



ICRA2018-RoboMaster

人工智能挑战赛

技术报告

哈尔滨工业大学

I Hiter战队

2018.04

目录

第1章 硬件部分	1
1.1 机械结构	1
1.2 传感器选型及参数	2
第2章 软件部分	3
2.1 软件部分整体架构图	3
2.2 装甲识别跟踪	3
2.3 全场定位	5
2.4 自动打击	6
2.5 路径规划	7
2.6 自主决策	8
2.6.1 有限状态机决策	8
2.6.2 基于强化学习的自主决策	9
附录	10
参考文献	10
展示视频	10

第1章 硬件部分

1.1 机械结构

ICRA2018人工智能挑战赛的比赛内容为双方的全自动步兵机器人在布有障碍的平面场地上进行对抗，这就需要机器人拥有可快速进行全向移动的底盘和传动精准、灵活轻便的云台。同时还应按需要在机器人上安装相应的传感器，并尽可能提高机器人的可靠性和可维护性，减少卡弹等意外情况的发生。根据以上需求，机器人整体机械方案如图1.1.1所示。

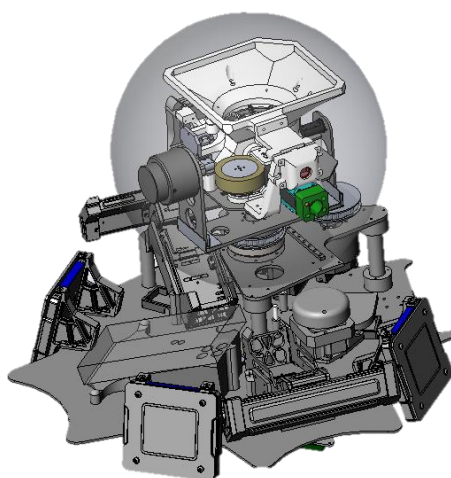


图1.1.1 机器人整体模型

机器人底盘选择了三轮全向底盘以减少重量和成本，同时利用电刷滑环设计了可以整周回转的云台，如图1.1.2所示，从而使机器人在射击目标时可以旋转底盘以减小被敌方命中的几率。

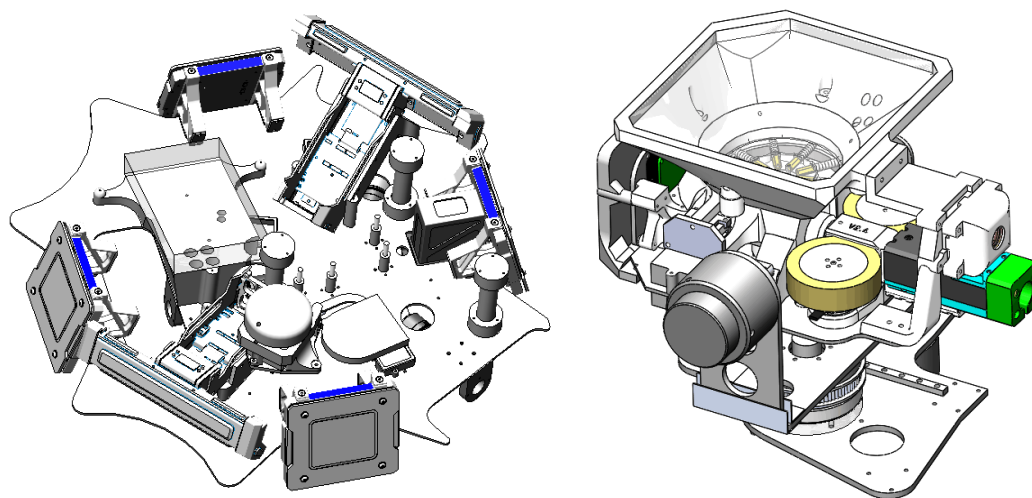


图1.1.2 机器人底盘和云台模型

1.2 传感器选型及参数

比赛中，需要全自动机器人自动搜索敌方目标并打击，对机器人的自定位能力，动态目标识别能力以及自主避障能力都有较高的要求。因此我们选择了激光雷达，里程计和陀螺仪解决机器人最重要的定位问题，选择了具有较高帧率的彩色相机解决目标识别问题，同时选用了Intel NUC为机器人提供强大的计算能力，主要传感器与处理器的具体参数见表1.2.1，每台机器人的主要设备如表1.2.2所示。

表1.2.1 主要传感器和处理器信息

类型	型号	具体参数
MiniPC	Intel NUC6I7KYK SkullCanyon	CPU: (四核) i7-6770HQ GPU: 锐炬Pro 内存+硬盘: 8G+128G固态硬盘
激光雷达	HOKUYO UST-10LX	测量范围: 0-10m, 270° 扫描周期: 25ms
相机	OV2710	分辨率(帧率): 1280×720 (60FPS), 640×480 (120FPS) 焦距(视角): 6mm (水平方向60°)
陀螺仪	LPMS-CURS2	测量范围: Roll: ±180°; Pitch: ±90°; Yaw: ±180° 分辨率: <0.01° 准确度: <0.5° (静态), <2°RMS (动态) 数据输出频率: 最大400Hz

表1.2.2 主要车载设备及用途

类型	数量	型号	用途
主处理器	1	IntelNUC SkullCanyon	运行装甲识别, 定位导航, 路径规划, 自主决策等算法
激光雷达	1	HOKUYO UST-10LX	采集环境数据进行机器人定位和避障
相机	1	OV2710	采集环境数据进行目标识别
陀螺仪	2	LPMS-CURS2	云台和底盘各一个, 用于获取云台和底盘的当前姿态角
微处理器	2	STM32	云台底盘各一个, 用于机器人底盘速度规划, 底盘和云台的运动控制
轮式里程计	2	1000线编码器	用于获得机器人当前xy坐标
通信模块	1	Zigbee模块	用于两车通信

第2章 软件部分

2.1 软件部分整体架构图

机器人的整体程序架构如图2.1.1所示，机载电脑(上位机)主要发布运动以及状态切换指令，运动控制则由云台和底盘各自的主控芯片负责。

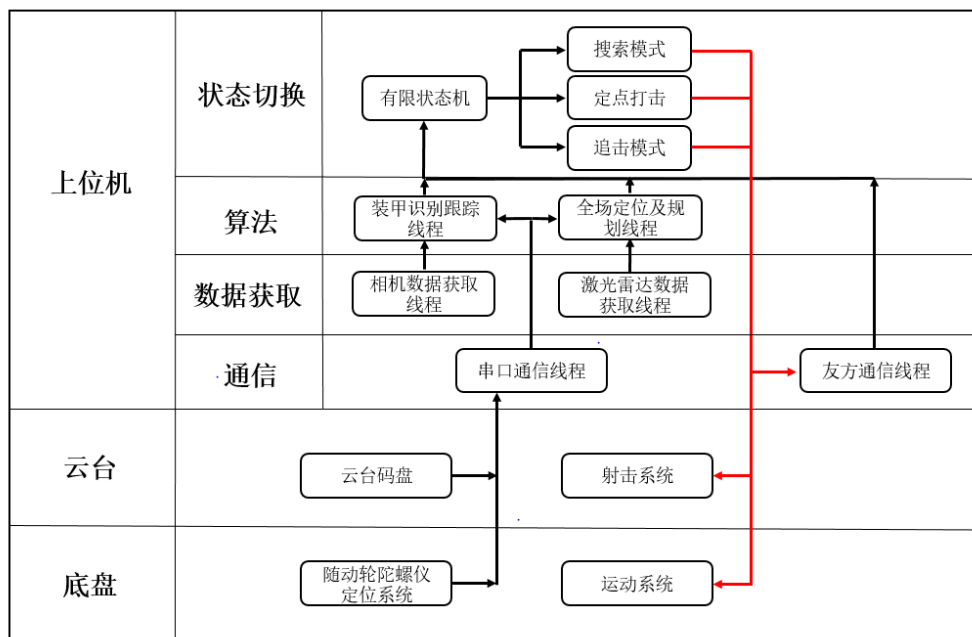


图2.1.1 整体程序架构

2.2 装甲识别跟踪

我们主要依靠装甲两侧的灯条对敌方进行识别，主要步骤如图2.2.1，其中的筛选主要依靠轮廓面积和平行度等先验。

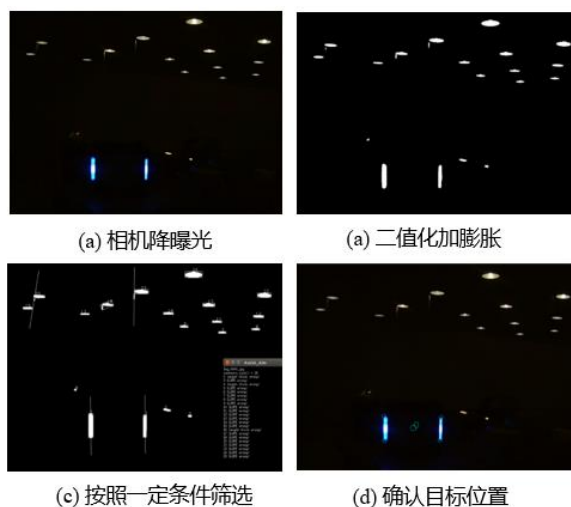


图2.2.1 装甲识别流程

检测到敌方装甲目标之后，还需要解算出其距离以求得射击角度。

求解装甲中心在相机坐标系下坐标的方法是：规定世界坐标系原点为目标装甲的中心，由于装甲的真实尺寸已知，则可得到装甲四个顶点在世界坐标系中的坐标；通过上述识别方法得到该装甲四个顶点的图像坐标，相机内参和畸变参数由事先标定得到，通过pnp算法可以解算出装甲中心点在相机坐标系下的坐标(图2.2.2)。然后利用一系列几何转换和简单物理学方程即可计算出目标在云台坐标系下的位置以及该位置对应的云台发射参数。

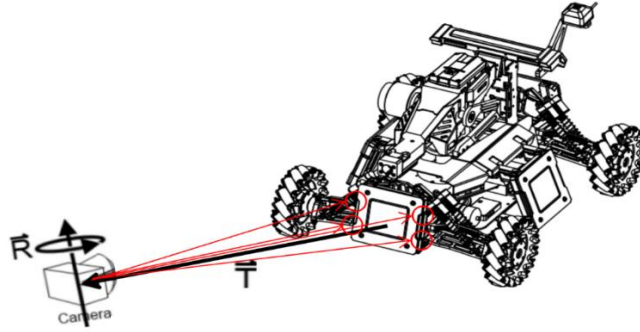


图2.2.2 SolvePnP对装甲解算原理示意图

经过实验测量，我们使用机载电脑通过上述方法进行识别的耗时为5ms左右，相机分辨率为720p的情况下可以达到50-60Hz，该算法对于静止正向装甲的有效识别距离为6m，有效测距距离是4m，基本满足了比赛的需求。但是由于计算、通信以及控制延时，对于高速移动的目标并不能达到理想的打击效果。于是我们采用基于Kalman Fliter改进的强跟踪滤波器进行敌方运动的预测。因为比赛时的机器人属于高机动目标，且我们认为装甲运动为时间相关模型，测量噪声为高斯白噪声，所以我们采用了Singer模型作为运动模型。

Singer 模型假设目标的机动加速度 $a(t)$ 服从零均值的一阶时间相关过程，其自相关函数为^[1]

$$R_a(f) = E[a(t)a(t+f)] = e_a^2 e^{-\alpha|f|} (\alpha \geq 0) \quad (2-1)$$

其中， e_a^2 为机动加速度方差， α 为机动时间常数 f_m 的倒数，即机动频率，通常取经验值。对时间自相关函数应用Wiener-Kolmogorov白化后，机动加速度可以用输入为白噪声的一阶时间相关模型来表示，即

$$a(t) = -\alpha a(t) + w(t) \quad (2-2)$$

其中 $w(t)$ 是零均值白噪声，方差为 $2\alpha e_a^2$ ，由此可以得到连续时间系统下该模型的状态微分方程为

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\dot{x}}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\alpha \end{bmatrix} \begin{bmatrix} x(t) \\ \dot{x}(t) \\ \ddot{x}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t) \quad (2-3)$$

该模型的离散形式为^[2]

$$x(k+1) = F \cdot x(k) = \begin{bmatrix} 1 & T & (\alpha T - 1 + e^{-\alpha T})/\alpha^2 \\ 0 & 1 & (1 - e^{-\alpha T})/\alpha \\ 0 & 0 & e^{-\alpha T} \end{bmatrix} \begin{bmatrix} x(k) \\ \dot{x}(k) \\ \ddot{x}(k) \end{bmatrix} \quad (2-4)$$

其中 T 为时间间隔，来预测下一次的 x 。

我们只使用了相机获取对方机器人的信息，所以观测矩阵为

$$Z(k+1) = H(k)X(k+1|k) = [1 \quad 0 \quad 0] \begin{bmatrix} x(k+1) \\ \dot{x}(k+1) \\ \ddot{x}(k+1) \end{bmatrix} \quad (2-5)$$

因为Kalman Fliter具有无限增长的记忆特性，当模型建立不准确或者系统存在输入偏差时，模型与所获得的测量值不匹配，新测量值对估计值的修正作用下降，可能导致滤波发散。为了使滤波器适用于高机动情况，需要

修改卡尔曼增益，利用预报残差的增大，使得式(2-6)成立，残差序列仍然保持正交，就可以实现滤波器对实际系统状态的跟踪^[3]。

$$E[\gamma(k+1+j)\gamma^T(k+1+j)] = 0 \quad k = 0, 1, 2, \dots ; j = 0, 1, 2, \dots \quad (2-6)$$

引入时变渐消因子 $\lambda(k+1)$ ，遗忘因子 β ，以及弱化因子 ρ ，构成新的强跟踪滤波器(图2.2.3)。

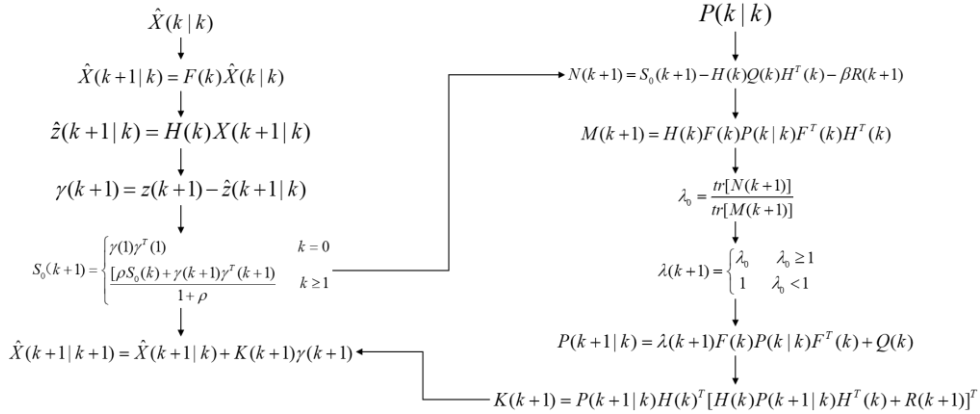


图2.2.3 强跟踪滤波器算法流程图

实际测验的效果如图2.2.4所示，可以看出，针对此次比赛内容，强跟踪滤波器相比于传统的卡尔曼滤波器对目标运动具有更好的预测效果。

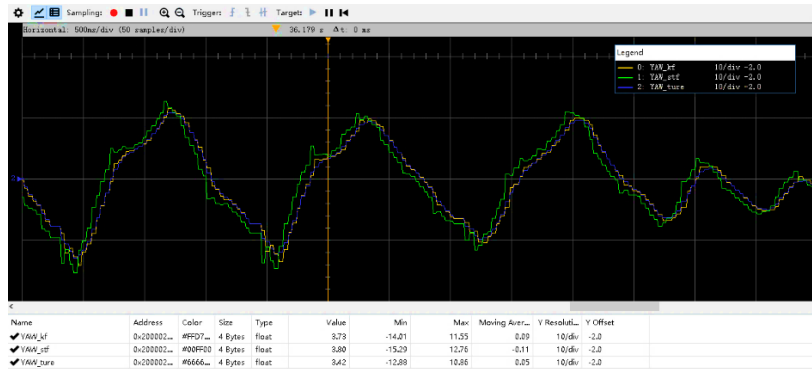


图2.2.4 强跟踪滤波器实际效果图

(蓝色为实际解算曲线、黄色为卡尔曼滤波、绿色为强跟踪滤波)

2.3 全场定位

自主定位是自动机器人的核心功能，定位的准确性直接决定了机器人的性能。本次比赛中机器人的全场定位方案采用的是陀螺仪、正交随动轮和激光雷达的融合定位方案，由于陀螺仪和里程计的采样频率远高于激光雷达，为了使机器人可以达到较高的灵活性和较大的运动刚度，故采用由stm32直接采集陀螺仪与里程计数据进行融合，并对数据进行积分得到机器人位姿，由机载电脑采集激光雷达数据与地图进行匹配，通过上下位机通信纠正累积误差。

随动轮和陀螺仪系统经过标定之后可以较准确的积分出机器人当前在世界坐标系中的位置。经测量，仅依靠该系统，机器人行进五米后的定位误差不超过5cm。完全可以将此位置作为激光雷达扫描匹配算法的先验，为了得到足够高的定位精度，我们选择了ICP (Iterative Closest Point) 算法进行激光雷达点云和地图的匹配，在有噪声的情况下，匹配结果如图2.3.1所示。

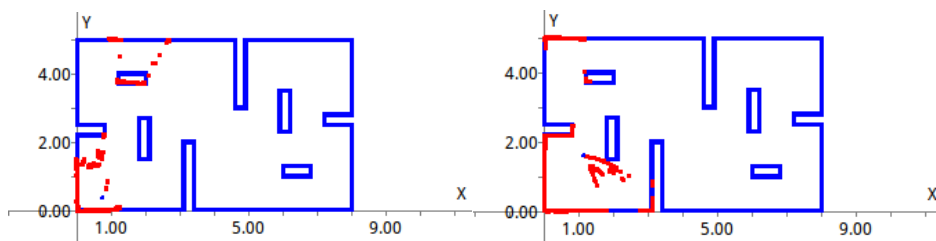


图2.3.1 部分噪声下的ICP匹配效果

由于直接进行SVD分解求解ICP问题的算法极为依赖于迭代初始值，容易陷入局部最优解，故我们采用了基于非线性最小二乘优化的LM_ICP算法来提高匹配的稳定性，目前最终的效果为全局定位误差不超过5cm。

2.4 自动打击

本次比赛中，自动打击功能主要依赖云台实现，云台上搭载相机、测速模块、发射机构等模块，在底盘运动的过程中处于整周旋转状态，同时当搜索到敌方的装甲之后通过调整自身的姿态对其进行打击。经过分析，我们认为影响自动打击效率的因素主要有以下三点：

- 1.视觉处理和上下位机通信的时间延迟；
- 2.云台的控制效果；
- 3.相机光轴与云台枪管轴线的平行度。

对于因素1，我们在视觉识别算法中添加了强跟踪滤波器进行运动预测以补偿延迟，具体方法可见本文第2.2节。

对于因素2，两自由度云台采用了串级双闭环PID控制，为了使云台的控制效果能达到预期，我们对控制系统的内外环均进行了系统辨识，得到系统传递函数后，加入相应的修正器或者控制器对整个控制环路进行改善，使得系统达到较为理想的控制性能指标，系统框图和对应的Simulink仿真框图如图2.4.1所示。

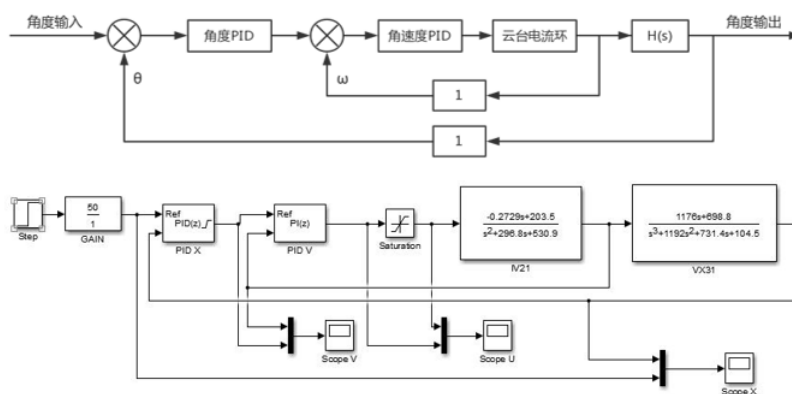


图2.4.1 云台控制系统框图及Simulink仿真
(IV21和VX31为辨识出的系统内外环传递函数)

云台最终的控制效果如图2.4.2所示，左上图为速度环(内环)在200阶跃指令仿真曲线，右上图为实测曲线，两曲线基本相符，响应时间约为80ms，超调量较小；左下图为位置环(外环)10度阶跃指令仿真曲线，右下图为实测曲线，两曲线基本相符，响应时间约为100ms，超调量约为15%，基本满足了比赛需求。

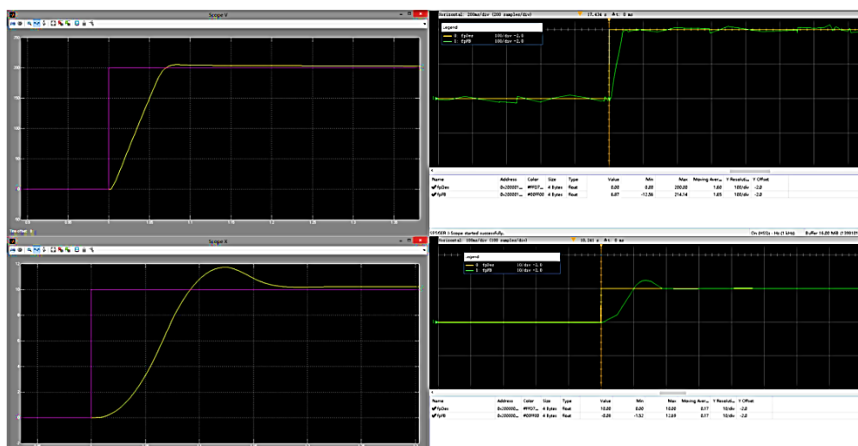


图2.4.2 最终控制效果

对于因素 3，在相机坐标系下进行目标位置解算之后，需要计算云台两自由度的偏角进行目标打击，程序计算过程中默认相机光轴和枪管轴线完全平行，但由于加工与安装误差无法避免，两者之间其实有一定夹角，这就导致了打击较远距离目标时的精准性下降，为了解决这个问题，我们利用计算相机外参的方法求出两向量间的旋转矩阵添加到程序中，经过实弹测验达到了较好的射击效果。

除上述 3 点因素外，我们在如何提高发射机构的射击精度和如何解决卡弹情况等问题上也付出了巨大的努力，并取得了一定成效。

2.5 路径规划

由于比赛场地是固定的，而且经过实际测试，各个障碍块之间可通过的最优路径比较固定，因此考虑在进行全局路径规划时将目标位置设置为人为选择的若干路径点之一，如图 2.5.1 所示，各个路径点用黑色空心圆圈表示，每个路径点与邻近点之间可以直线移动，整个地图中的路径点可以生成一个无向有权图，各条边的权重为路径点之间的直线距离，当机器人处于巡航或搜索模式时，可在该图中选择一顶点作为目标点对整个图进行 A* 搜索以寻找最短路径。若机器人的初始位置不在顶点集中(图 2.5.1 红色实心圆圈所示)，则在规划前，将初始位置当做一个顶点，并在保证路径可通过的前提下在原图中增加新边，再进行 A* 搜索。实际测验发现，这种方法相比于直接对栅格地图进行大范围 A* 搜索节省了大量的计算时间，且实际过程中可以反复调节各路径点的位置以达到更好的运动效果。

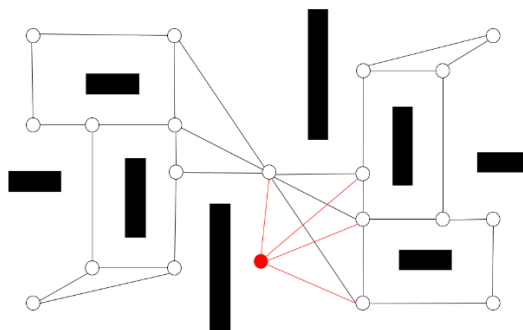


图2.5.1 各路径点生成的无向有权图

当机器人发现敌人，并切换到追击模式时，考虑到此时需规划的路径较短，目前选择基于栅格地图的动态 A* 算法进行短距离的路径规划，把识别部分计算出的敌方机器人在地图中的位置作为路径目标点，并令机器人进行步进运动(即把路径分为各个小段每次前进一段)，在运动过程中执行激光雷达的避障策略。

关于避障策略，考虑到本次比赛中的障碍物除场地元素外只有机器人，即认为障碍物大小已知。机器人运动过程中，激光雷达会实时测量下一个或几个路径点是否存在障碍物，如果存在，在排除障碍是对方存活机器人和己方机器人后，立即停止运动，重新规划路线(目前在此情况下采用栅格地图的A*搜索)，如图2.5.2所示。

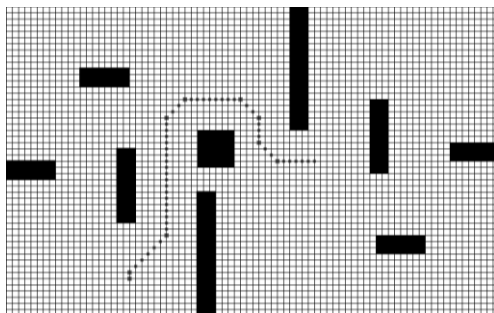


图2.5.2 发现障碍后的路径规划

2.6 自主决策

对于机器人的自主决策，我们在使用传统有限状态机进行决策的同时，也探索了通过强化学习在仿真环境中训练机器人进行主动搜索并跟踪敌方机器人的方法，具体成果见下述。

2.6.1 有限状态机决策

我们规定了机器人运行中的三种模式，分别为搜索模式，定点打击模式和追击模式。如图2.6.1所示，三种模式可以相互切换，机器人初始化完毕之后即进入搜索模式自动搜索敌方机器人，若发现目标并且目标在自动射击的最佳距离之内，则切换至定点打击模式，若目标超出最佳射击距离即切换到追击模式，跟踪并同时打击目标直至目标进入最佳攻击范围再切换到定点打击模式。在定点打击和追击模式下，若敌方机器人在我方视野中消失(如被障碍物遮挡)，随即切换至搜索模式继续搜索目标。

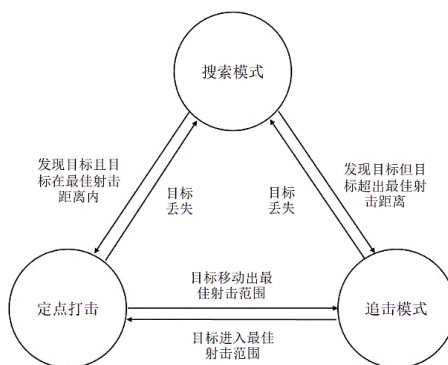


图2.6.1 机器人模式切换示意图

在上述自主决策方案中，如何自动搜索敌人是最为关键的环节，目前我们制定的搜索策略具体如下：

由于我方机器人的云台可以高速整周回旋，即可认为具有360度视野，且相机对敌方机器人的识别具有较高的成功率，故可根据机器人的视野范围确定目标位置。如图2.6.2所示，在机器人运动过程中决策程序会一直维护一张动态地图(分辨率为 500×800)用来表示敌方机器人的可能位置，我们认为当前视野范围内存在敌方机器人的概率较低，在动态地图中用较低的灰度值表示，各个像素的灰度值会随时间按一定系数增加。

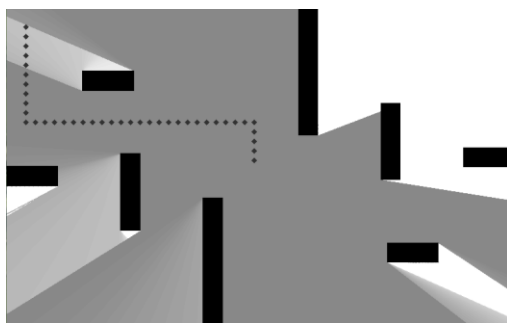


图2.6.2 机器人的动态视野

与此同时，程序会在一段较短的时间之内保存敌方机器人在世界坐标系下的坐标。在机器人需要确定下一个目标位置时，按照全局路径规划的路径点分布把地图分为若干块，计算每块地图的平均像素灰度并结合最近一帧敌方机器人的位置，选择目标出现可能性最大的地图块对应的路径点作为此次路径规划的目标，按规划出的路径继续搜索敌方机器人。

2.6.2 基于强化学习的自主决策

我们主要采用基于值函数的深度强化学习的方法，对此次比赛的最优策略进行探索。考虑到深度强化学习方法对奖励设置、网络结构设计敏感，且在动作状态空间较大、输入信息复杂时较难收敛，我们对问题做出了适当简化：

- 1.先考虑双方各有1辆机器人；
- 2.将地图离散化为 $40*25$ 个格点，机器人每步可以移动1格；
- 3.agent的视野为 360° ，在没有障碍物时，视野可以到达场地边界；
- 4.敌方在agent视野内时，可以自动射击，双方均会受到伤害，伤害量与双方距离具有一个固定的函数关系，并且己方的攻击力大于敌方；
- 5.状态输入：己方位置 (x, y) ；如果敌方在视野内，输入敌方位置 (x, y) ，否则输入 $(-1, -1)$ ；己方剩余血量；比赛剩余时间；
- 6.动作输出：上下左右或保持不动；
- 7.奖励设置：根据规则的计分公式给出回合结束的总奖励，考虑到总奖励的滞后性，每一步给予小惩罚量(激励agent更快的完成任务)；当agent做出撞向障碍物的选择时，保持agent的坐标不变并给予较大的惩罚量。

在上述简化假设下我们进行了如下实验：我们首先选用Deep Q Network(DQN)^[4]进行了尝试，然后使用Dueling DQN^[5]，并结合经验回放(Prioritized experience replay)机制^[6]。Dueling DQN 中使用Value和Action两个独立网络共同组成Q值，可以将两者估值区分开。经验回放机制则可以根据经验的重要程度做出排序，并增加重要经验被选中的概率，从而加快网络收敛，仿真环境如图2.6.3所示。

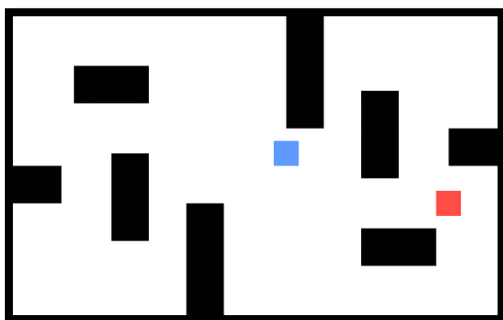


图2.6.3 仿真环境

状态输入为双方坐标，动作输出为蓝方的上下左右或者停止，奖励设置为：撞上障碍物或出界($r = -1$)、每一步动作结束($r = -1$)、蓝方碰上红方($r = 100$)，当达到最大步数或碰上红方时回合结束。

试验由易到难，具体参数设置如表2.6.1所示，网络结构及loss曲线如图2.6.4所示，试验4是在试验3训练好的模型上调优 (fine-tuning) 得到的，网络结构及参数设置相同。

表2.6.1 具体参数表

序号	红方初始位置	红方每步动作	蓝方初始位置	蓝方每步动作	状态 (state)	网络结构	网络层数	Batch Size param	e_greedy range step
1	固定	保持不动	固定	DQN输出	[2]蓝方坐标	Prioritized DQN	4层FC	64	0-0.9 5e-5
2	固定	保持不动	随机	DQN输出	[4]红蓝坐标	Deuling DQN	4层FC	64	0-0.9 5e-5
3	随机	保持不动	随机	DQN输出	[4]红蓝坐标	Deuling+prioritized	8层FC	256	0-0.9 1e-6
4	随机	随机	随机	DQN输出	[4]红蓝坐标	Deuling+prioritized	8层FC	256	0-0.9 1e-6

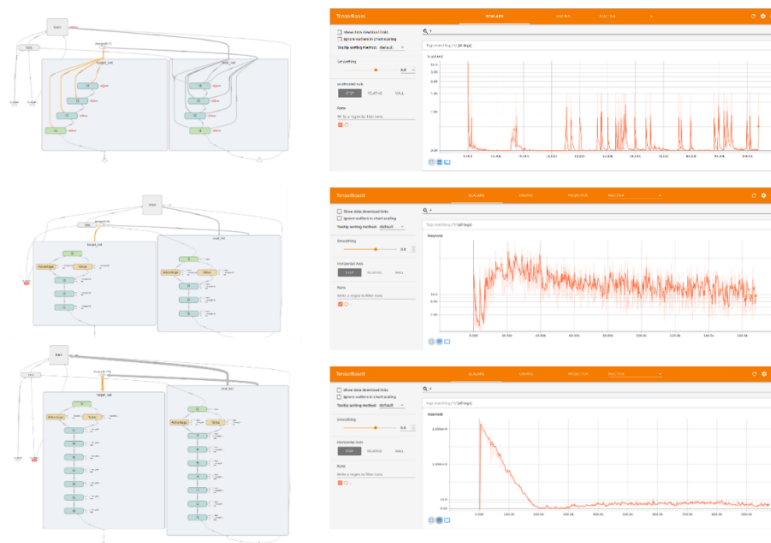


图2.6.4网络结构及Loss曲线（从上至下依次为1、2、3）

附录

参考文献

- [1] Jing Y, Yuan-fu Y. Application of Singer Tracking Model in Adaptive Gps Signal Tracking Algorithm[C] Instrumentation, Measurement, Computer, Communication and Control (imccc), 2012 Second International Conference on, [S.l.]: [s.n.], 2012: 33-38.
- [2] 张燕,柳超,李云鹏. 基于改进Singer模型的机动目标跟踪方法[J]. 火控雷达技术, 2015, 44(3): 37-40, 50.
- [3] 叶 斌,徐 毓. 强跟踪滤波器与卡尔曼滤波器对目标跟踪的比较_叶斌[J]. 空军雷达学院学报, 2002, 卷缺失(2): 17-19, 22.
- [4] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540):529.
- [5] Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning[J]. 2015:1995-2003.
- [6] Schaul T, Quan J, Antonoglou I, et al. Prioritized Experience Replay[J]. Computer Science, 2015.

展示视频

视频网址为: http://v.youku.com/v_show/id_XMzUyODAwOTkwNA==.html?spm=a2hzp.8253869.0.0

观看密码为: hitcrt2018