基于激光测距与单目视觉的微型无人机室内目标人物搜索方法研究

高天禹¹, 马雨婷¹, 韩成哲¹, 蔡博^{1,2}, 翁睿¹, 张立宪^{1*}

1. 哈尔滨工业大学航天学院, 哈尔滨 150001;

2. 系统控制与信息处理教育部重点实验室, 上海 200240

* E-mail: lixianzhang@hit.edu.cn

收稿日期: 2020-03-27; 接受日期: 2020-07-15; 网络版发表日期: 2020-07-19

系统控制与信息处理教育部重点实验室开放课题(编号: SCIP2020002)资助项目

摘要 微型无人机一般指尺寸在10 cm以下的超小体积无人机, 其载荷能力有限, 无法携带高性能传感器及机载计算机, 在执行自主飞行任务方面存在困难. 本文在导航和识别方面提出了一种在微型无人机上装配单目相机及激光传感器实现室内目标人物搜索的方案. 为在含有多房间的复杂室内环境中到达目标人物所在区域, 本文基于激光测距路径规划实现了微型无人机的“沿边”搜索, 完成了对室内主要区域的遍历式查找. 同时, 设计了一种基于地面站辅助云计算, 结合全卷积神经网络单目图像深度恢复与深度阈值图的路径规划方法, 在激光无法正常工作地完成导航任务并优化“沿边”飞行过程中冗余的搜索路径. 为检测并跟踪搜索过程中出现的目标人物, 本文采用YOLOv3目标检测算法对人脸、躯干、人手等多部位进行检测, 并采用剪枝化设计降低了其检测耗时. 为确保目标面部特征丢失时仍能对其进行持续跟踪, 本文提出了一种人脸识别与躯干检测切换式的跟踪方案, 同时能通过检测到的手部位置信息, 对目标的意图进行实时判断. 经实验比对, 本文的算法有效缩短了搜索过程的飞行距离, 提升了目标追踪的成功率.

关键词 微型无人机, 室内导航, 深度恢复, 深度阈值图, 目标识别跟踪

1 引言

随着无人机微型化技术的发展, 越来越多超小尺寸(10 cm以下)的无人机^[1]凭借其隐身性好、灵活多能、便于大批量部署的优点, 在环境监测、灾后救援、军事侦察、管道巡检等诸多领域表现出极大的发展前景和应用价值^[2-4], 微型无人机正越来越受到各国

军方和民用部门的重视. 然而微型无人机的尺寸限制, 也意味着其无法搭载体积或重量较大的高精度传感设备和高算力的运算设备. 如何利用有限的资源进行感知和计算、完成自主飞行任务, 成为了一个重要的研究课题. 而本文则针对微型无人机在室内环境下的目标人物搜索问题, 从导航避障和目标识别两个方面进行研究.

引用格式: 高天禹, 马雨婷, 韩成哲, 等. 基于激光测距与单目视觉的微型无人机室内目标人物搜索方法研究. 中国科学: 技术科学, 2020, 50: 971-982
Gao T Y, Ma Y T, Han C Z, et al. Applying a micro-UAV for searching an indoor target person using laser ranging and monocular vision (in Chinese). Sci Sin Tech, 2020, 50: 971-982, doi: [10.1360/SST-2020-0130](https://doi.org/10.1360/SST-2020-0130)

相较于一般的野外高空环境, 微型无人机在包含多个房间的室内环境中搜索目标人物时需要躲避多种复杂的障碍物, 这对无人机的导航避障能力要求较高. 现有的无人机室内导航避障方法通常基于即时定位与地图构建(simultaneous localization and mapping, SLAM)技术, 其中一类方法基于激光雷达、深度摄像头等高精度传感器建立精确地图^[5~7], 并在此基础上完成无人机的路径规划^[8,9]实现室内导航. 这种方法虽然精度高且较为可靠, 但严重依赖于传感设备, 难以在载荷较低的小型或微型无人机上应用. 另一类方法是基于单目相机的单目视觉SLAM, 该导航方法通过对多帧图像进行匹配, 解算无人机的姿态并重建周围环境. 其使用的单目相机体积较小, 可以安装在微型无人机上. 代表性的算法有VINS-Mono^[10]和LSD-SLAM^[11]等. 然而, 这类方法的建图与定位效果受画面质量因素的影响, 在性能较差的微型无人机机载相机上表现不佳. 除了上述两类基于SLAM技术的方法, 还有一类方法直接利用深度视觉信息, 通过搜寻深度图像中距离较远的区域进行避障^[12,13]. 由于不需进行建图, 该类算法的复杂度低、实时性好, 且不依赖历史数据. 但是该类方法仍然需要使用体积较大的深度传感器, 且只能进行局部避障, 无法进行长距离的导航. 本文针对微型无人机载荷较低的特性, 选择激光传感器与单目摄像机作为传感设备, 以激光测距“沿边”飞行作为搜索目标过程中的主要路径规划方法引导无人机对房屋中的各个房间进行搜索. 同时在云端利用基于全卷积神经网络的单目深度恢复的方式还原单目视觉图像中的深度信息, 并设计一种基于深度阈值图的导航方法, 在恢复的深度图中搜寻可通行的路径, 从而优化搜索线路并在激光传感器不能正常工作的环境中辅助避障. 相对于传统的导航方式, 本方案兼具传感器占用体积小、可靠性高的特点.

微型无人机对目标人物的搜索跟踪需要通过实时目标检测与识别算法实现, 而其中目标检测算法的速度对无人机的搜索实时性影响较大. 在目标检测领域, 近年来以各类神经网络算法为代表的深度学习方法成为了主流, 相较于需要进行手工特征设计的传统算法, 这类算法通过梯度下降等优化方法自动学习图像中的特征, 并对网络参数进行优化, 其可移植性较强, 检测准确率提升显著. 目前, 深度学习方法主要有两类: 一类是R-CNN^[14]、Faster R-CNN^[15]等区域提名算法, 其

在检测前先进行候选区域的提取, 检测准确率较高. 但其候选框数量众多, 导致运算量庞大, 在机载处理器上无法满足检测实时性要求. 另一类是以YOLO^[16]、SSD^[17]、RetinaNet^[18]等算法为代表的端到端方法, 其特点是通过回归机制, 一次产生既定坐标, 在检测准确率略为下降的前提下显著提高了检测速度. 在这一类算法中, YOLO算法因检测速度快、泛化性强等特点, 成为了工程中最常见的算法之一, 其经过三代版本^[16,19,20]的演变后, 在检测精度与小目标检测上有了明显的效果提升. 本文是在第三代YOLO算法(YOLOv3^[20])的基础上, 通过稀疏化训练及网络剪枝等方法降低其计算资源消耗与检测用时, 实现了在机载处理器上对人脸、躯干、人手等目标的实时检测. 此外, 对于特定目标人物的识别与跟踪问题, 传统的视觉方法一般采用DeepFace^[21]、FaceNet^[22]等人脸识别算法, 通过持续捕捉目标面部特征来实现. 但对于微型无人机的目标跟踪任务, 目标的运动、躲避及无人机的姿态调整都会造成人脸识别有效帧的降低, 无法实现对目标人物的持续跟踪. 为此, 本文提出了一种人体多部位检测与人脸识别相结合的目标识别算法, 该算法在人脸信息丢失时切换至对相应人体的追踪, 在人脸信息重现后再切换回对目标人脸的跟踪, 通过这种切换策略增强了微型无人机在人脸识别环境较差时对目标的持续跟踪能力, 并通过检测到的人的手部位置信息, 对目标的意图进行实时判断.

2 微型无人机室内路径规划

考虑到微型无人机尺寸小、载荷低的特点, 本文采用激光测距与单目视觉相结合的路径规划方法, 完成了在含有多个房间的复杂室内环境中对目标人物的搜索任务.

对于激光传感器, 本文设计了“沿边”搜索方案, 在无人机本地实时处理激光数据, 实现了对室内主要区域的遍历式查找. 与此同时, 云端利用单目相机数据, 结合全卷积神经网络进行图像深度恢复并使用深度阈值图规划路径, 给出了另一种实时性稍差但更具可靠性的搜索策略.

在无人机实际飞行过程中, 云端搜索策略将作为本地“沿边”搜索策略的补充, 即在激光传感器无法正常工作或“沿边”搜索路径可优化两种情况下, 无人机

遵从云端规划的指令;而在其他情况下,无人机按本地路径规划结果进行导航.

2.1 基于激光测距的本地路径规划

为满足本地路径规划实时性的需求,本文选用基于飞行时间法的激光测距传感器.考虑到无人机平台算力有限,本文设计了“沿边”搜索的方案来完成本地路径规划,即在初始化后,无人机快速贴近距离较近的一侧墙面或障碍物表面,并沿墙面或障碍物表面进行搜索飞行.为降低噪声对激光传感器测量距离的影响,提高距离采样密度,在硬件设计中,共用到8个激光传感器,分布于无人机四周.

当无人机进行“沿边”搜索时,为保证墙面或障碍物表面始终在激光传感器探测范围内且无人机不与其碰撞,需要保持无人机与墙面或障碍物的距离恒定,无人机飞行方向与墙面或障碍物表面切向方向一致.以图1中无人机沿平滑墙面飞行的情景为例,飞行方向及离墙距离的控制方法如下.

将激光传感器测得的距离按顺时针方向依次记为 $d_a, d_b, d_c, \dots, d_h$, 并取每个旋翼下方传感器数据最小值作为此旋翼与墙面的距离,记为 d_1, d_2, d_3, d_4 , 表示为

$$\begin{cases} d_1 = \min\{d_a, d_b, d_c\}, \\ d_2 = d_d, \\ d_3 = d_e, \\ d_4 = \min\{d_f, d_g, d_h\}. \end{cases} \quad (1)$$

无人机与墙面的相对位姿信息可由机体与墙面的

图 1 (网络版彩图)无人机沿直线墙面飞行
Figure 1 (Color online) The micro-UAV flies along the straight wall.

距离 d 和夹角 θ 表示, 记为

$$d = (d_1 + d_2) / 2, \quad (2)$$

$$\tan\theta = \sqrt{2}(d_2 - d_1) / l, \quad (3)$$

式中, l 为无人机轴距.

为保证无人机不与墙面发生碰撞,定义其飞行时机体与墙面的最小安全距离为 $d_{\min}$, 最小安全距离与无人机速度有关.假设无人机飞行速度为 v , 飞行方向与墙面夹角为 ϑ ($0^\circ < \vartheta < 180^\circ$), 当无人机与墙面距离有减少趋势时,取 ϑ 为锐角,反之取为钝角,水平面内产生的最大加速度为 $a_{\max}$, 则

$$d_{\min} = \frac{(v \sin \vartheta)^2}{2a_{\max}} + \frac{l}{2}. \quad (4)$$

激光传感器探测距离为 $d_{\max}$, 则无人机“沿边”飞行时与墙面的期望距离 d_r 满足

$$d_{\min} < d_r < d_{\max}. \quad (5)$$

根据无人机机体与墙面的相对位置关系,对无人机姿态和位置进行控制.当无人机与墙体距离较小且有指向墙面的速度分量,即 $d < d_{\min}$, $\vartheta < 90^\circ$ 时,机体将先以最大加速度减速至垂直于墙面的速度分量为零,随后朝远离墙面方向运动直至 $d > d_{\min}$.而在其他时刻,将以 $d = d_r$, $\theta = 0$ 为期望,基于串级PID算法计算电机输出转速,对无人机进行伺服控制.

在微型无人机飞行过程中,所贴平面的走向会发生变化,无人机四周可能出现障碍物,需要通过调整偏航角完成伺服目标的转换.图2以无人机通过内侧拐角和外侧拐角两种典型场景为例,展示无人机的控制过程,而各种复杂环境都可看作是这两种典型场景的变换和组合.对于内侧拐角,如图2(a)所示,当无人机尚未到达内侧拐点时,右前旋翼与墙面的距离由右侧传感器测得,即 $d_1 = d_c$.随着无人机机头与前方墙面的距离减小,前侧传感器测得的数值 $d_a < d_c$,无人机将减速并逐步调整偏航角,同时随着偏航角的改变,右前旋翼的距离又由 $d_1 = d_a$ 过渡至 $d_1 = d_b$ 再到 $d_1 = d_c$,最后无人机完成转弯,沿第二道墙面继续飞行.而当无人机遇到外侧拐角时,如图2(b)所示,前侧传感器数值会发生突变, d_1 增大且 d_1 远大于 d_2 ,无人机贴近墙角的同时航向角发生改变,直至完成转弯,沿第二道墙面继续飞行.

图2 (网络版彩图)无人机沿边飞行的典型场景。(a) 内侧拐角; (b) 外侧拐角

Figure 2 (Color online) Typical scenes of micro-UAV flying along the edge. (a) Inside corner; (b) outside corner.

2.2 基于单目视觉的云端路径规划

本文使用单目摄像头获取RGB图像, 并通过图传模块传输给云端进行数据处理。云端对接收到的图像采用全卷积神经网络进行深度恢复, 再通过本文提出的深度阈值图规划路线, 完成基于单目视觉的云端路径规划。

2.2.1 基于全卷积神经网络的图像深度恢复

为从微型无人机传回的单目图像中恢复深度信息, 本文采用Laina等人^[23]提出的一种基于全卷积层的深度恢复网络架构。该网络使用卷积层代替传统的全连接输出层, 实现了高分辨率的深度图像生成。具体结构见文献^[23]。

该网络从输入到输出可分为下采样压缩提取信息与上采样扩展输出尺寸两个部分。其中下采样部分使用去除全连接层的ResNet-50^[24]架构, 通过捷径(short-cut connections)结构实现较深的网络层数, 在更好提取输入图片深层特征信息的同时, 扩大神经元的感受野, 方便后续深度信息的生成。为了获得较高分辨率的深度图像, 网络的上采样部分通过将前一层神经元的输出映射到后一层上的多个神经元输入来扩张网络的空间大小。最终生成了 162×128 像素的深度图像。

使用Zed mini双目相机在室内不同高度采集样本制作训练集进行训练。最终的深度恢复效果如图3所示。

图3 (网络版彩图)深度恢复效果。(a) 输入图像; (b) 深度恢复
Figure 3 (Color online) Depth recovery renderings. (a) The input image; (b) depth recovery.

示。训练好的网络较好地还原了输入图像中的深度信息。

2.2.2 基于深度阈值图的云端导航

通过深度阈值图在全卷积网络深度恢复图像上搜寻所有不发生碰撞的飞行路径, 实现路径规划, 其原理及构建方法如下。

对微型无人机及相机进行建模。假设微型无人机体积小, 可认为无人机几何中心与摄像机光心重合。以光心为原点 O_c , 垂直于无人机升力平面向上为 Y_c 轴, 无人机正向为 Z_c 轴, 垂直 Y_cZ_c 平面为 X_c 轴构建相机坐标系 $O_c-X_cY_cZ_c$, 然后以 Z 轴正半轴上与相机坐标系原点 O_c 的距离为焦距 f 处为原点 O , 平行于世界坐标系 X_c 和 Y_c 轴方向建立像平面坐标系 $O-XY$, 如图4所示。

无人机在飞行过程中靠近地面、天花板等障碍物时会受到强气流的干扰, 因而需要比机体轮廓更大的空间才能完成正常的飞行。为了描述这一现象, 垂直飞行方向在地面上的投影进行切片, 形成一平面。然后在该平面内以飞行过程中光心穿过平面的位置 O' 为中心, 构造一个顶点为 $ABCD$ 的矩形区域 R , 称为安全飞行区, 其长 L 、宽 W 可由相应空气动力学公式或实验给出。其中将安全飞行区设置为矩形是出于简化计算考虑, 也可设置为其形状。当无人机以某一时刻光心所在位置为起点, 向某一方向飞行时其安全飞行区将构成一个连续的飞行包络面, 当其中没有障碍物时认为无人机沿此方向飞行过程中不受到障碍物影响。

深度恢复图像中的任意一像素点对应一条从光心射出的穿过成像平面的射线(图5), 其必与包络面交于一点 N 。若深度恢复图上该像素点深度值小于 N , 则说明有障碍物在包络面内, 反之则说明没有障碍物, 即飞行包络面内没有障碍物的充要条件是: 深度图像中每一点的深度值都大于其与包络面交点间的距离 ON 。

图 4 (网络版彩图)相机建模及安全飞行区

Figure 4 (Color online) Camera modeling and safe flight area.

图 5 (网络版彩图)深度阈值图的生成

Figure 5 (Color online) Generation of depth threshold graph.

通过计算上述每一个像素点对应的阈值深度 ON , 可以形成一张深度阈值图. 该图可由下列方法近似得到.

设无人机以向量 (a, b, c) 为飞行方向时, 点 O' 的坐标 (x, y, z) 在世界坐标系下的方程为

$$\frac{x}{a} = \frac{y}{b} = \frac{z}{c}. \quad (6)$$

当 z 取不同值时, O' 点的 X_c 和 Y_c 轴坐标为

$$\begin{cases} x = \frac{a}{c}z, \\ y = \frac{b}{c}z. \end{cases} \quad (7)$$

代入安全飞行区长宽 L 和 W 求得矩形上的一个顶点 A 的坐标为

$$A\left(\frac{a}{c}z + \frac{Lc}{2\sqrt{a^2+c^2}}, \frac{b}{c}z + \frac{W}{2}, z - \frac{La}{2\sqrt{a^2+c^2}}\right). \quad (8)$$

令

$$k_1 = \frac{2\sqrt{a^2+c^2}f}{2z\sqrt{a^2+c^2}-La}, k_2 = \frac{2\sqrt{a^2+c^2}f}{2z\sqrt{a^2+c^2}+La}, \quad (9)$$

可求得 OA 与像平面的交点 A' 的坐标为

$$A'\left(\frac{ak_1}{c}z + \frac{Lck_1}{2\sqrt{a^2+c^2}}, \frac{bk_1}{c}z + \frac{Hk_1}{2}, f\right). \quad (10)$$

同理, 可求得其他三个像平面交点 B' 、 C' 、 D' 的坐标:

$$B'\left(\frac{ak_2}{c}z - \frac{Lck_2}{2\sqrt{a^2+c^2}}, \frac{bk_2}{c}z + \frac{Hk_2}{2}, f\right); \quad (11)$$

$$C'\left(\frac{ak_2}{c}z - \frac{Lck_2}{2\sqrt{a^2+c^2}}, \frac{bk_2}{c}z - \frac{Hk_2}{2}, f\right); \quad (12)$$

$$D'\left(\frac{ak_1}{c}z + \frac{Lck_1}{2\sqrt{a^2+c^2}}, \frac{bk_1}{c}z - \frac{Hk_1}{2}, f\right). \quad (13)$$

深度阈值图中四顶点连线上各像素点与安全飞行区的边对应, 考虑到当无人机检测 0.5 m 以上距离的障碍物时, 光心与安全飞行区中点连线 OO' 的长度 L_O 满足:

$$L_O \gg L > W. \quad (14)$$

可以近似认为 R 上任意一点 q 与光心 O_c 的距离及深度图像上像素点的深度 L_q 与 L_O 相等. 这样就求出了阈值图上一个四边形上各点的阈值. 连续改变 z 值直到图上各点都有对应的深度, 最终生成沿某方向飞行的深度阈值图, 如图6所示.

由于在室内飞行时前向范围中总是存在墙体或障碍物, 为了避免无人机频繁的采取避障动作, 可根据需求对 z 进行限幅, 仅对某一距离范围内的障碍物进行避障.

将某一方向上的深度阈值图与深度恢复图进行比对, 即可确定沿某一方向前进是否会发生碰撞以及碰撞将会发生的区域, 在预知保持当前飞行方向可能发生碰撞时, 微型无人机将搜索其他方向上的深度阈值图, 找到新的不发生碰撞的前进方向, 并将新的期望方向作为底层控制的期望航向角 θ 输入, 实现无人机基于单目视觉的云端路径规划(图7).

图7(b)中红色部分为保持当前航向飞行 3 m 将发生撞击的区域, 绿色圆点为改变后的飞行方向. 整个避障过程在Gazebo上的仿真如图7(c)所示.

为了加快求解速度, 提升算法实时性, 可预先计算好深度阈值图以减少生成时间. 同时在生成深度阈值图时对 z 采用较大的步长, 减少深度阈值图中点的数量. 并对多个方向的阈值图共用的点进行同时比对减少重复操作.

图6 不同方向的深度阈值图. (a), (b) 0°和22°方向阈值图; (c), (d) 0°和22°方向等值线

Figure 6 Depth threshold graphs in different directions. (a), (b) Threshold graphs at 0° and 22°; (c), (d) contours at 0° and 22°.

图7 使用深度阈值图进行避障. (a) 深度恢复图; (b) 碰撞预测及避障; (c) Gazebo仿真避障过程

Figure 7 Obstacle avoidance based on depth threshold graph. (a) Depth recovery renderings; (b) collision prediction and obstacle avoidance; (c) simulation of obstacle avoidance on Gazebo.

采用不同步长的深度阈值图进行计算时的计算时间和因为过长步长导致错误检测到的可飞行方向的数量如图8所示. 可以发现, 当步长为3 cm时生成的深度阈值图可以在大大减少计算量的同时保证路径检测的正确率不会明显下降.

2.3 总体路径规划算法

本方案以激光传感器“沿边”飞行作为首要路径规划方式, 确保无人机能够对整个屋内环境进行无遗漏

图8 不同间隔的深度阈值图的计算速度和准确率

Figure 8 Calculation speed and accuracy of depth threshold graph at different intervals.

的遍历搜索, 同时依靠其较快的响应速度对可能出现的障碍物进行实时避障. 另一方面, 由于单目视觉云端导航更新频率低、可视距离远, 在方案中用于以下两种情况.

第一种情况, 在某些特殊情境下, 由于环境中存在反光材质或镂空结构, 可能导致激光传感器失效或数据剧烈变化. 此时, 为了保证微型无人机飞行的稳定性, 将依赖单目视觉搜寻避障路径.

选取阳台作为此类特殊环境对无人机路径规划效果进行测试(图9(a)). 由于场景含有玻璃和防护网, 激光传感器测得的数据无法反应与障碍物表面的真实距离且数据会剧烈跳变, 当无人机检测到此类剧烈变化时停止贴墙位置伺服并减速, 待停稳后转换为单目云端路径规划的方式继续前向搜索直到传感器的数据恢复正常.

第二种情况, 单目视觉判断能够优化飞行线路时. 无人机在飞行过程中可以通过云端路径规划跳过较为平整的大段墙面. 在这些墙面范围内不包含门或其他能够隐藏目标的障碍物, 因而梯度连续变化. 减少对这些区域的搜索能够大大提升整体速度.

在通过深度阈值图完成对不发生碰撞的飞行方向的搜索后, 对其中连续跨度最大的一组飞行方向上的深度依次求梯度 G , 直到梯度大于设定阈值 $G_{\max}$, 使用该点对应方向代替初始点方向继续搜索达到优化路线的目的, 分别对有门的墙体和无门的墙体进行搜索优化(图10). 无人机在不错过门的前提下尽可能缩短了

图 9 (网络版彩图)特殊情境下采用单目选取路径。(a) 场景模型; (b) 激光数据; (c) 单目图像; (d) 碰撞预测及避障

Figure 9 (Color online) Monocular vision is used to select the path under special circumstances. (a) Special scenario; (b) the laser data; (c) monocular image; (d) collision prediction and obstacle avoidance.

图 10 (网络版彩图)搜索路线优化

Figure 10 (Color online) Search path optimization.

飞行距离。

总体路径规划算法由感知信息获取、本地信息处理、云端信息处理和导航方式选取四部分组成。无人机实际飞行过程中,本地端以30 Hz的频率更新基于激光数据的导航结果,云端以15 Hz的频率更新基于单目图像的导航结果。在激光传感器无法正常工作或“沿边”搜索路径可优化两种情况下,无人机遵从云端规划的指令,而在其他情况下,无人机按本地路径规划结果进行导航。总体路径规划算法流程如图11所示。

在Gazebo仿真环境中构建包含多个房间的室内场景,并放置桌椅、书架、栅栏、梯子等常见元素,分别模拟无人机单独使用激光、单目视觉以及使用激光和

图 11 (网络版彩图)总体路径规划算法流程图

Figure 11 (Color online) Flow chart of overall path planning algorithm.

单目相结合时的导航效果。对比不同方法的导航性能,结果见表1。

在单独使用激光进行导航时,经过桌腿、梯子、栅栏区域时激光数据的跳变导致无人机不能及时发现障碍物完成减速,而排除这些因素后,无人机可以完成对整个区域的搜索任务,但无人机完成搜索任务的耗时较长。在单独使用单目视觉进行导航时,由于算法更倾向于对视深较大区域的探索,面对含多房间的室内环境,虽然能在较短时间内搜索房间大部分区域,但经常会遗漏某些小房间,无法实现全空间的搜索。而使用激光与单目相结合的导航方法,无人机在保证全空间搜索的前提下尽可能减少耗时、缩短路径。图12为其飞行路线,其中红色段表示无人机基于激光数据进行导航,黄色段表示无人机基于单目视觉进行导航。相比单独使用激光导航,其优化飞行线路的作用主要体现在无人机飞至拐角和小房间时。经统计,在这些区域中使用激光与单目相结合的导航算法比单独使用激光导航可平均节省48.78%的时间,缩短42.43%的飞行距离。

进一步在真实室内环境中基于微型无人机平台测试路径规划算法。测试平台为一轴距6.5 cm、重量40 g的微型四旋翼无人机,以STM32F405为主控芯片,搭载激光传感器、惯性测量单元、单目相机、无线模块等。其中,激光传感器采用V15311X,测量范围为

表 1 不同方法的导航性能

Table 1 Navigation performances of different methods

	平均耗时 (s)	平均路径 (m)	搜索覆盖率 (%)
仅使用激光导航	64.11	121.48	100
仅使用单目导航	24.25	43	53.1
使用激光与单目导航	50.25	98.72	100

图 12 基于激光与单目视觉的导航效果

Figure 12 Navigation results based on laser and monocular vision.

4~400 cm, 数据采集频率30 Hz, 通过I2C总线与主控芯片相连, 通过选择不同总线地址实现多激光数据的同步采集. 单目相机采用高清微型UVC相机, 图像分辨率为1280×720, 采集频率为30 Hz, 通过无线模块MT7628DN实现无人机端与云端图像和指令数据的实时交互.

通过记录无人机的侧向位移和偏航角, 验证算法对无人机的实际控制效果. 图13(a)为某次实际测试的场景, 在约7 s的飞行时间中, 无人机环绕障碍块搜索目标, 飞行路径示意图如图13(b)所示. 图13(c), (d)记录了在此过程中无人机侧向与障碍物表面的距离及无人机偏航角. 由起始点至终止点, 无人机经过三个转弯, 转角分别为180°、-90°、90°, 可以看出, 在连续转弯的过程中, 无人机偏航角的改变较为平缓, 无人机与障碍块间始终保持10 cm以上的安全距离. 多次实验验证, 上述算法能够在有效规划路径的同时, 保证无人机的飞行安全.

3 微型无人机实时目标搜索方案

在微型无人机应用路径规划算法持续进行区域搜

图 13 无人机实际测试效果. (a) 测试场景; (b) 飞行路径示意图; (c) 无人机与障碍物侧向距离; (d) 无人机偏航角

Figure 13 Actual experimental results of UAV. (a) The test environment; (b) flight path; (c) lateral distance from the obstacle; (d) yaw angle.

索的过程中, 需要通过单目相机判断视野中是否存在目标人物, 为此要建立一套具有检测、识别与跟踪功能的图像目标搜索方案. 而微型无人机的工作特点给这一方案提出了如下要求.

(1) 在微型无人机对目标进行搜索时, 检测算法的帧均耗时是影响其机动性的主要因素, 而检测算法的检测准确率直接决定了无人机对目标人物搜索的成功率. 然而微型无人机所搭载的处理器处理能力有限, 难以同时保证算法的实时性与准确率. 因而所设计的算法应能针对此类低算力设备, 较好地平衡检测速度与目标召回率.

(2) 在微型无人机对目标进行跟踪时, 目标人物可能会做出移动、躲避等动作, 且无人机也会进行自身的姿态调整, 使其难以持续、稳定地捕捉目标的面部特征. 为此要求无人机的目标检测与追踪算法具有较强的智能性, 能通过对目标的多处部位(如面部、躯干、手部等)进行综合检测与关系推断, 在视野中目标面部信息丢失后仍具有持续追踪能力.

基于以上两点考虑, 本文设计了基于剪枝YOLOv3的人体多部位检测算法以及基于人脸识别与躯干检测的切换式跟踪算法进行实时的目标搜索. 其特

点有以下三个方面.

(1) 通过稀疏化训练与网络剪枝的方法对YOLOv3网络框架进行优化, 在保障检测精度的同时显著提高其检测速度, 使其更适用于搭载低算力设备的微型无人机.

(2) 所设计的目标检测网络在检测人脸的同时检测躯干与人手, 在不额外增加检测耗时的基础上获得了更多的相关信息用于目标追踪.

(3) 当目标的面部信息在视野中丢失时, 检测是否有邻近的躯干以供持续跟踪, 通过人脸与躯干的切换式跟踪算法解决了目标人脸丢失导致的跟踪失败问题. 此外, 通过对目标手部位置的检测, 实现了对其动作意图的判断.

3.1 基于剪枝YOLOv3的人体多部位检测算法

为了对人脸、躯干、人手三类人体部位进行检测, 以供后续进行目标识别与跟踪, 在谷歌OpenImages V5数据集^[25]中等量选取了带有这三类物体标签的图像, 以保障训练过程中的样本均衡. 此外, 由于该数据集存在躯干标签不统一的问题, 即将躯干这一物体标注为person、man、woman、body中的一个或多个, 导致其标注质量较差. 为此, 通过标签整合、交并比计算等方法进行标签清洗, 解决数据集标注中的漏标问题. 最终获得了10000张带有标注的图片作为训练集.

在网络的训练过程中, 设置动态的学习率以同时保障训练速度与训练精度. 通过在1080ti显卡上进行50个epochs的训练后, 网络能够对面脸、躯干、人手三类目标进行检测, 将训练后的网络在GTX 960M显卡上进行测试, 对于输入尺寸为416×416的图片, 其平均精度均值(mAP值)为0.526, 网络前向推断时间为160 ms.

对于无人机实时目标检测, 160 ms的前向推断时间将严重影响无人机的目标跟踪性能, 因此为了降低深度学习算法的计算资源消耗, 采用了网络剪枝的方式将权重较低的冗余参数去掉, 以降低网络的复杂度.

在原始训练出的YOLOv3模型中, 权重参数分散程度较高, 对模型进行网络剪枝会较明显地改变网络最终输出结果. 为保证网络剪枝后的目标召回率不明显下降, 采用稀疏化训练的方式, 在卷积层后插入一个批标准化(batch normalization, BN)层, 通过其归一化激活值的作用诱导模型中许多权重值向趋于0的方向进行训练. 通过这种训练方式, 降低后续剪枝带来的网

络性能影响.

在进行模型稀疏化训练后, 网络中的大部分参数都趋于0. 随后剪掉网络中较小的85%的权重所连接的网络通道. 实际测试中, 在0~1的权重范围内, 只有小于0.007的那一部分权重被剪掉了, 因此与剪枝前的网络模型对比, 剪枝后的网络模型的性能几乎没有受到影响, 而网络通道数量由13376个下降到了2007个.

最终对剪枝后的模型再次进行微调, 以达到其最佳检测性能, 原版模型、剪枝后模型及微调后模型的性能参数见表2.

由表2可知, 通过剪枝及模型微调, 模型的通道数下降到了原来的15%, 前向推断耗时减小了一半, 而模型在各类检测物体上的平均精确率与原模型相当, 实现了在不损失检测精度的前提下对原始模型的速度优化.

3.2 基于人脸识别与躯干检测的切换式跟踪算法

为确定无人机图像中是否出现了目标人物, 需要通过人脸识别技术, 将图像内人脸与数据库中的目标人脸进行比对. 为此选取了目前分辨能力较强的FaceNet人脸识别算法. 该算法通过在训练过程中对网络参数持续优化, 使得相同人脸在空间内进行聚类, 从而将不同人脸区分开. 通过对学习率、batch大小等超参数进行调整, 并引入TensorFlow机器学习库中的tensorboard可视化工具进行训练的监视, 在经过50万次图片的训练后, 网络的识别精确率变化如图14所示.

最终构建的人脸识别算法检测精度达到0.953, 在GTX 960M的GPU上测试后, 160×160大小的人脸识别平均耗时为0.018 s, 能够持续、稳定地将目标人脸成

表2 原版模型、剪枝后模型及微调后模型的性能参数

Table 2 Performance parameters of original model, pruned model and fine tuned models

	原版模型	剪枝后模型	微调后模型
参数数量	61.53M	12.36M	12.36M
通道数	13376	2007	2007
前向推断耗时 (ms)	160	78	78
mAP	0.534	0.532	0.539
AP(face)	0.768	0.764	0.769
AP(hand)	0.405	0.403	0.411
AP(body)	0.428	0.429	0.437

图 14 (网络版彩图)进行50万张图片训练后的准确率变化曲线

Figure 14 (Color online) Variation curve of accuracy rate after training with 500000 pictures.

功地从50个目标人脸中区分开来, 满足了无人机目标识别的性能要求。

然而微型无人机在进行目标追踪时, 目标人物可能会做出遮挡面部、逃离等举动, 此时人脸目标在图像中丢失, 会导致无人机失去对目标的识别能力。为此, 我们设计了一种在人脸丢失情况下的切换式目标搜索方案: 无人机在捕捉到目标面部特征后, 会根据位置关系判断出图像中与人脸对应的躯干, 并对其位置信息进行实时更新。而当目标的人脸图像丢失时, 无人机会切换到对该躯干的持续跟踪模式, 避免无人机因为目标丢失, 而重新对整个室内空间进行搜索的情况出现。当人脸目标重新出现在图像中后, 无人机会切换回对目标人脸的跟踪, 这是因为通过人脸推断目标身份的方式更为可靠。通过这种切换式的搜索算法, 无人机实现了在高帧率图像下对目标的连贯性搜索与跟踪, 为无人机的机动性与抗干扰性提供了保证。此外, 通过对目标手部特征及其位置的检测, 无人机能够判断出目标求救、投降等行动意图, 并据此可以做出相应动作决策, 增强了无人机的自主性、智能性, 整体搜索效果示意图如图15所示。

在GTX 960M上对算法进行测试, 测试视频时长15 s, 分辨率为416×416, 视频内容为一段典型应用场景: 前7 s目标面部正对镜头, 后8 s目标遮挡面部并开始移动。在多种算法下统计目标被成功识别的帧数如图16所示。

通过对YOLOv3算法进行剪枝降低网络冗余度, 同时使用人体多目标检测算法辅助人脸识别算法对目标人物进行持续搜索, 既实现了较高的追踪帧率, 又不

图 15 (网络版彩图)目标切换式搜索算法示意。(a) 面部特征未被遮挡时的目标追踪;(b) 面部特征被遮挡时的目标追踪;(c) 对目标投降手势的识别

Figure 15 (Color online) Schematic diagram of target switching search algorithm. (a) Object tracking when facial features are not occluded; (b) object tracking when facial features are occluded; (c) recognition of surrender gesture of target.

图 16 (网络版彩图)多种算法的目标追踪效果对比测试

Figure 16 (Color online) Comparison test of target tracking effect of several algorithms.

会因目标的遮挡造成目标搜索的失败。证明了该算法在低算力、高机动性的微型无人机目标搜索这一作业环境下, 实现了算法实时性、搜索准确率与抗干扰能力兼具的追踪效果。

4 结论

本文针对微型无人机室内目标搜索任务提出了一种基于激光测距与单目视觉的解决方案。在微型无人机的室内导航方面, 通过激光测距“沿边”飞行作为搜索目标过程中的主要路径规划方法, 同时使用基于全卷积神经网络深度恢复结合深度阈值图进行辅助避障的方式, 解决了激光传感器状态异常时的导航问题, 并优化了微型无人机的飞行路线, 提高了搜索效率。相较于现有的无人机导航方案, 该屋内导航方案具有占用空间小、可靠性高的特点。在微型无人机的目标识别方面, 本文对原始YOLO网络进行稀疏化训练及网络剪枝, 使其参数数量压缩, 网络冗余度降低, 加快了网

络模型对人体多种部位的检测速度, 之后使用FaceNet人脸识别框架对检测到的目标进行身份识别, 在目标面部特征被遮挡时通过人体多目标检测的结果进行目标位置推断, 增强了无人机对目标追踪的鲁棒性. 最后, 经仿真和实验验证, 该算法缩短了搜索过程的飞

行距离, 提升了目标追踪的成功率, 具有较高的可行性. 未来工作中, 将进一步解决单目视觉深度恢复在与训练集相差较大的环境中的泛化问题, 并提高对目标人物意图的判断能力, 开展对人群中特异性目标搜索方法的研究.

参考文献

- 1 Kawamura Y, Souda S, Nishimoto S, et al. Clapping-wing micro air vehicle of insect size. In: Kato N, Kamimura S, eds. Bio-mechanisms of Swimming and Flying. Tokyo: Springer, 2008. 319–330
- 2 Bürkle A, Segor F, Kollmann M. Towards autonomous micro UAV swarms. *J Intell Robot Syst*, 2011, 61: 339–353
- 3 Longhi M, Marrocco G. Flying sensors: Merging nano-UAV with radiofrequency identification. In: 2017 IEEE International Conference on RFID Technology & Application. Piscataway: IEEE, 2017. 164–168
- 4 Fregene K. Unmanned aerial vehicles and control: Lockheed martin advanced technology laboratories. *IEEE Contr Syst Mag*, 2012, 32: 32–34
- 5 Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D LIDAR SLAM. In: 2016 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2016. 1271–1278
- 6 Santos J M, Portugal D, Rocha R P. An evaluation of 2D SLAM techniques available in robot operating system. In: 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics. Piscataway: IEEE, 2013. 1–6
- 7 Cole D M, Newman P M. Using laser range data for 3D SLAM in outdoor environments. In: Proceedings 2006 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2006. 1556–1563
- 8 Goldberg A V, Harrelson C. Computing the shortest path: A search meets graph theory. In: Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Philadelphia, 2005. 156–165
- 9 Trivedi M H, Rush A J, Wisniewski S R, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *Am J Psychiat*, 2006, 163: 28–40
- 10 Qin T, Li P, Shen S. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans Robot*, 2018, 34: 1004–1020
- 11 Engel J, Thomas S, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM. In: Fleet D, Pajdla T, Schiele B, et al., eds. European Conference on Computer Vision. Switzerland: Springer, Cham, 2014. 834–849
- 12 Zhang Y, Jiang X, Luo Y, et al. Dynamic obstacle avoidance algorithm for mobilerobots based on depth image (in Chinese). *Control Eng China*, 2013, 4: 78–81, 90 [张毅, 蒋翔, 罗元, 等. 基于深度图像的移动机器人动态避障算法. *控制工程*, 2013, 4: 78–81, 90]
- 13 Yang W, Zhu W Q, Zhang C L. A fast autonomous obstacle avoidance algorithm based on RGB-D camera (in Chinese). *J Hunan Univ Technol*, 2015, 29: 78–83 [杨维, 朱文球, 张长隆. 基于RGB-D相机的无人机快速自主避障. *湖南工业大学学报*, 2015, 29: 78–83]
- 14 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014. 580–587
- 15 Sang J, Guo P, Xiang Z, et al. Vehicle detection based on faster-RCNN. *J Chongqing Univ*, 2017, 40: 32–36
- 16 Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 779–788
- 17 Liu W, Anguelov D, Erhan D, et al. SSd: Single shot multibox detector. In: Leibe B, Matas J, Sebe N, et al., eds. European Conference on Computer Vision. Cham: Springer, 2016. 21–37
- 18 Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017. 2980–2988
- 19 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017. 7263–7271
- 20 Redmon J, Farhadi A. YOLOv3: An incremental improvement, 2018. arXiv preprint, arxiv: 1804.02767
- 21 Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014. 1701–1708

- 22 Schroff F, Kalenichenko D, Philbin J, et al. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 815–823
- 23 Laina I, Rupprecht C, Belagiannis V, et al. Deeper depth prediction with fully convolutional residual networks. In: 2016 Fourth International Conference on 3D Vision. Piscataway: IEEE, 2016. 239–248
- 24 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 770–778
- 25 Kuznetsova A, Rom H, Alldrin N, et al. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale, 2018. arXiv preprint, arxiv: 1811.00982

Applying a micro-UAV for searching an indoor target person using laser ranging and monocular vision

GAO TianYu¹, MA YuTing¹, HAN ChengZhe¹, CAI Bo^{1,2}, WENG Rui¹ & ZHANG LiXian¹

¹ School of Astronautics, Harbin Institute of Technology, Harbin 150001, China;

² Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240, China

A micro-unmanned aerial vehicle (UAV) generally refers to an ultrasmall unmanned aerial vehicle <10 cm in size. Its load capacity is not high enough to carry high-performance sensors and an onboard computer. Therefore, performing autonomous flight missions using micro-UAVs is difficult. In this study, an indoor target person searching scheme is proposed for navigation and recognition by assembling a monocular camera and laser sensors on a micro UAV. To locate the target person in a complex indoor environment with multiple rooms, an “along the edge” searching strategy based on path planning using laser sensors is implemented for the micro-UAV to search the house. In addition, a path planning algorithm based on depth prediction using convolutional residual networks and a depth threshold graph is designed to complete the navigation task in case the lasers malfunction and to optimize the redundant search path during the “along the edge” flight. For the purpose of detecting and tracking the target person during the search process, the YOLOv3 object detection algorithm is used to detect the human body, face, and hands. The proposed method helps reduce detection time through sparse training and network pruning design. To ensure that a target can be continuously tracked when its facial features are lost, this study presents a switching tracking scheme for face recognition and body detection. Thus, the intention of the target can be estimated in real-time by collecting hand position information. Experimental results prove that the proposed algorithm effectively shortens the flight distance of a searching process and improves the success rate of target tracking.

micro-UAV, indoor navigation, depth prediction, depth threshold graph, target identification and tracking

doi: [10.1360/SST-2020-0130](https://doi.org/10.1360/SST-2020-0130)