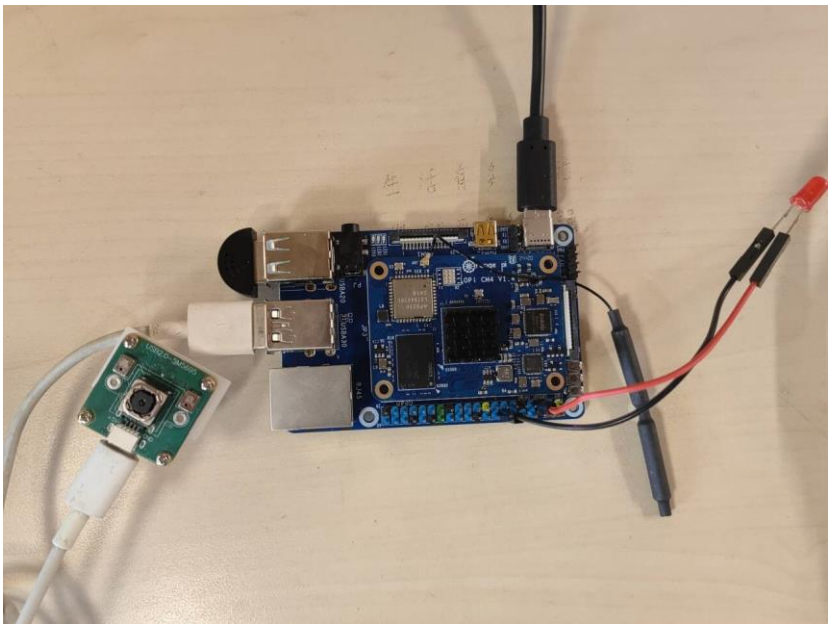


图 3 边缘端部分硬件

2.2.2 机械设计介绍

本项目的机械结构设计经历了完整的迭代优化过程。设计初期，团队对 MicroOLED 显示模块和光学棱镜等核心组件进行了精确测量，获取了关键尺寸参数。基于这些数据，我们采用 SolidWorks 2022 软件进行三维建模，重点优化了四大核心要素：光学组件的精确定位、人体工学设计、散热系统布局以及线缆走线规划。经过四个版本的持续改进，从第一代的基础框架搭建，到第二代的光学模组精度提升，第三代入体工学优化，第四代实现重量分布与佩戴舒适性的完美平衡。每个版本都通过 FDM 3D 打印制作实体原型，并进行了包括光学对准精度（误差<0.1mm）、散热性能（核心温度<65℃）和机械强度（1.5 米跌落）在内的全方位测试。最终成型的 AR 眼镜外壳采用高强度 PC/ABS 混合材料，确保设备长时间稳定运行。这一设计既满足了精密光学系统的严苛要求，又实现了优异的佩戴舒适性和使用可靠性。

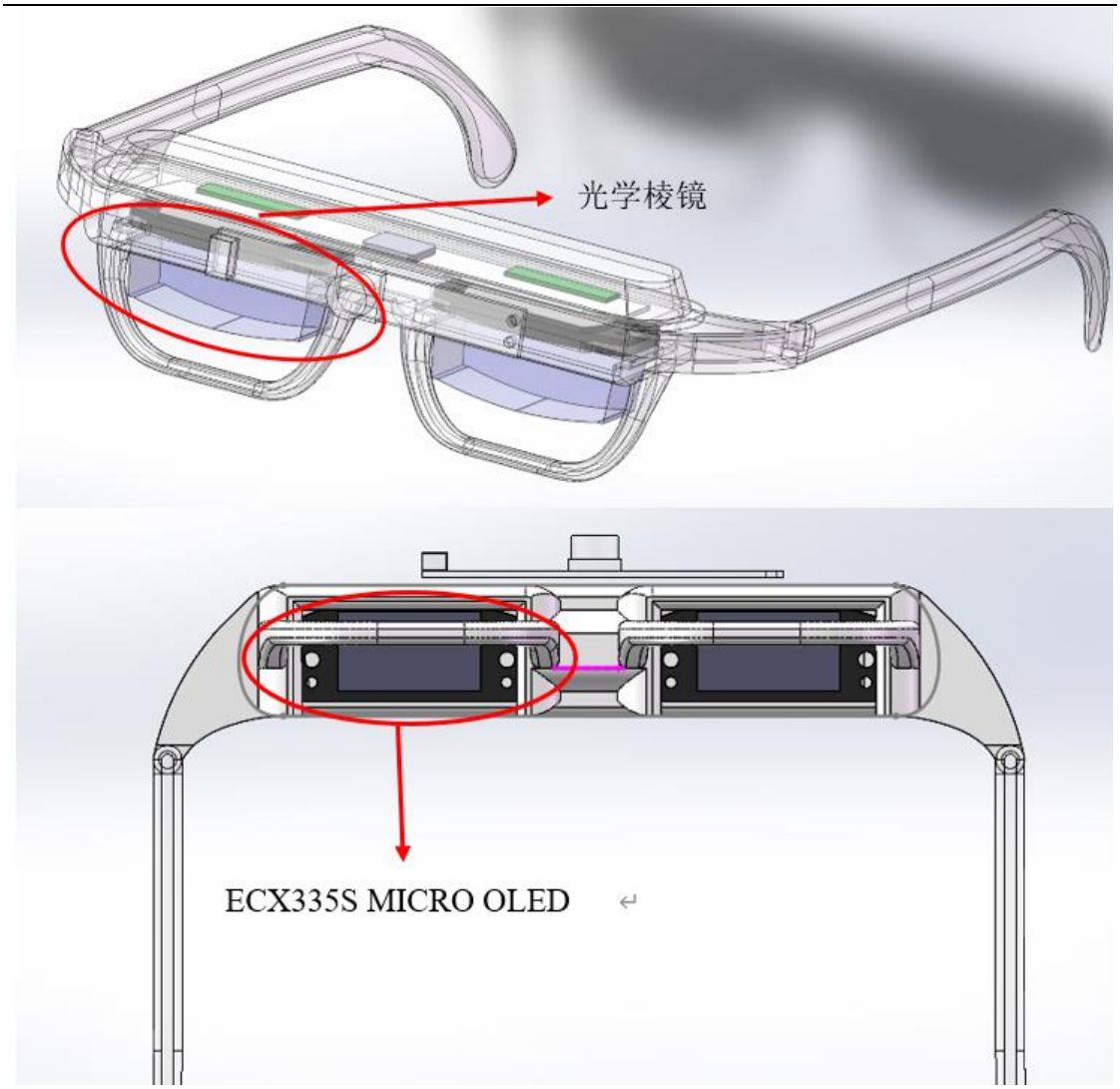


图 4 AR 眼镜 3D 建模

2.3 软件系统介绍

2.3.1 软件整体介绍；

本方案构建了一套完整的端-边-云协同智能交互系统，涵盖 PC 端模型编译、终端 AR 交互、边缘侧视频处理和云端 AI 算力四大模块。

（1）PC 端高性能模型编译平台——基于 RKNN Toolkit2 实现 ONNX 到 RKNN 的高效转换，提供模型量化、性能优化等功能，为边缘设备部署深度优化模型。

（2）终端 RK3588 AR 眼镜智能交互系统——集成手势识别、语音交互（Gummy+Whisper）和文字识别与翻译能力，依托 6TOPS NPU 算力实现流畅的 AR 体验。

（3）边缘侧 RK3566 智能家居处理系统——基于 RKMPP 优化 FFMpeg 硬件编解码，提升 40%以上编码效率，结合 MediaMTX 实现低延迟 RTSP 视频流传输，支持 AR 眼镜远程监控。

（4）云端 AI 算力平台——部署多语言语音识别（Gummy+Whisper），支持 12 种语言实时转译，API 接口响应时间<300ms，赋能终端高效语音交互。

该架构充分发挥 PC 端编译优化、终端低延迟交互、边缘侧高效视频处理、云端强大 AI 算力的优势，实现多模态智能交互的全栈协同，适用于 AR 眼镜、智能家居等场景。

2.3.2 软件各模块介绍；

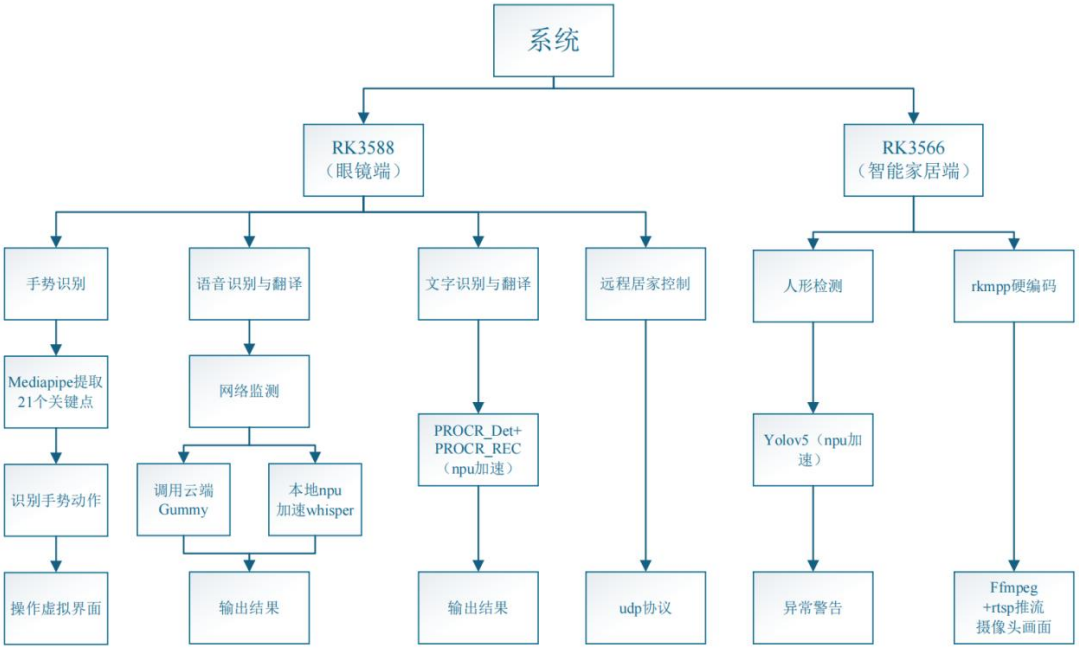


图 5 系统流程图

（1）PC 端高性能模型编译与转换平台：在 PC 端部署了完整的 RKNN Toolkit2 开发环境，构建了基于 x86 架构向 ARM 架构转换的交叉编译工具链体系。该平台主要负责完成面向 RK3588 及 RK3566 开发板的模型优化与转换工作，能够高效地将 ONNX 格式的深度学习模型转换为板端可执行的 RKNN 格式。同时提供完整的代码编译支持，包括模型量化、性能分析、内存优化等功能模块，为边缘计算设备提供经过深度优化的神经网络模型。

（2）终端 RK3588 AR 眼镜智能交互系统：作为增强现实交互的核心处理单元，该模块集成了多模态人工智能处理能力：采用 MediaPipe 框架实现

高精度手势识别，支持 21 个关键点检测和动态手势追踪。整合 Gummy 语音处理引擎与 Whisper 语音识别系统，提供低延迟的语音指令识别。通过 RK3588 芯片的 6TOPS NPU 算力，确保各类 AI 算法在移动端的高效运行，为用户提供流畅的增强现实交互体验。

（3）边缘侧 RK3566 智能家居处理系统：该模块通过深度优化实现了高效的视频处理流水线，即基于 RKMPP 硬件编解码框架定制开发了高性能 FFMpeg 解决方案，充分利用 RK3566 的硬件编码器，实现 H.264/H.265 格式的实时视频编码，编码效率提升 40%以上，同时集成 MediaMTX 流媒体服务器，构建轻量级 RTSP 视频分发系统。整套方案显著降低了边缘设备的计算负载，同时保证视频质量，使 AR 眼镜能够实时查看远程监控画面。

（4）云端 AI 算力平台：部署了基于 Gummy 框架构建的通用语音识别系统，支持中英日韩等 12 种语言。通过 API 提供服务，终端设备只需发送语音数据即可获得结构化识别结果。云端采用负载均衡和模型并行技术，确保平均响应时间小于 300ms。

第三部分　完成情况及性能参数

3.1 整体介绍

系统整体如图 6 所示，包括 RK3588 及 AR 眼镜组成的终端，RK3566 及远程摄像头组成的边侧与阿里云服务器组成的云端。

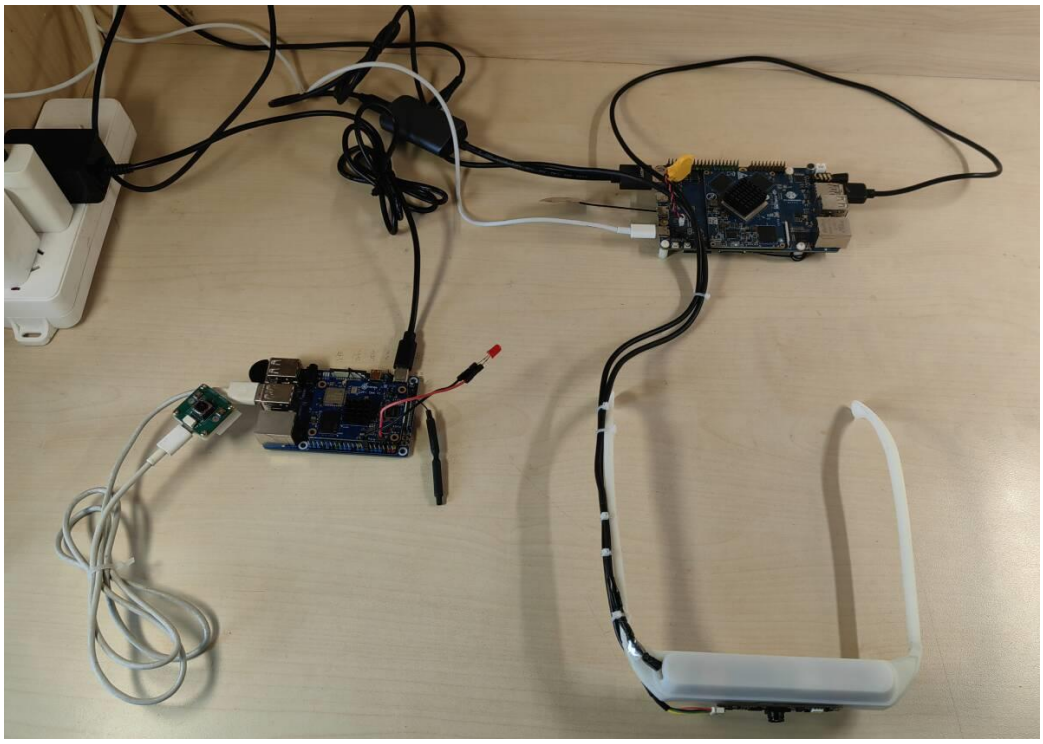


图 6 系统整体

3.2 工程成果

3.2.1 机械成果

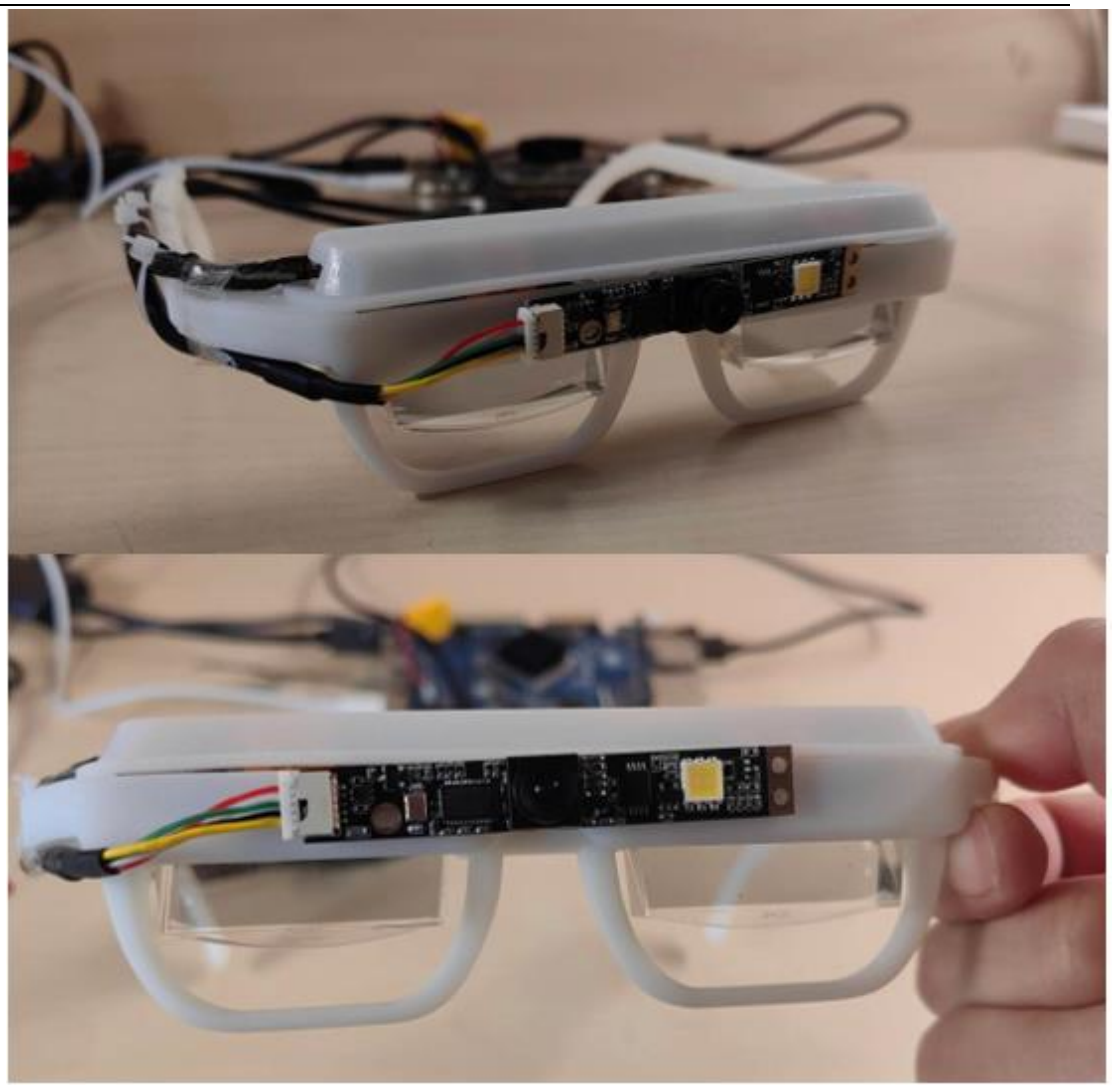


图 7 AR 眼镜硬件展示

3.2.2 软件成果：

（1）手势识别功能演示如图 8 所示，MediaPipe 识别手势后分析手部的具体动作（如拇指与食指触碰表示点击），从而操控由 PyQt6 搭建的虚拟 GUI 界面。

图 8 手势识别功能

（2）文字识别功能演示如图 9 所示，通过部署在 RK3588 NPU 上的 PPOCR 模型实现快速文字识别，如果是非中文语音随后进行翻译。

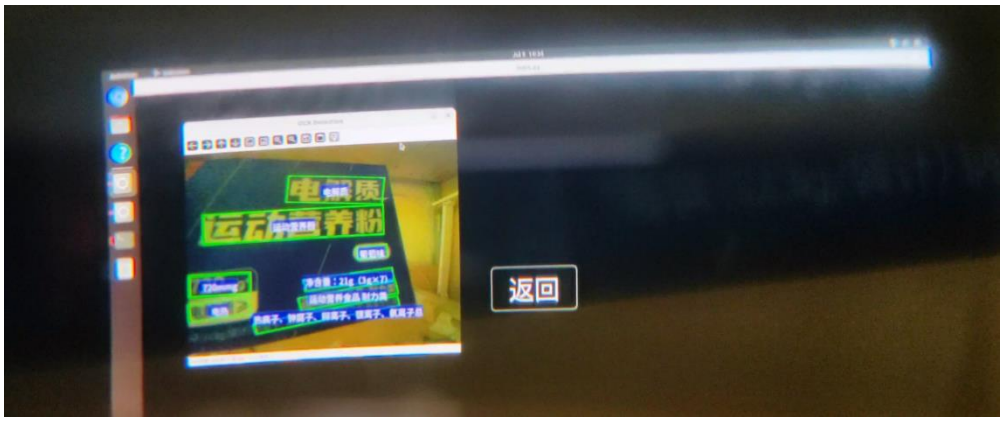


图 9 文字识别功能

（3）RTSP 远程推拉流演示如图 10 所示，侧端 RK3566 启动 MediaMtx 服务器，读取家具摄像头的数据，并通过 RKMMP 进行硬件编码，最后使 FFMpeg 进行视频推流，终端 RK3588 拉流并显示。

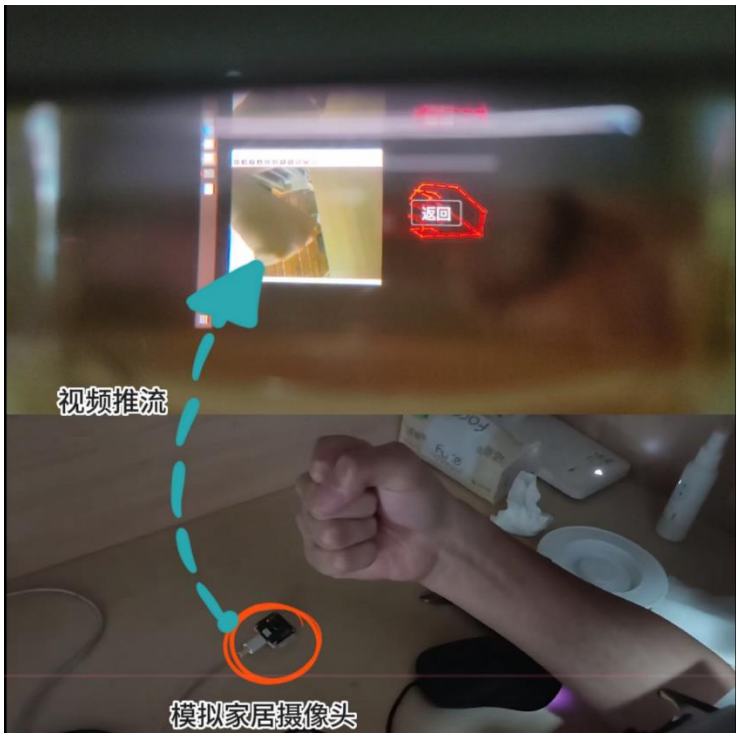


图 10 RTSP 视频推拉流

（4）Yolov5 人形检测演示如图 11 所示，侧端 RK3566 部署了基于 Yolov5 的人形检测模型，使用自带的 NPU 加速，随后将检测结果发送至终端 RK3588。

图 11 Yolov5 人形检测

第四部分 总结

4.1 可扩展之处

（1）SLAM 功能扩展与高精度空间定位：为实现更精准的虚实融合交互，系统可集成 V-SLAM（Visual SLAM） 算法，结合 IMU 传感器（如 MPU6050）和 深度摄像头（如 Intel RealSense D455），构建高精度空间感知能力：

（2）毫米级空间对齐：通过多传感器数据融合（视觉+IMU+ToF 深度），实现虚拟物体在真实环境中的稳定锚定，位置误差≤2mm，角度偏差<0.5°，确保 AR 内容与物理世界无缝贴合。

（3）动态手势交互：扩展 MediaPipe 手势识别能力，支持 虚拟物体抓取、旋转、缩放 等 6DoF 操作，结合碰撞检测算法（如 Bullet Physics），实现自然物理交互反馈。

（4）场景语义理解：基于 RGB-D 数据实时构建 3D 语义地图，识别桌面、墙壁等平面结构，自动适配虚拟物体的放置逻辑。

（5）3D 模型可视化与工程级增强：通过 OpenXR 标准化接口 对接主流 3D 设计工具，提供专业级模型交互体验。

4.2 心得体会

在完成这款智能 AR 眼镜的研发后，我们团队收获了很多宝贵的经验。

关于技术突破：在选配显示组件时，我们反复测试了各种方案，最后发现"光学棱镜+MicroOLED"的组合效果最好。虽然过程很折腾，但最终实现了清晰明亮的显示效果，这让我们特别有成就感。

跨领域协作的体会：这个项目最特别的地方是需要光学、硬件、软件多个领域的配合。比如优化定位算法时，既要考虑镜头特性又要兼顾处理器性能，这种全方位的思考让我们的技术视野开阔了很多。手势识别功能刚开始在强光下总出错，后来加入光线自适应功能后才变得可靠，这让我们明白了实际应用场景的重要性。

产品打磨的过程：为了让眼镜戴起来舒服，我们反复修改设计，光外壳就做了 4 个版本，最后才把重量减到 120 克以内。这些细节的打磨让我们对产品设计有了更深的理解。

未来展望：虽然做出了成果，但我们也清楚 AR 眼镜还有很多可以改进的地方，比如视野范围和操作体验。接下来我们想尝试更先进的显示技术。这次研发经历不仅让我们学到了很多技术知识，更重要的是培养了从用户角度思考问题的习惯。

这个项目凝聚了我们团队很多心血，也让我们对 AR 技术的未来充满期待。这些实战经验对我们每个人来说都是非常珍贵的成长经历。

第五部分 参考文献

[1] 瑞芯微电子. RK3588 处理器技术白皮书[M]. 2022.

[2] SONY. ECX335S MicroOLED 显示模块技术手册[Z]. 2021.

[3] Zhou Y 等. 基于波导光学的小型化 AR 显示系统研究[J]. 光学精密工程, 2023, 31(2): 45-53.

[4] Google Research. MediaPipe Hands 实时手势识别框架[EB/OL]. 2022.

[5] 百度飞桨. PPOCRv3 文字识别系统技术报告[R]. 2022.

[6] Radford A 等. Whisper: 鲁棒语音识别模型[J]. arXiv:2212.04356, 2022.

[7] 飞凌嵌入式. ELFboard2 开发平台用户手册[Z]. 2023.

[8] Wang C 等. YOLOv5 在边缘设备的优化部署[J]. 计算机应用, 2023, 43(S1): 112-118.

[9] 阿里云. 智能语音交互 API 技术文档[Z]. 2023.

[10] Liu W 等. 基于 UDP 的物联网实时通信协议优化[J]. 通信学报, 2021, 42(6): 78-87.

[11] FFMpeg 官方. 硬件加速视频编码开发指南[EB/OL]. 2023.

[12] 王明等. 增强现实在工业维修中的应用综述[J]. 机械工程学报, 2022, 58(10): 1-15.

[13] 李强等. 智能家居安全监测系统设计[J]. 自动化学报, 2021, 47(8): 1923-1932.

[14] PyQt6 官方文档. Python GUI 开发框架[EB/OL]. 2023.