基于局部语义地图的半自主导航

# 摘要

围绕智能轮椅和助行器的智能导航控制系统研制任务，针对部分用户使用语音控制等低通量人机交互接口在无先验信息、变化的、有障碍物遮挡的环境中移动不便甚至出现碰撞等危险的问题，提出了一种基于局部语义地图的半自主导航系统。用户前期使用语音指令控制轮椅，Mask-RCNN框架处理RGB图，输出对象实例作为视觉反馈，待识别到要到达的目标后，与对应的经过处理的深度图进行数据融合实时计算目标位姿，经由语音模块发布选择目标的指令后，系统切换为时间弹性带（Time Elastic Band，TEB）局部路径规划算法进行自动导航和避障，以安全到达目标处。该系统已经部署到智能轮椅平台并在实际环境中进行了测试。

# 1介绍

随着我国老龄化状况不断加深，越来越多的老年人需要使用轮椅帮助他们移动。尽管大多数辅助机器人系统（轮椅、操纵器）都配备了操纵杆接口，但由于各种健康状况，仍有部分老年人被剥夺了基本的操作能力而无法利用操纵杆接口控制轮椅【1】，。

对于不适合通过传统方式（如操纵杆）控制轮椅的残疾用户来说，使用语音【2】【3】【4】指令等低吞吐量接口操作轮椅是可替代的方案。但是，由于低吞吐量接口以低频率和大延迟生成控制命令，因此很难通过该接口准确高效地控制轮椅。另一方面，在未知的环境中智能轮椅通常不知道目标的位姿，无法完全使用自动导航系统。因此，如何使用低吞吐量接口控制智能轮椅是一个有待解决的问题。

共享控制【5】可以将用户控制和机器自动控制相结合，减少用户操作，使得用户更方便高效地使用低吞吐量接口控制轮椅。自20世纪90年代以来，人们提出了许多共享控制。这些方法根据控制方式可分为两类：直接共享控制【6】和间接共享控制【7】。直接共享控制在进行操作时融合用户和轮椅的命令。间接共享控制是轮椅朝着使用者指出的方向移动，辅助系统提供避障算法以确保安全。但当前共享控制把环境物体当做障碍物，未考虑物体的语义信息。

近年来，语义地图【8】越来越多地应用于机器人，尤其是智能轮椅等辅助机器人。语义地图可以帮助智能轮椅实现与对象相关的导航，例如停靠到桌椅等家具处。基于语义地图的导航有助于减少用户操作并提高安全性，但目前基于语义导航的方法物体的位置都是固定的，并不适应变化的环境。

针对部分用户使用语音控制等低通量人机交互接口在无先验信息、变化的、有障碍物遮挡的环境中移动不便甚至出现碰撞等危险的问题，提出了一种基于局部语义地图的半自主导航系统。本文的主要贡献是：

1.针对语音等低通量交互接口控制轮椅效率低下的问题，设计了共享选择策略提高效率，减轻用户负担；

2.针对无先验信息的环境，融合RGB图和深度图信息，实时提取对象位姿；

3.针对无先验信息环境中的障碍物带来的安全问题，结合Aloam定位，采用改进的TEB局部路径规划算法进行避障，更安全地对接目标；

4.该算法已部署到轮椅上，并在实际环境中进行测试

# 2相关工作

共享控制可以将用户控制和机器自动控制相结合，减少用户操作，使得用户更方便高效地使用低吞吐量接口控制轮椅。

Sheng【9】等人提出一种基于SSVEP的BCI系统，该系统带有视觉伺服模块，用于操作服务机器人。用户选择目标后，由服务机器人自动导航到指定位置。但是该系统需要预先设定路径。

Zhang【10】等人设计了一套自主导航系统，根据现有环境自动生成候选目的地和航路点。用户使用基于运动图像（MI）或基于P300的BCI选择目的地。根据确定的目的地，导航系统规划一条短而安全的路径，并将轮椅导航至目的地。使用该系统，用户的精神负担可以大大减轻。此外，该系统能够适应环境的变化。但是，该系统需要在配备摄像机的房间内使用，限制了使用范围。

语义地图可以帮助智能轮椅实现与对象相关的导航，例如停靠到桌椅等家具处。

【11】提出了一种基于图的语义建图的用于室内机器人导航的方法，该方法利用机器人特定的、语义、拓扑和几何信息扩展了OpenStreetMap（OSM）。并根据这些信息生成路线点，根据“交通规则”的建模语义，路线点的位置确保机器人路径保持在走廊的右侧。

【12】引入了SSCNav，这是一种算法，它使用置信度感知语义场景完成模块显式地建模场景优先级，以完成场景并指导代理的导航规划。给定对环境的部分观察，SSCNav首先使用未观察场景的语义标签以及与其自身预测相关的置信度图推断出完整的场景表示。然后，策略网络从场景完成结果和置信图推断动作。

以上的基于语义地图的导航都需要事先建立全局的语义地图，且不适应变化的环境。

魏【13】等人开发了一个基于语义地图的脑控轮椅导航系统。语义地图不仅包括作为通用导航系统的可穿越区域的导航点，还包括可以与人类用户交互的语义目标，例如停靠或门通道。用户只需一个命令就可以选择目标来执行操作。因此，减少了向用户请求的操作。然而文章采用几何特征判别物体类别，识别种类有限，且未考虑障碍物遮挡和避障问题。

# 3方法

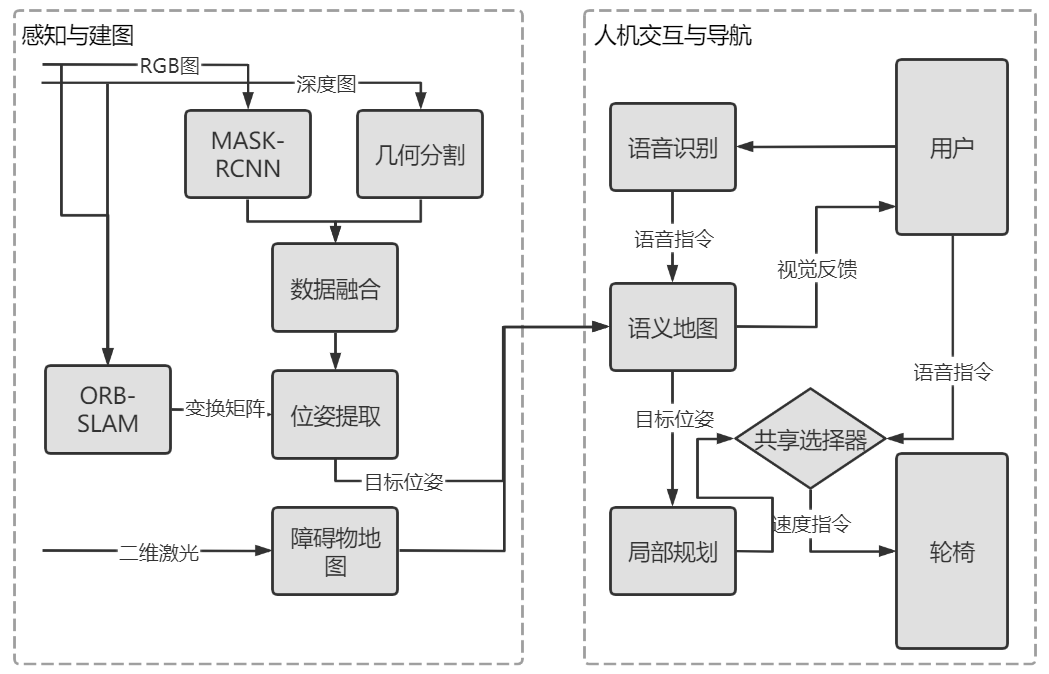


图1 半自主导航系统框架图

系统架构如图1所示，包括共享选择器与人机交互、语义建图和局部规划避障三部分。用户前期使用语音指令控制轮椅，Mask-RCNN【14】框架处理RGB图，输出的对象实例作为视觉反馈，待识别到要到达的目标后，与对应的经过处理的深度图【15】进行数据融合实时计算目标位姿，经由语音模块发布选择目标的指令，系统切换为TEB局部路径规划算法【16】进行自动导航和避障，以安全到达目标处。

## 3.1 共享控制

共享选择器在前期选择由用户使用语音指令控制轮椅，待视觉反馈识别到要到达的目标并计算得到目标位姿后，经由语音模块发布选择目标的指令后，切换为TEB局部路径规划算法进行自动导航和避障，以安全到达目标处。

人和机器在操作助行器设备时各有优劣。譬如在障碍物检测方面人眼往往不如摄像机或者激光测距仪精确，在控制轮椅的运动速度、最终位姿、轨迹曲率等方面人也不如DSP运动控制器可靠；但是在任务规划与决策方面，人往往能作出远优于机器的决策，人的控制灵活性和鲁棒性仍然没有任何智能控制器能够匹敌，另外对于纯反应式的机器控制器，还存在着控制陷入局部极小、轮椅震荡、控制不流畅等问题。

为了融合人和机器的不同优势，本文为智能轮椅设计了一种人机共享控制方法，以便为用户提供更有效的辅助出行服务。

其中，是轮椅运行的线速度和角速度，是用户操作时给出的控制指令，是切换到自动局部路径规划时算法给出的控制指令，切换时间由用户决定，通过语音交互接口发出。

本文的共享策略如下所述：

（１）在手动控制模式，由用户通过操纵杆或者语音交互接口发出控制指令操纵智能轮椅，同时语义识别模块检测周围环境的目标状况并通过视觉反馈模块呈现给用户；

（２）用户在视觉反馈模块观察到要到达的目标后，利用语音交互接口发出到该目标的切换指令，切换到自动导航模式；

（３）在自动导航模式，智能轮椅根据算法结合二维激光传感器获得的局部代价地图，规划出到指定目标的路径，并在移动过程中避开障碍物，直到安全地到达指定目标处；

（４）在自动导航模式下，用户只需监控智能轮椅运行状态，而不必亲自控制，减轻了用户的工作量，此外，用户可以随时取得智能轮椅的控制主导权，切换到手动控制模式；

本文设计的共享控制系统中的自动导航模式的目标点的位姿和局部规划的路径以及避障是基于智能轮椅配备的传感器实时自动生成的，所以系统可以适应环境的变化（比如，新增加的家具等），也可以部署到实验室，医院，养老院等不同的场景。

## 3.2 语义识别

针对存在遮挡状况下的点云语义识别不准确的问题，本文提出了一种基于语义分割的实时目标位姿估计。该模块可以向共享控制系统提供所识别目标的类别和位姿信息，以辅助智能轮椅完成人机交互和路径规划功能。

此模块包括目标语义识别和基于语义分割的目标位姿估计两部分，处理RGB-D相机的信息。首先，我们利用语义分割模型Mask-RCNN对输入的RGB图片进行处理，输出目标类别和目标在RGB图片的位置；然后对相应位置的深度图进行处理，获取语义目标的位姿信息。

语义图是连接用户和轮椅的桥梁，参考语义图上显示的反馈，用户可以通过语音交互模块选择目标，然后再利用局部路径规划模块将轮椅导航到目标。

## 3.2 人机交互

人机交互分为语音识别和视觉反馈两部分。语音识别模块设计为网页版，调用科大讯飞的API【17】来识别用户发出的指令，使用RosBridge【18】与轮椅通信，轮椅接收到指令后做出对应的处理，如图2所示。

目前可识别并发布的语音指令包括“前进”，“沙发”等10种。如表1所示。

表1 语音识别指令表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 指令 | 停止 | 前进 | 左转 | 右转 | 后退 | 桌子 | 沙发 | 椅子 | 床 | 人 |
| 编码 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

视觉反馈界面是Mask-RCNN输出的标注对象类别的语义图，识别到目标后，用户利用语音指令选择目标，然后轮椅自动导航到该目标的位置。如图3所示。

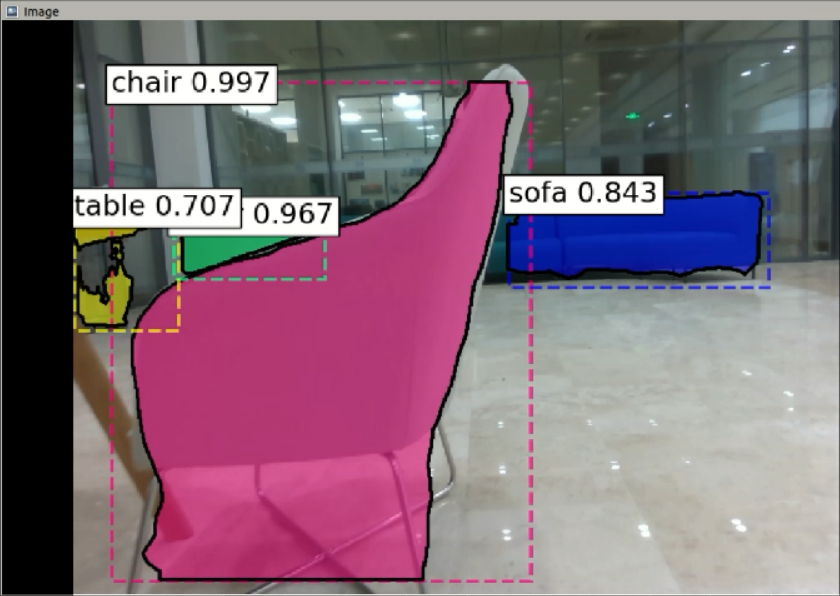


图3 语义识别示意图

语音是人类最自然的交流方式，尤其对于老残用户而言，语音接口能够更便捷地与智能轮椅进行交互，本文结合上述语义识别得到的目标语义信息，开发了网页版的语音交互界面。语音交互界面的主要作用是接收并识别用户的语音输入指令，并对其进行编码，然后与语义识别模块得到的语义信息进行匹配，实现目标的选择，然后由自动导航模块完成到达指定目标的任务。如何把人类信号变成轮椅识别的信号，过程再详细描述一下语言的符号处理和物理动作的低级控制之间的差距

该交互界面使用Rosbridge功能包实现网页端和轮椅ROS 端的通信，调用科大讯飞的语音识别接口，可部署到不同的平台上。语音交互界面主要包括三部分，一是界面设计，二是语音识别，三是与ROS端的通信。

界面如图所示，上方为主题“智能轮椅语音交互界面”，中间为语音识别的主体，显示语音识别后输出的文字信息。

语音识别部分本文调用了科大讯飞的语音识别接口，可以识别60秒的语音内容，在语音输入间隔3秒后，会停止识别并输出识别后的文字信息。

与ROS端的通信是利用ROSBridge功能包，ROSBridge是一个轻量级的用于ROS系统和非ROS系统进行通信的功能包，本文主要利用其支持的Websocket通讯方式实现网页端和ROS端的通信功能。

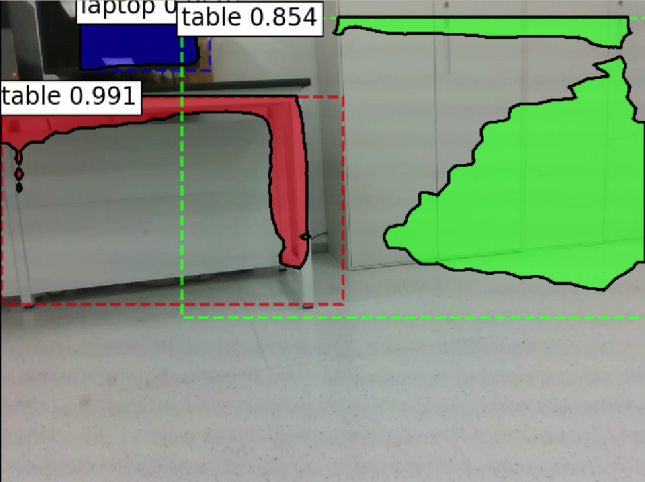
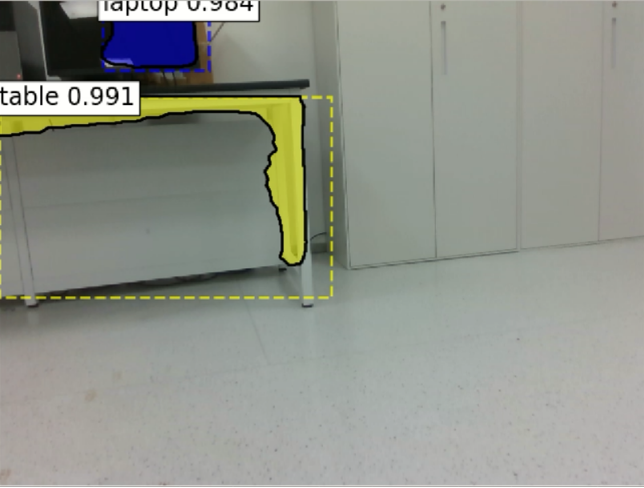
本文根据ROSBridge的协议在网页端部分定义了通信的话题，把语音识别模块识别到的指令信息按照协议格式进行了编码，在和ROS端建立连接后将指令以话题的形式发送给智能轮椅，进行交互。语音交互接口已经编码实现了前进、后退、左转、右转、停止、桌子、椅子、床、沙发、人等10种指令，并且后续可以根据需要进行扩充。

### 3.2.2 视觉反馈模块

视觉反馈模块的功能和目的是为用户提供有效的周围环境的映射，并提供可到达的目标供用户选择，减少不必要的信息。具体来说，语义分割模块识别到目标后，会在视觉反馈模块上显示目标的语义信息如桌子等，然后用户利用语音输入选择目标。

然后，使用Mask R-CNN框架对相应的RGB图像进行处理以获得目标的语义信息。具体来说，对于每帧输入RGB图像，输出一组对象实例，对于第k个检测到的实例，由ck表征对象类别，并将语义目标反馈给用户。

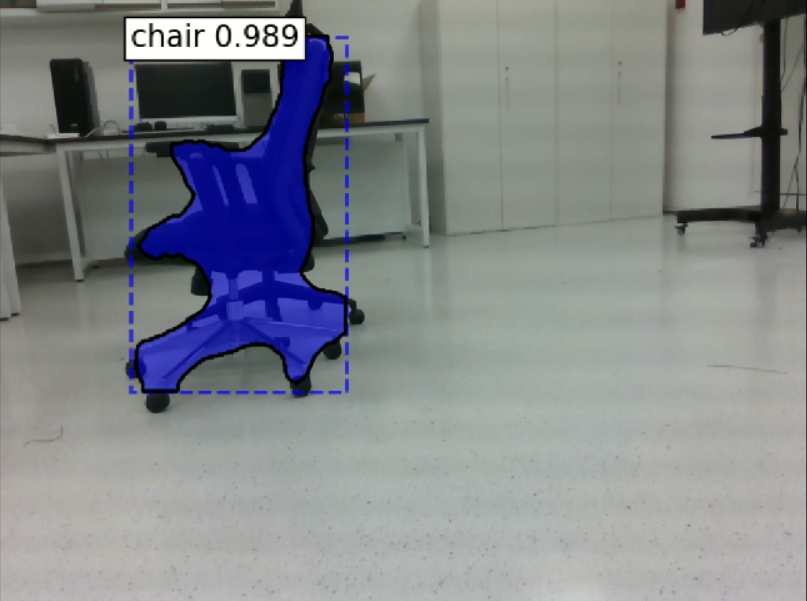
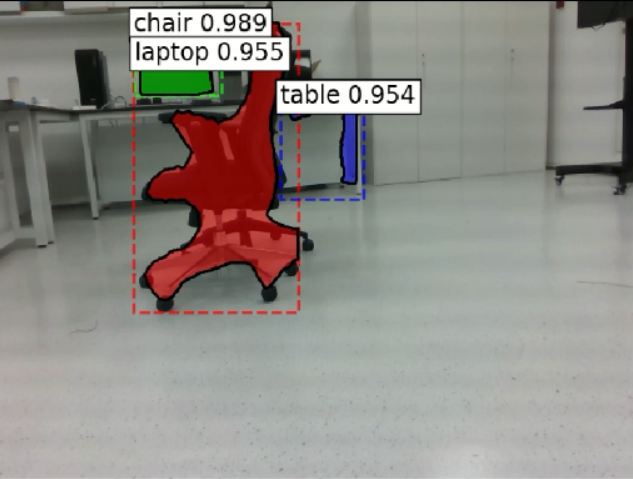
不过由于Mask R-CNN模型可以在较远的距离识别到目标，但深度相机在距离比较远时测得的深度值的误差比较大，从而影响到目标位姿的计算。此外，实验环境中存在一些和目标相似的物体，这会导致一定的误识别。为了克服这两个问题，本文提出了两项改进措施，一是设定识别目标阈值，当识别目标的准确度得分低于给定阈值时，目标将不会显示在视觉反馈模块，如图所示；二是设定显示目标的距离阈值，计算目标的平均深度值，当平均深度值大于给定距离阈值时，也就意味着目标距离较远，传感器获得的深度值误差较大，此时目标也不会呈现在视觉反馈模块，如图所示，以确保目标位姿估计的准确性。

图a 未设置分数阈值 图b 设置分数阈值

图 设置分数阈值的对比图

图中可以看到，在实际的实验环境中，存在一些和目标相似的物体，它们的得分会比较低，此时，如果设置一个得分阈值，便可以过滤掉它们的干扰，减少对用户的影响。

图a 设置距离阈值 图b 未设置距离阈值

图 设置分数阈值的对比图

图中可以看到，设置距离阈值后，在同样的位置，目标距离超出阈值时将不会显示给用户，减轻了距离较远时传感器精度不足时导致的位姿估计的偏差的影响。

## 3.3 目标位姿估计

首先，本文对深度图像进行预处理，以获得具有有效深度值的更连续的区域。接着，基于局部像素邻域估计每个深度图像点的表面法线以计算目标姿态;

前面流程通过语义分割模型Mask R-CNN和视觉slam系统ORB-SLAM2分别获得了目标的语义信息和每帧相机位姿，并定义了几个相关的坐标系，接下来，本文将重点描述一下如何获取目标的位姿信息。

通过 RGB-D 相机我们可以获得周围环境中目标的RGB图和深度信息，但在路径规划时我们需要目标在地图坐标系下的位姿。具体来说分为以下几个步骤，

（1）计算语义目标的深度值，利用Mask R-CNN网络预测输出目标的边界框在RGB图上的精确的高宽（w，h）和起始坐标（x，y），然后在空间对齐的深度图上计算目标在相同范围的深度值；

（2）语义目标深度值滤波：为了减少计算量以及语义分割模块输出目标的位置偏差和其它物体遮挡的影响，我们选择位置框中心1/8大小的区域作为目标的核心区域进行后面的计算，然后再滤除掉0值和异常值，对余下的深度值求平均值作为计算目标位姿的深度值d。

（3）获取到滤波后的目标的深度值后，通过相机的针孔成像模型，使用相机固有矩阵计算得到目标在相机坐标系下的位置，并利用平面法向量计算目标的姿态，最终得到目标在相机坐标系下的位姿

（4）利用Aloam模型得到轮椅的里程计信息，将目标的位姿由相机坐标系转换到世界坐标系；将相机坐标系的朝向变换到与世界坐标系朝向一致的变换矩阵为，设目标在世界坐标系下的位置为

## 3.4 局部路径规划与对接

得到目标在地图坐标系下的位姿后，利用二维激光获取的信息生成代价地图用于避障，本文采用TEB局部路径规划算法进行自动导航和避障，由于避障需要设置轮椅和障碍物保持一定距离， TEB规划方法到达目标处会出现震荡以致不能完成任务，所以本文在TEB算法的基础上做了改进，在轮椅到达目标附近时，将避障算法切换为运动控制算法完成最后的对接任务，以安全顺利地到达目标处。

然后根据当前位姿和最终的目标位置，切换为运动控制方法，设当前位置，最终目标位置为，其运动控制速度为：

其中，是参数，是当前轮椅的姿态。

# 4实验

## 4.1轮椅原型和实验设计

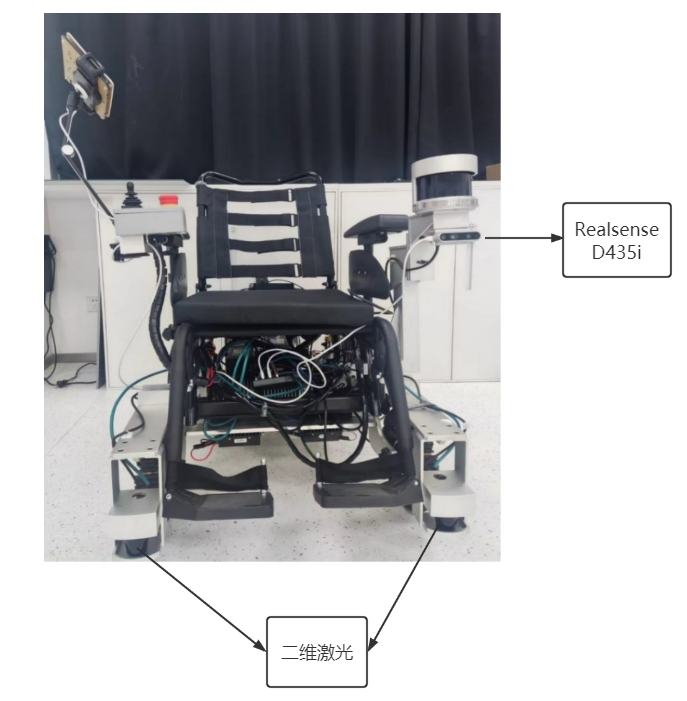


图4 轮椅原型和传感器配置示意图

如图4所示是普通电动轮椅的定制模型，其配备了视觉传感器Intel Realsense D435i，最大感知距离约为10m，以及二维激光sick-tim571，最大感知距离约为25m。计算机处理器为Intel i5，智能运动控制器（SMC）通过串行与计算机通信，用于控制轮椅的速度。系统软件基于ROS开发。

实验场景如图5所示，为上海交通大学转化医学大楼一楼的休息吃饭处，模拟用户饭后到沙发处休息的场景。用户操控轮椅识别到远处沙发后，发布语音指令“沙发”，然后轮椅将自动导航到沙发处。对比实验方案为完全由语音控制操纵轮椅到沙发处。

图5 实验场景示意图

## 4.2实验结果

本文设计了以下指标评估算法性能：

1）路径长度：完成任务所需的距离；

2）任务完成时间：完成任务所需的时间

3）对接精度：完成任务时与目标的距离

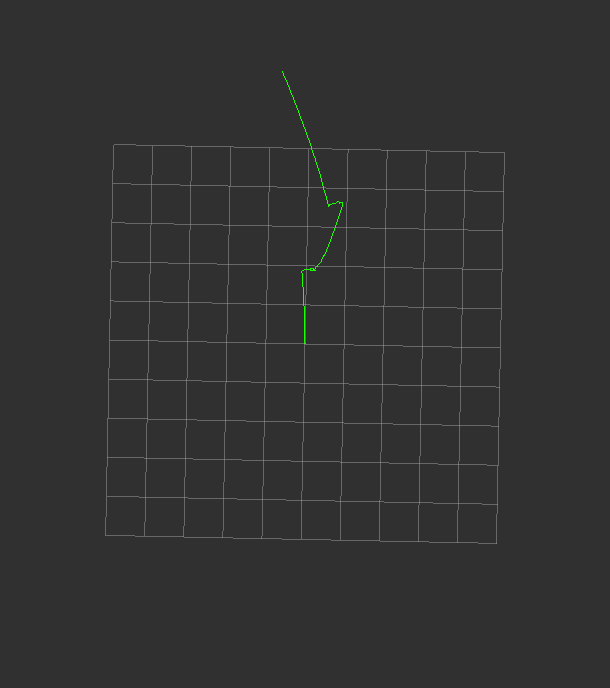
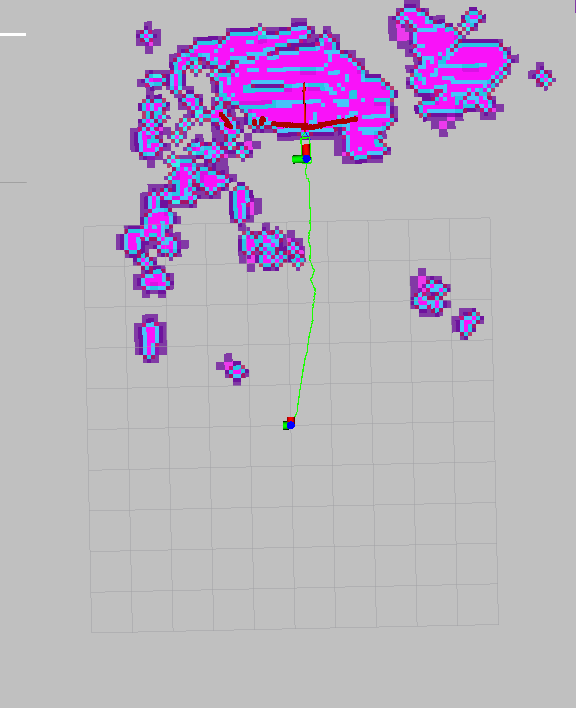
4）安全性，与障碍物碰撞次数；

5）舒适性：速度和角速度变化幅度。

表2 语音控制方式和半自主方式指标

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 与障碍物相碰次数 | 任务完成时间（s） | 对接精度（cm） | 语音指令数（个） |
| 语音控制 | 1 | 144 | 46 | 6 |
| 半自主导航 | 0 | 78 | 37 | 1 |

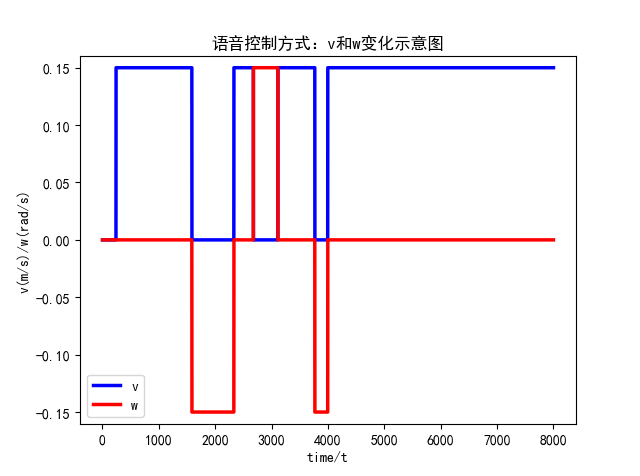
从表2中数据可以看出，完成相同的任务，半自主方式要比语音控制方式耗时短，由用户发布的指令数少，且对接精度较高，与障碍物相碰次数少。因此，半自主方式可以大大减轻用户负担，提高移动效率，增强安全性。

图a 图b

图6 运动轨迹示意图。图a为语音控制方式，绿色线为轨迹，图b为半自主方式，绿色线为轨迹，红紫区域为障碍物

运动轨迹如图6所示，半自主系统到达目标的轨迹比使用语音控制的轨迹要平滑，且半自主方式的轨迹长度为6.06m,语音控制方式的轨迹长度为8.40m，半自主方式的轨迹长度短于语音控制方式的轨迹长度。



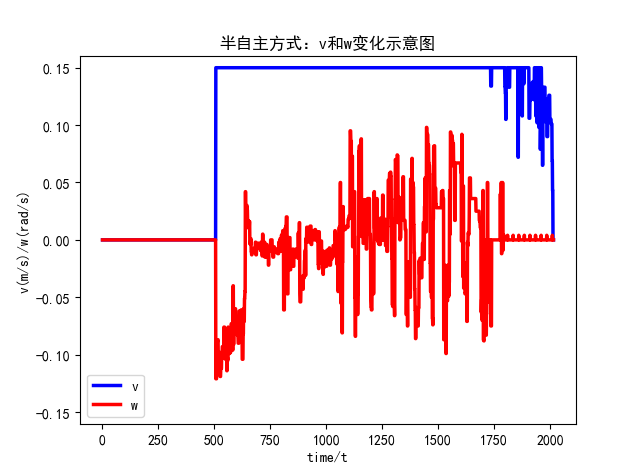


图7 线速度和角速度变化示意图

轮椅速度的变化率如图7所示，从图中可以看到，语音控制方式角速度和线速度的变化是完全陡峭的，而半自主方式的线速度和角速度变化相对比较平缓，因此半自主方式控制轮椅的舒适性更高。

# 5结论

围绕智能轮椅和助行器的智能导航控制系统研制任务，针对部分用户使用语音控制等低通量人机交互接口在无先验信息、变化的、有障碍物遮挡的环境中移动不便甚至出现碰撞等危险的问题，提出了一种基于局部语义地图的半自主导航系统。用户前期使用语音指令控制轮椅，Mask-RCNN框架处理RGB图，输出对象实例作为视觉反馈，待识别到要到达的目标后，与对应的经过处理的深度图进行数据融合实时计算目标位姿，经由语音模块发布选择目标的指令后，系统切换为TEB局部路径规划算法进行自动导航和避障，以安全到达目标处。该系统已经部署到智能轮椅平台并在实际环境中进行了测试。

未来，可以进一步优化语音识别的准确度以及位姿估计的精度，在更多样的场景中进行实际测试，以满足不同用户的需求。

# 6.参考文献