自动瞄准的过去、现在以及未来

RoboPioneers 视觉组 2021赛季第1阶段 阶段性技术发展报告

## .总述

自动瞄准，年纪轻轻，就已老态龙钟。

# 过去

很多人认为自动瞄准诞生于2017年，因为那一年能量机关只得自动激活，而那一年以前的资料则又稀缺而不可考据；但事实上，自动瞄准与自动单位一并诞生于2015年。自动瞄准真正被发扬光大是2019年：这一年，搭载第二代自动瞄准系统（工业相机、卡尔曼滤波、数字识别）的二代步兵，把使用第一代自动瞄准系统（USB相机）的一代和二代步兵，彻彻底底地碾在车轮下。从年龄上看，从2017年到2021年，第二代自动瞄准系统才3岁，尚且年轻，但其成长状况又与之自相矛盾。

2017年，大疆开源了其自动瞄准系统的程序：使用BGR色彩空间，通过红色通道和蓝色通道的颜色值相减，来过滤出需要的颜色的区域，对这些区域进行边缘检测，筛选包含两个小轮廓的外轮廓，再用旋转矩形匹配这些轮廓，从而得到这些轮廓的几何特征，对这些特征进行计算和筛选，从中匹配出能够组合出装甲板的灯条，再对其外接矩形进行特征匹配和筛选，最终选择一个面积最大、最接近中心的“最合适的装甲板”——这就是“几何阈值法”的鼻祖。2018年，大连交通大学使用三个USB相机构建了地盘双目、炮管绑定单目的相机系统，并在几何阈值法中加入了角度解算、数字识别，使得识别准确率大大提升，并且双目系统能够提供三维的目标坐标，从而辅助云台打击角度的计算。从结果上来看，大连交通大学的方案的打击效果非常好；同年也有很多学校使用这样的双目系统。但是在2019年，价格高昂、性能优异的工业相机的广泛引用，使得双目方案多数被替换回了单目方案，理由很简单：并非所有的学校的技术都足以从双目中受益、工业相机高昂的价格使得双目方案的造假过高、单个工业相机方案的性价比远超双目普通相机方案。2019年“几何阈值法”的代表是上海交通大学，在这一年，大部分大学都已经认识到了“几何特征法”实际上是“几何阈值法”，其核心关键在于几何特征的选择和阈值的设定；同期，视觉组研发人员们也有意识地开始研究三灯条、四灯条难题，也有越来越多的人意识到：几何阈值法的核心，几何特征和阈值，已经近两年没有变化了；就连这些难题的解决方案——挑选几何特征、调整阈值，也近两年没有变化了。

换言之，几何阈值法已经近两年没有发展了：人们在做的，只是调整。但这不足以成为宣判几何阈值法“死刑”的理由；真正的理由是：在这两年的时间里，几何特征法最大的弊端，也没有变。

几何特征对于光照具有极高的依赖性，当曝光时间过长时，灯条过曝会加重，从而导致灯条和场地在画面中具有相同的亮度，而这会直接重创几何阈值法的根基——灯条识别。因而，要么选用贵的离谱的相机，要么降低相机的曝光时间，而这么做也不是没有代价：画面整体会变得昏暗，甚至于除了发光、反光物外都几乎不可见的程度。而且战斗单位主体本身也是黑色的，这会导致单位的轮廓等信息几乎不可用。而且就算是同一个地方，每天的光照情况、场内氛围光的情况，也很难是一致的；这一是要求开发人员选择颜色条件的时候，要考虑排除氛围光的干扰，二是要求几乎每天调整一次阈值。而很可惜的是，研发人员研究几何阈值法已经接近两年了，尽管每个学校都积累了自己独特的选择几何特征的经验，但对于几何阈值法正确率较低、阈值难于确定等问题的解法，一直都没有质的飞跃。各项高级技术也受到了几何阈值法的影响，比如“反小陀螺”和“运动预测”。这些技术本身不难于实现，但在几何阈值法的观点中，画面只有概率的灯条和概率的装甲板，不借助时序很难确定哪些装甲板属于同一个战斗单位，单凭对二值化的灯条区域的几何分析，又很难单凭PnP解算出目标战斗单位的运动状态。

在过去的拓荒年代，这些都是几何阈值法的提升空间；但是这么些年后，这枯涸的泉眼已经无法给研发人员提供名为希望的甘露了。

## 现在

在2019年赛季，我队视觉组研发了自有版本的几何阈值法，但是选择的几何条件和设定的阈值过于宽泛，导致最终的解决方案是相机的曝光和光圈均设置得只能看到近处的灯条；在2020年赛季，视觉组的任务是移植上交在2019年开源几何阈值法。经过2020年一整年的研究和解析，我队视觉组已经实现了独立研发的几何阈值法版本 “Prometheus” 系列（P系列），而实现方案也极其地简单且高效。

只有知道几何阈值法的优点与缺点，才能够做到扬长避短；而其优点，是开发简单、运行速度优异，缺点的根源，是难以快速确定合适的阈值。“Prometheus Mk.1”（后称“P1”）以长宽比为核心特征，结果不适合应对旋转目标；“Prometheus Mk.2”以角度为几何特征，结果旋转矩形法匹配所得的角度的不稳定，直接导致装甲板识别结果不稳定；经过对前两款的优缺点的分析，后推出了更加纯粹的几何阈值法版本“Prometheus Mk.3”，该版本不进行任何复杂的几何特征计算，只使用最简单有效的几何特征去筛选目标，并且随附可视化的实时参数调整工具，使得阈值设定的流程变得极为快速高效。

P3的出现可以说是划时代性的。这种划时代性体现在三个方面：一是阈值法最大的弊端——阈值难以确定、因而正确率低的问题，被可视化实时调参工具的出现而近乎完美地解决了；二是可视化实时调参工具的出现，极大地增强了几何阈值法对于不同光照环境、不同镜头的适应性，从而近乎完美地规避了几何特征法泛用性差的问题；三是可视化实时调参工具，降低了几何特征法的使用门槛，从而使得不具备高性能、高检出率自动瞄准系统的队伍，能够在拿到该系统后的数日内，快速具备与设备高度适应的自动瞄准能力。

但可视化校准工具的出现，并非标志着几何阈值法即将进入鼎盛时期的黎明；相反，更像是几何阈值法的丧钟。从2020年开始众多视觉水平较高的高校纷纷转向以机器学习为首的第三代自动瞄准系统的现象来看，几何阈值法即将普及、并同时走入升级优化的死胡同。

正如几何阈值法结合工业相机最初只被少数队伍研发，并随后被战斗胜利带向四面八方一样，各种迹象都在表明，当初的几何阈值法的先驱者们都在走向半机器学习或全机器学习的道路。比如上交的哨兵，2021赛季采取的方案是，使用几何阈值法识别图片，生成训练集，并使用这些训练集来训练神经网络。当然，结合的方法还有很多；甚至不使用机器学习的方法也有很多，比如以运动检测和运动追踪为核心、以双目视觉的立体空间感知为核心等。

但这次“技术大拓荒”与上一次不同的是，几乎所有的队伍都明白：先掌握次世代技术的队伍，可以爆锤未掌握的队伍。同时，随着各种战术需求的产生，视觉组也逐渐“偏离”最初的“视觉组”。

# 未来

就如同大家普遍认识到了当下自动瞄准优于人工瞄准一样，可以预见，在未来的某一天、也许就是2021赛季，人们同样会发现自动取弹药箱、自动驾驶等在某些方面同样占有压倒性优势。

理由是极其简单的：机器，拥有人类无法靠经验匹敌的反应速度、动作精度；而其唯一落后于人类的，就是智能和决策深度。也就是说，在操作手无法得到极其充分的练习、智能技术相对较完备的队伍中，在较为依赖反应速度的方面，智能系统对于人类的控制具有压倒性的优势。

除此以外，2020年开始的比赛规则赋予了智能系统更多的优势：比如操作手之间没有心电感应，但单位间可以进行较高速度的通信，凭借以雷达为核心的战术数据网，可以做到单位与单位间共享信息、直至做到无死角地预警周边环境，从响应速度上提高单位的战斗能力；再比如血量、枪口热量、功率等高到夸张的全自动步兵，在一对一的战斗中可以无悬念地爆锤普通步兵；而在论坛等交流平台上，大家则畅想了智能系统更多的可能：主动让底盘棱角面对敌方枪管的主动防御、雷达和哨兵的协同防御、自动步兵的自动巡逻与支援……

正如2017年的大能量机关，主办方在规则中设置的一切都在暗示同一个方向：智能化、网络化、一体化；即便没有如此的暗示，所有人也都在不同程度上意识到了这个大方向，因为机器控制机器比人控制机器有着明显的优势：更完全的状态掌握、更实时的运动控制；而阻碍这个潮流的壁垒，如智能程度等，在近年来飞速进步的算法和运算设备的攻势前逐渐瓦解。未来的方向是如此的明显、如此的简单：人将负责决策，机器将负责执行，各司其职、各取所长。

在这个大方向下，自动瞄准的技术体量、战略地位不会萎缩，但同时，它却不会再独立地作为一个关键技术，而是成为一个更高层次的系统的一部分——智能感知的一部分。

智能化需要单位周边、战场环境的战斗数据，而仅靠自动瞄准去定位装甲板的位置是远远不足的，这样的需求不仅需要知道敌方单位的位置，还需要知道该位置上地方单位的类型、运动状态、可能正在进行的动作，以及我方单位的位置、运动状态和可能进行的动作。作为整个智能系统的基石、数据输入的唯一来源，智能感知只有做到并且做得更多，才有可能使得智能系统发挥出其真正应有战斗力。而无论其可行的实现方案是什么，都可以确定这样一点：专攻灯条的几何阈值法不会位列其中。

从这样的角度看，以几何阈值法为核心的自动瞄准已经完成了其历史任务，但却无法再继续满足未来的需求了。接下来，目光不能只放在灯条上了——我们需要看的更多，我们需要“感知”。

# 下一阶段发展方向

尽管几何阈值法的未来已经是板上钉钉的事，但目前仍有很多高校尚未掌握几何阈值法的精髓也是不可否认的事。更何况，P3作为几何阈值法具有划时代意义的系统，在面对第一代瞄准系统和不完善的几何阈值法时仍具有压倒性的优势。所以，在短期内，视觉组不会放弃维护使用几何阈值法的系统；但是，同样也可以确定，P4将会是最后一版使用几何阈值法的系统，而这，也是发展的必然。

正如广角镜头的作用被重新审视一样，自动瞄准的性能也在被重新审视。在过往，评价自动瞄准性能好坏的指标仍处于蒙昧状态，唯一有具体数字可读的项目只有寥寥几个：识别距离，和运行帧率。而至于识别距离，其界定标准更是模糊不清：也许是“能看到”，也许是80%的正确率等。虽然硬性标准必然是看枪管能否顺畅地跟随着装甲板运动，但软性指标的要求几乎只是帧率，即每秒钟处理了几次的图像。这点有着深刻的原因：因为电控组使用PID的方法控制炮台跟随目标，每秒接受的数据包数量越多，卡尔曼滤波等处理方法的效果就会越好。但事实上，帧率的要求如此之高的原因，在于现今自动瞄准的控制方式是实时控制，即使用识别出的目标坐标实时地调整云台角度，而云台角度的调整频率是500Hz，要提高总帧数，才能提高下位机调整云台的实时性。总而言之，帧率其实不是指标，下位机云台的实时性才是；提高控制效果的方案不只有提高总帧率，尝试研究并使用指令式控制方式也是解决方案之一。

从实时控制到指令方式必然是未来的大方向，这一点也是非常好理解的。自动瞄准、云台控制、发射系统，结合在一起，归根结底只有一个目标：让弹丸命中目标。注意，是让弹丸命中目标。

在自动瞄准相对较长的发展史中，弹丸总是最后一个阶段的测试和调整内容，在这之前，总是以激光红点取代弹丸——这一点其实也好理解，只有瞄的准，才有可能打的准；因而，总是要求红点持续跟踪装甲板的正中央位置。但是在过去数年的发展中，尽管“提前量”和“补偿角”在最初就被提出，但我们几乎所有人还是忽略了一个事实：红点跟随和弹丸命中是表面相同、根本不同的两件事。、

如果我们能够以激光红点跟随装甲板的中心位置，就能做到以激光红点跟随相对于装甲板的任意位置，比如考虑了提前量和补偿角的位置。在实时控制下，激光红点应当实时地跟随这些位置，从而在任意时刻发射弹丸，弹丸能命中正好达到这些位置的目标。

但事实上，跟随并不是从头至尾的要求——至少第一发弹丸需要的不是跟随，而是“截击”，而且，首发跟随抑制了自动瞄准的性能发挥。自动瞄准总是能在13ms以内找到目标在整个摄像头画面中的位置，而目标并不总是在当前云台指向位置附近，云台指向精准地抵达该位置，总是需要百毫秒级别的时间；云台运动的速度，即使肉眼无法察觉，事实上也是经历了加速-高速-减速的过程，才能做到精准指向，而不至于未达或过调；而在这个加速-高速-减速的三阶段过程中，强行提高高速阶段的占比，极易容易导致云台过调，其指向在目标附近高速抖动；唯一的解法是增大单次运动路径：总运动路径相同的情况下，单次运动路径越长，单次高速阶段占比就越大，相应地总用时就越短。而事实上，云台首次指向目标，总是要经历较长的总运动路径，而若该阶段使用实时控制，则云台总是试图逼近目标13ms（启用卡尔曼滤波后2ms，具体取决于云台控制频率）的位置，而该位置总是在不断变化，因而云台总是无法长时间处于高速运动阶段；继而导致虽然云台的运动速度上限事实上很大，实时控制时却总是无法追上目标的位置——因为速度总是根据需要的运动路程确定的，速度过快，则会在目标点附近“振荡”，反之，速度过慢，则总是追不上目标。

但是，云台需要一直跟随目标的某个位置吗？即使弹丸的射频是40Hz（爆发优先，三级，400热量上限），相邻两发弹丸的时间间隔也有25ms；而在25ms里，自动瞄准系统已经识别出了2次坐标，云台已经控制了12次；而非常不幸的是，即使考虑到预测，这12个控制坐标里只有第12个这1次的坐标是有效的。这也就是说，云台事实上的每一次发射的工作流程应当是：估计出25ms（假设25ms内重叠了发弹的控制延时）后目标加上弹道补偿的位置，在这25ms里将指向运动到该位置，打出弹丸。别小看了这2ms和25ms的差距——在2ms实时控制下，要求25ms内的12个控制阶段的结尾阶段云台指向都是精准的；而在25ms实时控制下，只要求这最后的数毫秒内云台指向是精准的即可；而精准的代价，是受限制的速度。

在这25ms甚至42ms（全自动步兵，三级，240热量上限）里，技术能有哪些突破？至少，对于视觉系统而言，25ms足够完成一轮机器学习的识别的和分类，从而实现对视野内环境的感知，以及目标运动状态的识别，和使用复杂的弹道模型反解云台的发射角度，从而实现更为智能的打击点的选择、更为精准的命中。

因为常识中，帧率总是越高越好，因为这样红点才能更好地跟随装甲板；所以必然会有人对其做出质疑。可以肯定，这样的方案必然有代价，但代价，不会是战斗能力的降低。在这样的方案下，机械的供弹系统、电控的发射系统、视觉的瞄准系统，它们的稳定性的重要程度提高了不止一个档次；只要任何一个环节的出现了异常，比如供弹系统中的拨盘、弹路卡弹、空弹、发射系统的转速不稳定、瞄准系统的误识别，都会导致估计的打击点失效，从而弹丸无法精准命中。但这里也必须要指出，即使不采取这样的控制方案，这些问题也是要解决的、也是要避免出现的。如果不管怎样。最终目标都是避免这些问题出现，那可以说，这些不属于该方案需要额外付出的代价。

如果我们继续维持原来的“红点跟随”方案，要如何提高远距离的命中率呢？假如自瞄的帧率达到了200fps，即达到了相机的硬件上限，可以有效提高命中率吗？

这个时候自瞄反馈的是目标5ms前的坐标；但是从5ms到13ms这8ms里，目标能运动多少？有人也许会认为这8ms的运动偏差会积累，这种想法是错误的：摄像头捕获的画面总是1ms前的，自瞄处理后得出结果总是13ms前的，从摄像头捕获画面到得出这个画面的结果这12ms中所捕获的12个新的画面，都被丢弃了，并不存在排队的现象，自然也不存在误差的积累。总而言之，200fps、100fps、30fps的差别，在于结果是5ms前的结果、10ms前的结果和33ms前的结果，这最大28ms的时间差，在3m/s的运动速度下，体现的偏差是8.4cm。可能有人会惊呼不可接受：大装甲板的宽度只有23.5cm，而小装甲板的快递才14cm！但我们仔细来算一算这笔账，即使采用17mm发射机构的单位选择弹速优先，弹丸的速度为30m/s，8m的距离也需要弹丸飞行266ms；而若是选择爆发优先的15m/s弹丸速度，则需要弹丸飞行522ms！相对于距离预测这百毫秒级的飞行时间差估计，这26ms的估计几乎不算难题——如果这26ms的提前量都无法预估，则可以直接放弃8m外的动态目标的攻击。对于8m外的大能量机关，在266ms的时间里旋转了19.89度，在522ms的时间里旋转了39.77度！这也就是说，无论怎样，提前量估计都是无法绕开的技术；弹路通畅无阻塞、发射延迟稳定、射速稳定、自瞄坐标稳定，都是无法避免的、必须要达到的指标。而若我们的发射机构达到了这样单发命中8m外目标的能力，即具备了能够驾驭522ms的延时的对攻击系统的精准控制能力，驾驭命令式控制相对于实时控制那至多25ms的延时简直轻而易举。

总结一下：如果按照目前各大高校追求的方向，我们所追求的极致是不完全符合要求的极致：它永远无法保证对高速的、运动状态不断变化的目标的高命中率：要么抖动明显、要么无法跟随、要么这一要求只能在特定的距离满足，其他距离无法企及或是一塌糊涂。

而本部分希望强调的重点其实只有一个：供弹模块、发射模块、自瞄模块，其实构成了一个整体，而且也只能以一个整体的形式发挥作用，而这个整体就是攻击系统；而检验这一个系统的指标只有一个，就是弹丸能否命中；如果弹丸在结果上无法命中，那么这三个模块无论哪一个模块看上去表现甚佳，其实质上也只是花拳绣腿；而为了实现近距离、中距离、远距离一致的高命中率，对各个模块的指标也是一致的：稳定。稳定是连贯的基础、稳定是预测的基础、稳定是结果按照预测发展的基础。

在过往，对于这些模块的整体性认知一直处于蒙昧状态，对于预测系统的重要性和难度的评估也处于模糊状态，也并未实质上自觉地意识到提前量控制能力的需求高达500ms；因而，过往的对这些模块的评测指标一直是零散的、看似一环套一环但却缺乏整体性的。

虽然如此，但有一点事实上早就被感知到了：那就是最终要求是命中目标。在数年的各自为战、埋头苦干中，各个模块的指标成功地并显著地影响了发展方向，但既然今年实现了对这一历史根源久远的问题的再发现，那就应当再次强调一遍：终极要求是命中目标，而且是连贯的、精准的、以最高射频命中近、中、远距离高速、高变化可能的运动目标；而那些技术指标，都是确定了实现方案后，才根据作战需求确定的；若是技术指标限制了更接近终极要求的实现方案的发展，则是本末倒置。

因而又必须着重重申一点：作为攻击系统的一个环节的研发部门之一，视觉组的终极要求也是弹丸命中目标。在此要求下，所有的实现方案没有一个是固定的、必然的，只要存在更可能的、更接近终极目的的方案，视觉组就有义务、就有责任去研究实现；若是其他组忘记了终极目的不是纸面的参数，而是弹丸在实际上命中目标，视觉组就有义务、有责任重申这一点；而若无形的、墨守成规的惯常方案限制了这一终极目的的实现，视觉组就有义务、有责任解构其中不合逻辑、不合实际的“常识”。

归根结底，所有的组都是服务于战斗，而非服务于所谓的“常识”，也非所谓的“主流”。再此基础上，我们也更有底气去重申这一坚定且明确的结论：以几何特征法为主的传统自动瞄准死期将至，而未来，将是以视觉感知、运动状态识别为基础来实现的瞄准模块、供弹模块和发射模块三位一体的、协同工作的、高度智能的自动攻击系统的天下。

贾浩宇 2021/1/30