**专利技术交底书编写模板**

发明名称： 基于AR眼镜的机器狗多模态融合控制系统

申请人：柳敏德

发明人：

联系人：技术交底书执笔人： 技术问题联系人及联系方式：

**该专利技术所属技术领域：增强现实（AR）人机交互、多模态融合控制、机器狗控制、语音识别、手势识别。术语解释：**

**1.详细介绍技术背景,并描述已有的与本专利技术最相近似/最接近的现有技术, 以及客观指出现有技术的不足/缺陷**

（说明：1）复杂的技术，除了介绍最直接的背景，也可以介绍技术大背景，便于审查员理解本专利的技术背景；2）介绍最接近的现有技术，有出处或引用最好，审查员可以去调取查看；3）现有技术的缺点可能很多，主要举本专利能解决的不足/缺陷；4）如果现有技术和本专利技术方案看起来比较近似，更需要好好介绍它，以把区别点明示出来，免得审查员看不出来区别认为本专利无创造性）：

1.1 技术大背景​​

​​增强现实（AR）技术​​ 是近年来快速发展的人机交互技术领域的重要分支。其核心在于通过​​计算机实时计算​​和​​多传感器融合​​，将计算机生成的虚拟信息（如文字、图像、三维模型、空间标注等）​​精准、实时地叠加并融合​​到用户感知的真实世界环境中，实现对人的视觉、听觉、触觉（可能）等感官的增强，从而提供超越纯物理现实的交互体验。AR技术在工业维检、远程协作、教育培训、智慧导航、娱乐互动等诸多领域展现出巨大的应用潜力。实现高沉浸感、低延迟、强交互性的AR体验，关键在于高精度的​​环境感知（SLAM技术等）​​、​​虚实融合的空间注册技术​​ 以及自然高效的​​多模态人机交互接口​​。

同时，​​移动机器人技术​​，尤其是​​四足机器人（机器狗）​​，凭借其在复杂非结构化地形中卓越的运动能力（如爬楼梯、跨越障碍、适应崎岖地面），已成为执行巡检、侦查、运输、协作操作等任务的重要载体。为实现有效控制，机器狗需要强大的​​环境感知​​（激光雷达、深度摄像头、IMU等）、​​自主决策规划​​（定位、建图、路径/轨迹规划）、以及​​鲁棒的执行控制系统​​。​​如何让人能更直观、高效、安全地指挥和控制这些复杂机器人系统，尤其在动态、多变的真实场景中，是人机交互与控制领域持续面临的关键挑战。​​

​​1.2 最接近的现有技术​​

在机器狗控制领域，现有主流的技术方案主要包括：

​​1.物理遥控器/掌机控制：​​

​​具体描述：​​ 例如，云深处绝影Lite3机器狗通常配备专用的物理手柄或与智能设备（手机/平板）通过APP配合形成“掌机”。操作者需​​双手握持物理控制器​​，通过按动摇杆、按钮或触摸屏来向机器狗发送预先定义好的控制指令（如前进、后退、转向、起立、趴下等）。这种控制通常基于​​UDP/IP协议​​进行无线通信，指令编码相对简单、直接（如发送特定代码代表特定动作）。

​​出处/代表性：​​ 这是目前绝大多数消费级和工业级机器人（包括无人机、轮式机器人）​​最基本、最普遍的控制方式​​。相关技术广泛应用于如DJI无人机控制器、各种机器人遥控器以及云深处提供的官方控制器中（例如，绝影系列自带的遥控器）。

​​2.搭载AI主机实现语音交互控制：​​

​​具体描述：​​ 通过在机器狗本体（如背部扩展接口）上​​额外搭载一个计算单元（如小型AI主机或高算力嵌入式平台）​​，集成语音识别模块（如麦克风阵列）。操作者​​通过语音口令​​（如“前进”、“左转”、“趴下”）来控制机器狗。系统识别语音命令后，将其转换为控制指令发送给机器狗的运动核心控制器执行。

​​出处/代表性：​​ 这种方法借鉴了智能音箱（如Amazon Echo, Google Home）和语音助手（如Siri, Alexa）的交互模式。​​越来越多的智能设备尝试集成语音控制作为辅助交互手段​​。在移动机器人领域，部分研究型平台和高端产品（如Boston Dynamics的Spot在某些配置下）也已支持语音指令输入。

​​3.基于环境感知的自主导航：​​

​​具体描述：​​ 机器狗利用​​自身搭载的多传感器系统​​（激光雷达、深度相机、视觉摄像头、IMU等），在未知或部分已知的环境中创建地图（或更新地图）、确定自身位置（定位）、并基于给定的目标点，​​自主规划全局路径和局部实时避障轨迹​​，最终到达目标点。整个过程无需操作者实时操纵移动细节。

​​核心技术：​​ 该方案依赖于​​同步定位与建图（SLAM）​​、​​路径规划算法（如A\*, RRT）​​、​​轨迹优化与控制（如MPC）​\*\*​。

​​出处/代表性：​​ 这是高级移动机器人（如云深处绝影系列、Unitree的Go1/Go2/B2、Boston Dynamics Spot、ANYmal）的核心能力之一。相关技术在​​机器人操作系统（ROS/ROS2）​​中有大量成熟的导航功能包（如move\_base）和应用实例作为支撑。

​​1.3 现有技术的缺点（聚焦与本专利解决方案相关的）​​

尽管上述现有技术各有其适用场景，但​​从提升复杂场景下人-机器狗交互的自然性、效率和鲁棒性角度来看，存在以下显著不足与本专利旨在解决的核心问题：​​

​​1.交互方式单一，缺乏多模态融合与灵活性：​​

​​物理遥控器​​依赖纯手动操作，​​仅提供单一维度的控制输入​​（物理按键/摇杆）。它完全​​无法满足​​在诸如抢险救灾、协同作业、复杂操作等需要​​复合指令​​的场景下的人机交互需求。例如，操作者无法同时方便地用手势指向目标位置再用语音说明任务、或在操控行走姿态的同时用语音调整机械臂。

​​语音控制​​虽然提供了免提交互，但​​无法在需要精确空间定位（如指着一个具体位置说“去那里”）、或需要视觉辅助确认/引导的情况下单独有效地完成任务​​。手势控制的缺失也限制了交互的自然度。​​现有技术缺乏将手势、语音、视觉标注等多模态输入灵活组合应用的能力。​​

​​2.环境适应性差，鲁棒性不足：​​

​​语音控制​​的核心问题在于其在​​噪声环境下（如工厂、户外、灾害现场）的识别准确度会急剧下降​​，导致指令误识别或失效。此外，​​它难以区分“有效指令源”​​：当环境中有多人说话或存在背景广播指令时，机器狗​​无法可靠识别并响应其唯一指定操作者（“主人”）的命令​​，存在误执行或拒绝执行正确指令的风险，造成混淆甚至安全隐患。

​​3.人机交互体验受限，操作负担高：​​

​​物理遥控器控制的最大局限性在于强制要求操作者持续用双手握持设备​​，极大地限制了操作者在控制过程中同时执行其他任务（如手持工具、操作其他设备、进行手势引导）的能力。用户体验被局限在狭小的屏幕或按键上，​​缺乏身临其境的直观感与空间感​​。

​​缺乏对现实世界的直接可视化交互。​​ 现有方法很少能让用户在物理环境中直接看到“规划路径”、“任务标记”或“避障区域”等叠加在真实世界上的虚拟信息，人-机器狗-环境的交互存在信息割裂。

**2.针对上述现有技术的缺点，说明本专利技术的目的，指出本专利技术要解决的技术问题。** ​​ 针对以上现有技术在交互模态单一、环境鲁棒性差、操作负担重及体验直观性不足等方面的核心缺点，本专利提出了一种基于高性能AR空间计算平台的机器狗多模态融合控制方法。​​

​​本专利的技术目的在于解决以下技术问题：​​

​​ 如何实现自然高效的多模态融合控制？​​ 解决现有单一模态控制方式（物理遥控或纯语音）无法满足复杂场景下用户灵活组合使用语音、手势甚至视觉环境标记对机器狗下达复合指令的需求。

​​ 如何提升在复杂声学环境下的指令识别准确性和可靠性？​​ 显著克服传统语音控制在噪音环境下的失效问题，并精准区分指令来源，确保机器狗仅响应并执行其指定操作者的命令，避免误操作。

​​ 如何提供解放双手、直观沉浸的控制体验？​​ 消除对物理遥控器双手握持的依赖，允许操作者解放双手进行其他任务，并利用AR技术将虚拟控制界面和机器狗相关关键信息（指令反馈、导航路径、任务目标点等）直接、精准地叠加到用户视野中的真实世界相应位置，大幅提升人机交互的空间感知力、直观性和操作效率。

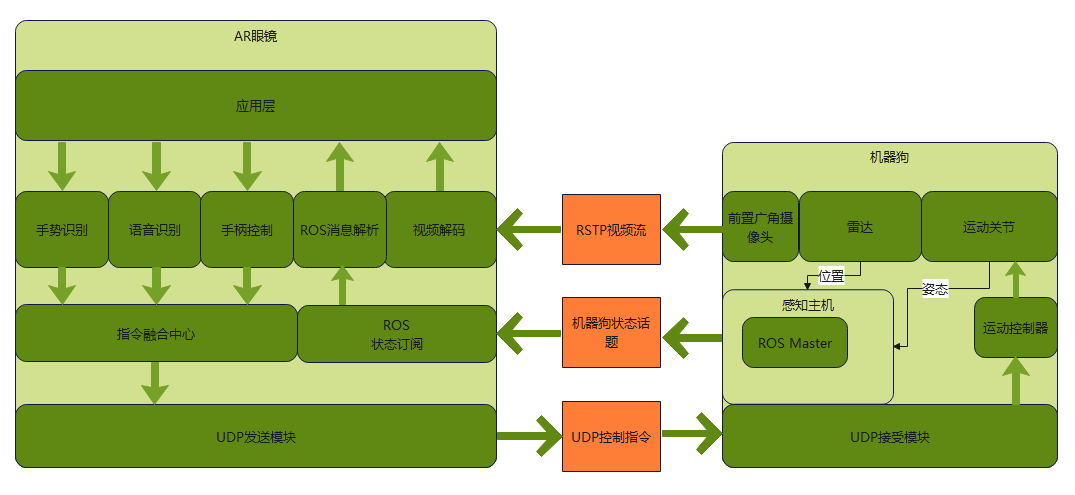
通过实现这些目的，本专利技术方案旨在为人与四足机器狗在复杂现实场景（如工业巡检、安防巡逻、抢险救援、影视制作）中的协同工作，提供一种更自然、鲁棒、高效且直观的新型交互范式。

**3.对本专利技术方案进行详细阐述，并结合结构图、流程图、原理框图、电路图、时序图等，进行详细说明。**

（说明：1）这是本专利技术内容的最关键部分，要求清楚、详细、完整、以及技术的可实现/可重现性。要注意条理、逻辑。2）为了讲清楚，可实现，要多进行一些举例。）

（这部分内容很重要，要达到 让除了你之外，其他技术人员——其实主要是专利代理师和审查员能看懂你的技术 的标准，撰写时，要注意：1）一定 要结合附图，详细介绍本发明技术是怎么做、如何实现的，包括工作流程、工作原理等，要把条理、流程、逻辑写清楚；2）附图，不能仅是“丢”了上来，而是要在这个板块里对附图也做详细介绍，因为附图是辅助技术文件，需要解释或翻译；3）为了让除了你本人以外的其他人员如专利代理师、审查员能更容易看懂你的专利技术，请尽可能举一些具体实例进行说明）

本系统采用​**​双端分层架构​**​，实现AR眼镜与机器狗的高效协同，工作原理图如下：



**一、AR眼镜端分层架构（左半部分）**

**1.应用层（顶层）**

**功能​**​：用户交互入口，显示机器狗图传视频流等

**2.感知处理层**

| 模块 | 功能描述 | 技术实现 |
| --- | --- | --- |
| 手势识别 | 解析深度摄像头数据，识别手势动作 | 3D卷积神经网络(CNN) |
| 语音识别 | 离线语音指令解析 | 端侧LSTM语音模型 |
| 手柄控制 | 物理按键指令解析 | Unity输入系统 |
| ROS消息解析 | 转换机器狗状态数据 | ROS#协议解析库 |
| 视频解码 | RTSP流媒体实时解码 | FFmpeg硬件加速 |

##### 3.决策融合层

功能：

1.执行多源指令优先级仲裁，例如优先执行急停指令，防止机器狗意外动作。

2.多条指令融合，例如手柄前进和语音前进，发送速度调整指令和前进指令，达到加速前进的目的。手势后退和语音前进相互抵消，不发送指令等。

##### 4.通信层（UDP发送模块）

协议：小端序UDP二进制协议

格式：[指令码4字节][参数4字节][类型4字节]

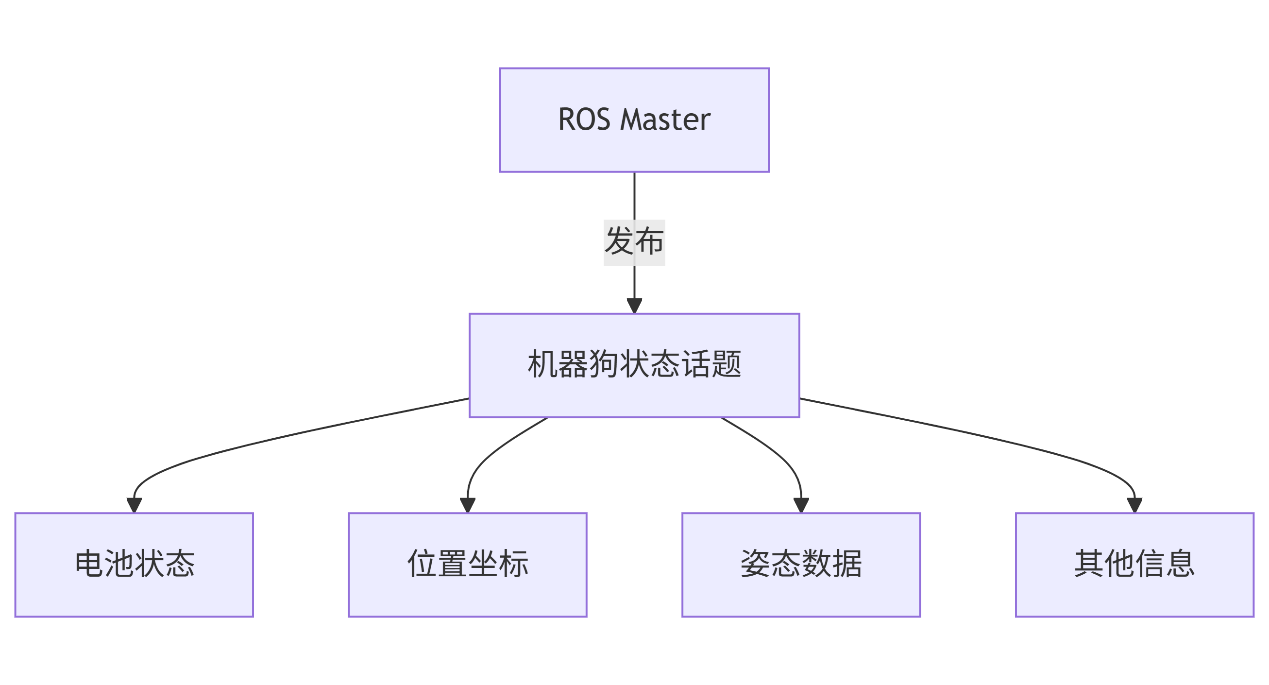
频率：25Hz高速发送（40ms间隔）

#### 二、机器狗端分层架构（右半部分）

##### 1. 物理层

| 组件 | 数据输出 | 技术规格 |
| --- | --- | --- |
| 前置广角摄像头 | 1080P@30fps视频流 | 120° FOV镜头 |
| 雷达 | 3D点云数据 | 10米探测范围 |
| 运动关节 | 实时位姿信息 | 6轴IMU传感器 |

##### 2. 感知主机（核心处理层）



##### 3. 运动控制器

* ​**​功能​**​：
  + 轨迹规划和步态控制
  + 传感器数据融合处理
  + 安全监控（碰撞预测/急停响应）

##### 4. 通信层（UDP接收模块）

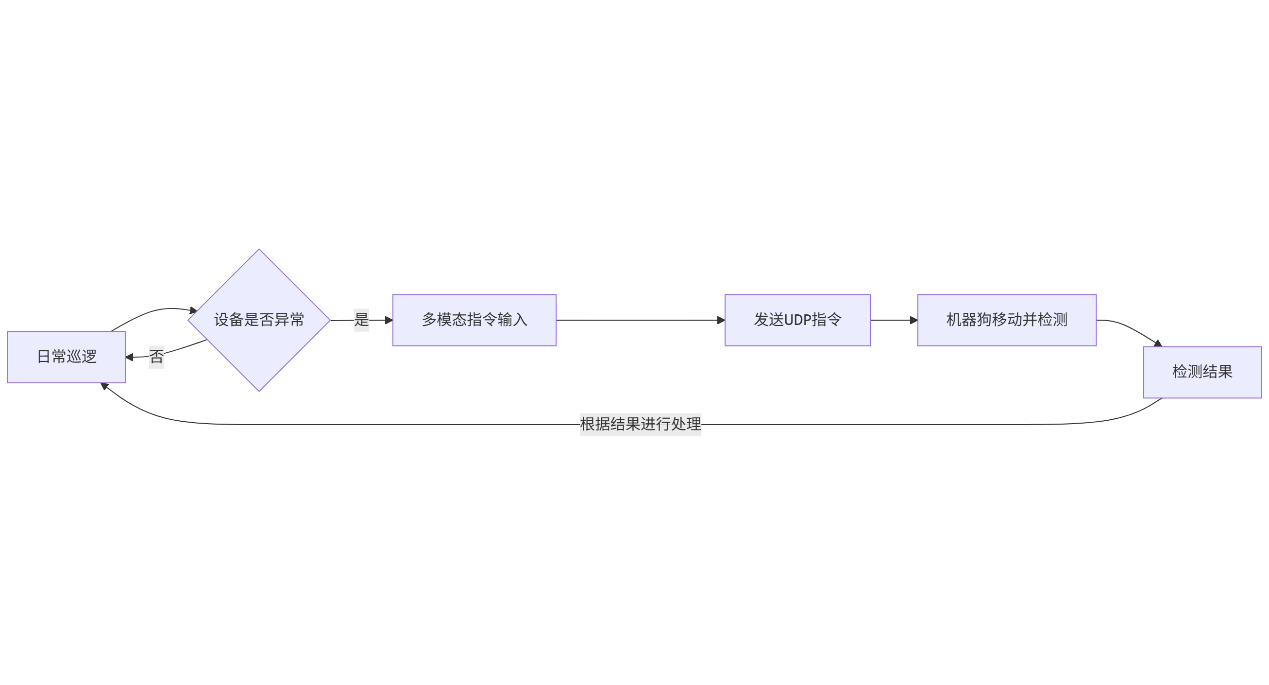
指令处理流程：

接收字节流 → 小端序解析 → 指令校验 → 指令分发 → 执行反馈

### 案例：危险环境，使用AR眼镜引导机器狗巡检设备

​**​场景描述​**​：

操作员在高危（例如高温，辐射等）环境中，通过AR眼镜操控机器狗巡检设备。发现某台设备可能存在异常，需要机器狗靠近检查。操作员使用手柄控制机器狗移动到设备前，然后发出语音指令"检查该设备"，机器狗接收到指令后执行对应和通过自身传感器执行检测任务，并将检测结果实时回传到AR眼镜。如果遇到常见的故障（如开关关闭等），使用手势控制机器狗按下开关。



**4.本专利技术的技术关键点和创新点是什么？**

**5.与现有技术相比，本专利技术的优点/技术效果是什么？**

（说明：1）要尽量和现有技术的不足对应；2）要有理有据，技术效果要尽量和创新点对应，来自于创新点）

**6.是否还有别的替代方案同样能完成本发明目的？**

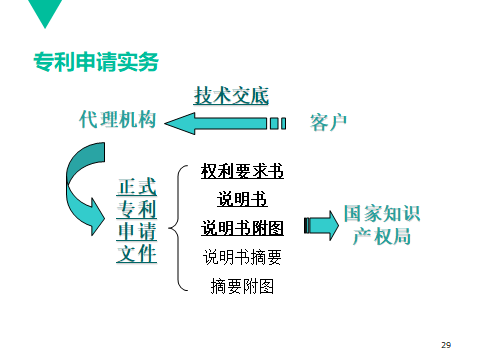
**7.其他有助于审阅、代理人员理解本技术的资料.**

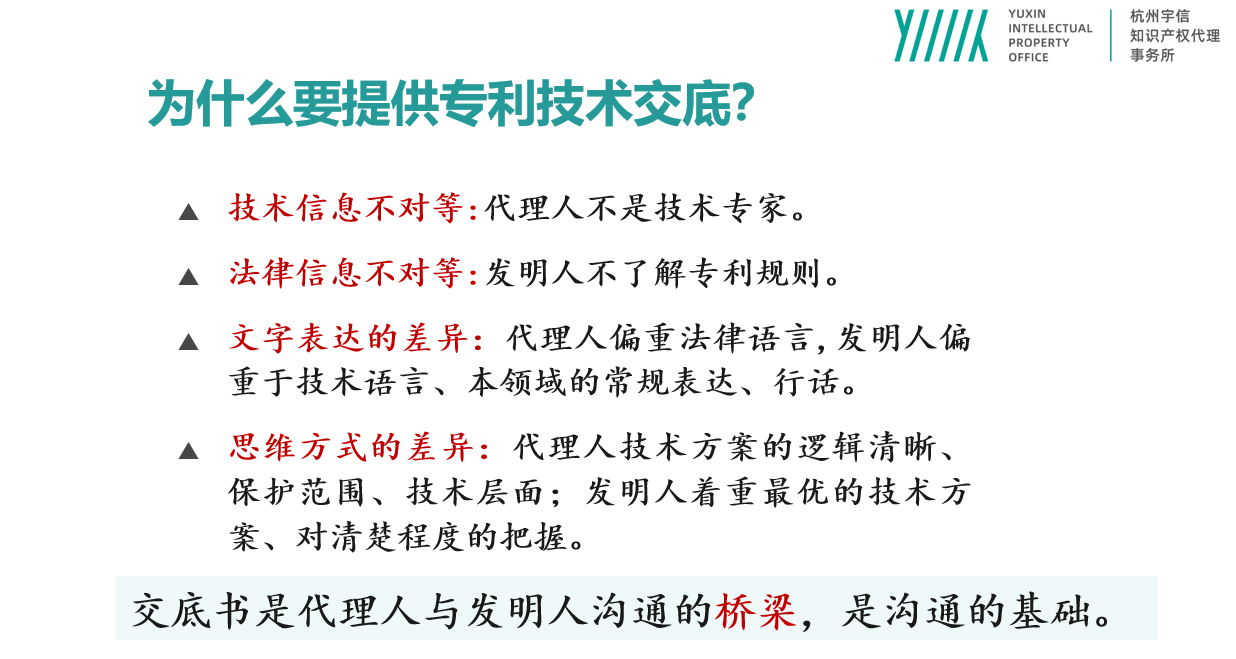
**8. 附图**

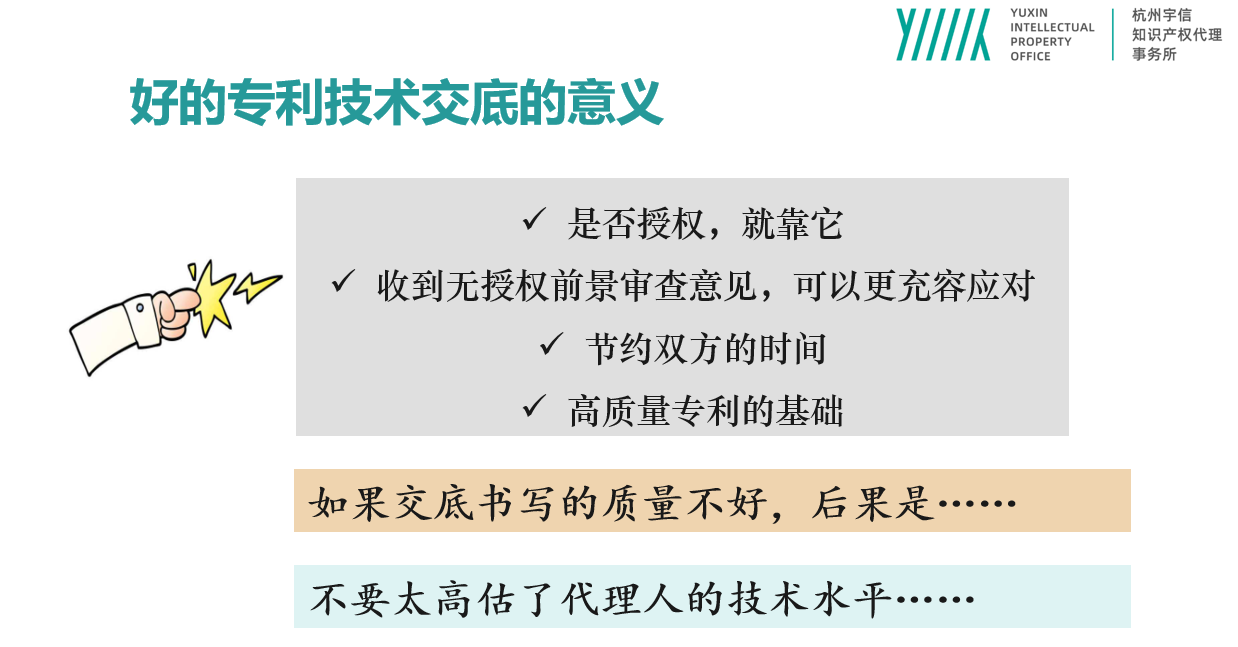
（附图是辅助性技术文件，它与说明书的文字内容一起，将本发明创造讲述清楚。可以是结构示意图、爆炸图、剖视图、局部放大图、原理图、流程图等。说明：附图可以在对本专利技术方案进行详细阐述时放在一起做介绍。）

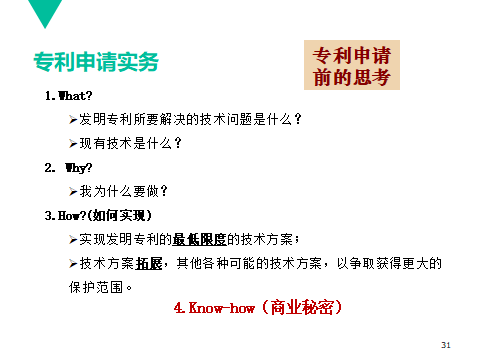
**以下几张幻灯片图是关于专利技术交底书的介绍，供参考：**

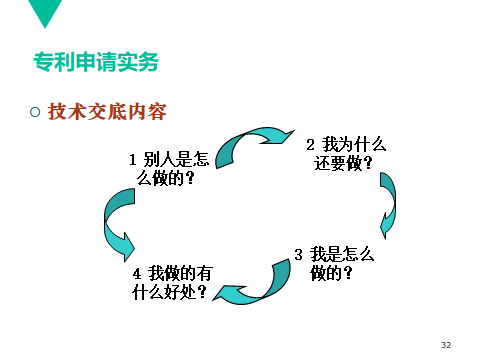
****

****





****

****

