

# 第十二届中国研究生电子设计竞赛

## 技术论文

论文题目：

面向野外搜救的智能无人机系统

The Intelligent Unmanned Aircraft System for Wilderness  
Search and Rescue

参赛单位： 武汉大学

队伍名称： 空中骑士队

指导老师： 杨文 张海剑

参赛队员： 王金旺 余淮 程文胜 何婉君 张恒

完成时间： 2017.04.30

## 摘要

我国幅员辽阔，野外灾害频发，搜救难度极大。通常来说，野外搜救就是和时间赛跑，分秒必争。搜救行动往往牵动大量人力物力。传统的地面救援方式效率低下，极易错过黄金搜救时间。近年来无人机 (Unmanned Ariel Vehicles, UAV) 的快速发展，使快速搜救成为可能。UAV 搭载图像采集设备以及各种传感器，可将空中的俯瞰视野实时传回地面。此外，UAV 敏捷，快速，并且可以以低操作成本执行难以由人类执行的动作。这些特点使得 UAV 很适合应用到搜救领域。目前，在搜救领域中，所用 UAV 大多只进行图像采集，其自主性没有得到很好的发挥。因此，本文研究了野外搜救无人机系统中的关键技术：自主避障、路径规划和自动返航；同时，结合计算机视觉中的目标检测识别技术，进行待救援人员的检测与定位。通过模块化的设计，旨在充分利用无人机平台资源的基础上，大幅度减少野外搜救耗时，为关键的生命救援争取时间。

在无人机目标搜索算法设计上，采用结合红外和光学图像的深度卷积神经网络检测方法。先使用阈值分割和形态学处理方法获取红外图像中的显著区域，再通过红外图像与光学图像配准，找到光学影像中的候选区域。基于候选区域，使用改进的深度卷积神经网络模型 SSD 算法进行实时目标检测。改进的 SSD 算法通过自建数据集进行模型训练。

关于无人机自主避障，采用了基于双目视觉的避障方案。主要是利用双目点云得到无人机前方障碍物的深度信息。然后根据障碍点云信息更新代价地图，进而不断更新避障路线，从而实现自主避障和路径规划。

关于自主移动降落，引入了增强现实技术中的视觉基准系统 AprilTag。通过编解码 AprilTag 标志，检测算法可以大大降低虚警，同时由于编码时引入容错机制，漏检率也保持在一个很低的水平。然后通过求解 PnP 问题，无人机能够自主动态调整其位姿，从而实现精准移动降落。在静止降落的基础上，利用分阶段处理，实现了降落在移动的平台 (如车顶，卡车后部) 上。

基于上述基础关键技术，本文创新性的实现了野外无人机搜救系统的原型搭建。系统采用模块化开发，模块之间耦合度低，可移植性强。配套的软件程序包括安卓应用程序以及基于 Qt5 的多线程图形界面程序。软硬件完整系统实验结果表明，实现的野外无人机搜救系统原型稳定性良好，可以完成搜救任务，为打造集成度更高功能更强效率更高的搜救系统提供了良好的基础。

**关键词：**野外搜救；无人机；目标检测；自主避障；移动降落

## ABSTRACT

Wilderness search and rescue (WiSAR) is very necessary and difficult for our country for its vast territory, frequent field disasters. In general, WiSAR is the race with time, every second counts. Search and rescue operations often need a lot of manpower and resources. The traditional rescue method is inefficient and can easily miss the gold search and rescue time. In recent years, the rapid development of Unmanned Ariel Vehicles (UAV) has made it possible for rapid search and rescue. UAV equipped with image acquisition equipment and a variety of sensors, transport the obtained video to the ground station. In addition, the UAV is agile, flexible, and can perform actions that are difficult to perform by humans. These features make UAV more suitable for WiSAR. However, currently UAV in WiSAR mainly uses the image acquisition module, cares little about the automation, the application scenarios are relatively limited. Therefore, the key technologies in the design of UAVs for WiSAR are studied, i.e., the autonomous obstacle avoidance, path planning and automatic mobile landing. Meanwhile, we utilize the target detection and recognition technology in computer vision to detect and locate survivors.

For target detection, we use the infrared images for assistance. Morphological image processing methods are used to obtain the salient region. The salient region is then converted into the target candidate region in the optical image by infrared and optical image registration. Based on the candidate region, the improved deep convolution neural network model SSD is used for real-time detection of target. And the model is trained with the built data set. With regard to autonomous obstacle avoidance, we adopt the obstacle avoidance scheme based on binocular vision. Mainly use the parallax principle to carry out motion estimation, so as to obtain the depth information of the obstacle in front of the UAV. And then update the cost map based on the obstacle cloud, and further update the obstacle avoidance route according to the cost map, so as to achieve the purpose of avoiding obstacles. The results of the obstacle avoidance experiments for static obstacle (wall, tree, etc.) and moving obstacle (pedestrian) show that the vision avoidance scheme based on binocular vision is more suitable for the realization of low cost UAV SAR system than LIDAR.

With regard to the autonomous mobile landing, we introduce the visual fiducial system AprilTag from the Augmentation Reality filed. By encoding an AprilTag, the detection algorithm can greatly reduce the false alarm, and because of the introduction of fault

tolerance mechanism, the miss rate is also maintained at a very low level. After detecting the mark, by solving the PnP problem, the UAV can dynamically adjust its pose in order to achieve accurate mobile landing. Experimental results show that the position deviation of landing is not more than 10 cm, which hopefully meets the requirement of real landing scenario where a UAV is likely to need to land on a platform with limited area, such as roof and truck rear.

Based on the above, we design the prototype construction and a supplementary demonstration system of WiSAR. The system adopts modular development strategy, leading to low coupling degree and high portability. The accompanying software applications includes an Android mobile application and a multi-threaded graphical interface program based on Qt5. Both applications respond quickly and interactively. The experimental results on the complete system of hardware and software show that the prototype of the WiSAR system functions well and the simulation search and rescue task can be successfully accomplished, which provides a good foundation for creating a WiSAR system with higher integration and higher efficiency.

**Key words:** Wilderness Search and Rescue; Unmanned Aircraft Vehicle; Object Detection; Autonomous Obstacle Avoidance; Mobile Landing

# 目录

<b>第 1 章 引言</b>	1
1.1 研究背景与意义 .....	1
1.2 难点与创新点 .....	1
1.3 本文研究工作与内容安排 .....	2
1.3.1 研究工作 .....	2
1.3.2 内容安排 .....	3
<b>第 2 章 系统需求与方案设计</b>	4
<b>第 3 章 系统硬件设计平台</b>	6
3.1 无人机平台 .....	6
3.2 机载计算设备 .....	6
3.3 无人机云台相机 .....	7
3.4 热成像仪 .....	7
3.5 双目视觉模块 .....	7
3.6 地面站 .....	8
<b>第 4 章 软件设计</b>	9
4.1 无人机地目标搜索 .....	9
4.1.1 红外图像目标检测 .....	9
4.1.2 光学图像目标检测 .....	11
4.2 无人机自主避障 .....	14
4.2.1 使用双目视觉获取深度信息 .....	14
4.2.2 使用代价地图 .....	15
4.3 无人机自主移动降落 .....	16
4.3.1 AprilTag 视觉基准系统 .....	17
4.3.2 AprilTag 检测 .....	17
4.3.3 AprilTag 识别 .....	18
4.3.4 相机三维姿态解算 .....	20
<b>第 5 章 系统测试</b>	22

5.1 红外图像候选区域提取实验结果 .....	22
5.2 光学目标检测结果 .....	24
5.3 自主避障实验结果 .....	26
5.4 移动降落实验结果 .....	27
5.5 配套演示系统 .....	29
<b>第 6 章 总结与展望</b>	<b>31</b>
<b>参考文献</b>	<b>33</b>

## 第1章 引言

### 1.1 研究背景与意义

我国幅员辽阔，随着人民生活水平日益提高，人们开始注重精神生活的丰富，年轻人更是勇于探索各处未知山野。然而由于各种因素，报纸新闻中也不乏“驴友”迷失，救援队伍耗费大量人力物力展开救援的报道。此外，一些特殊工作岗位人员，如野外考古学家，地质工作者等因为天气、通讯故障等原因也有可能发生人员失踪的事故。通常来说，野外迷失人员精神和生理压力骤升，体能消耗加速，尤其是遭遇自然灾害后，人体可能已经遭受损伤，因此野外搜救就是在和时间赛跑，分秒必争。但是阻碍救援因素却不少，恶劣的天气，通讯故障，交通不便等都是致命因素。搜救行动往往牵动大量人力物力。传统的救援方式是地面搜索，在某些特殊环境比如山林或者山谷，救援人员和设备运输就会受阻，导致搜救效率就十分低下。

如果能够从空中进行搜索，救援行动就会简便高速很多。随着近几年无人机(Unmanned Ariel Vehicles, UAV)行业的快速发展，这已经成为可能。无人机能够搭载图像采集设备以及各种传感器，可以将空中的俯瞰视野实时传回地面，同时，无人机行动敏捷，快速。基于以上考虑，通过多传感器数据融合实现无人机的一些自主行为(如自主避障，自动路径规划，自动返航等)，同时，利用机载计算设备，结合目标检测识别技术，进行待救援人员的检测与定位，能够大幅度减少野外搜救耗时间，为关键的生命救援争取时间，所以本系统具有很广的应用前景。

### 1.2 难点与创新点

为了达到野外搜救的目的，系统设计需要实现无人目标搜寻、路径规划、自主避障以及移动降落功能。针对上述四个功能需求，本系统实现有以下难点：

1. 野外搜救场景较为复杂，目标表观姿态较多，识别目标具有很大难度；
2. 目前尚无公开的无人机视角下人体数据库，需要自己建立无人机视角人体数据库；
3. 无人机航拍图像分辨率较高，无法实时对视频进行目标检测，同时，单独使用光学相机不能保证系统全天候运作；
4. 无人机自主避障技术尚无成熟的技术框架；
5. 为了保证搜救行动的效率，需要保证无人机能够自主降落在移动平台上。

针对上述难点的分析与研究，本文成功设计了一种面向野外搜救的智能无人机系统。其创新点在于：

1. 针对本系统中，野外搜救环境较为复杂的特点，使用目前综合检测性能最好的基于深度卷积神经网络的目标检测算法 SSD 算法作为基础算法，并针对本系统的应用场景进行了优化改进，提高了检测性能；
2. 建立了无人机视角人体数据库，便于对改进后的 SSD 算法进行重新训练，提高此应用场景下的检测性能；
3. 使用热成像仪和光学相机两路视频源进行目标检测，使用热成像仪得到目标的候选区域，再进行光学相机目标检测，既提高了搜索效率，又保证了本系统能全天候正常工作；
4. 针对无人机自主避障这一技术难题，使用双目相机进行障碍物位置测算，并通过合理的路径规划策略规避障碍物；
5. 针对无人机移动平台自主降落这一技术难题，使用 AprilTag 视觉基准系统辅助无人机进行自主降落。

### 1.3 本文研究工作与内容安排

本文致力于研究一套面向野外搜救的智能无人机系统。本文研究工作与内容安排如下：

#### 1.3.1 研究工作

本文主要研究工作如下：

- 1) 无人机目标搜索。采用双源视频流分析方法，天空端使用机载计算设备进行热成像仪红外图像处理，根据红外图像的特点得到目标候选区域。地面端通过高清图传接收器获取实时的光学相机视频流。考虑到无人机视角下人体表观变化巨大，传统的基于手工设计特征的目标检测方法效果不佳，因此基于目前最先进的基于深度学习的检测算法 SSD，并进行了相应的改进。此外，针对野外搜救这一应用场景，本系统构建了一个无人机视角下的人体数据集，并利用该数据集训练改进 SSD 算法模型。
- 2) 无人机自主避障。使用双目相机，根据双目视差原理获取障碍物的点云图，借助机器人操作系统 ROS (Robot Operating System) 下的导航工具包，实现了基于代价地图的无人机自主避障。

- 3) 无人机自主移动降落。本系统还实现了基于增强现实的基准视觉系统 AprilTag 的识别算法，并将其应用于无人机相对位姿获取。无人机根据得到的 6 个 DOF (Degree of Freedom) 坐标信息进行自身的姿态调整，从而实现精准移动降落。
- 4) 完整软硬件系统的搭建与测试。基于以上三个关键技术，本论文实现了面向野外搜救的智能无人机系统。系统通过大量实验调试来提高整体稳定性。此外，还编写了用于必要交互的移动应用程序以及方便地面工作人员可视化搜救结果的 PC 端图像界面程序。

### 1.3.2 内容安排

根据本论文工作的内容，本论文的各章节安排与主要内容如下：第一章主要介绍了无人机野外搜救系统的研究背景与意义，最后给出论文的主要工作以及整体结构；

第二章介绍系统需求与方案设计；

第三章介绍无人机搜救系统的硬件组成。包括无人机平台和各种使用到的传感器，并对各自的特点及参数进行介绍；

第四章介绍无人机搜救系统的软件设计。包括无人机地面对目标搜索的软件实现，基于双目视觉的无人机自主避障技术的软件实现以及基于视觉辅助的无人机自主移动降落技术的软件实现；

第五章介绍系统测试结果，包括软硬件系统各个功能模块以及完整系统的实验结果以及分析；

第六章进行总结与展望。总结工作的主要贡献和不足，并对下一步的研究方向进行讨论。

## 第2章 系统需求与方案设计

根据本系统的应用需求，系统需要具备目标搜索、自主避障、移动降落等功能。

根据模块化设计原则，本系统的模块划分如图2-1所示：

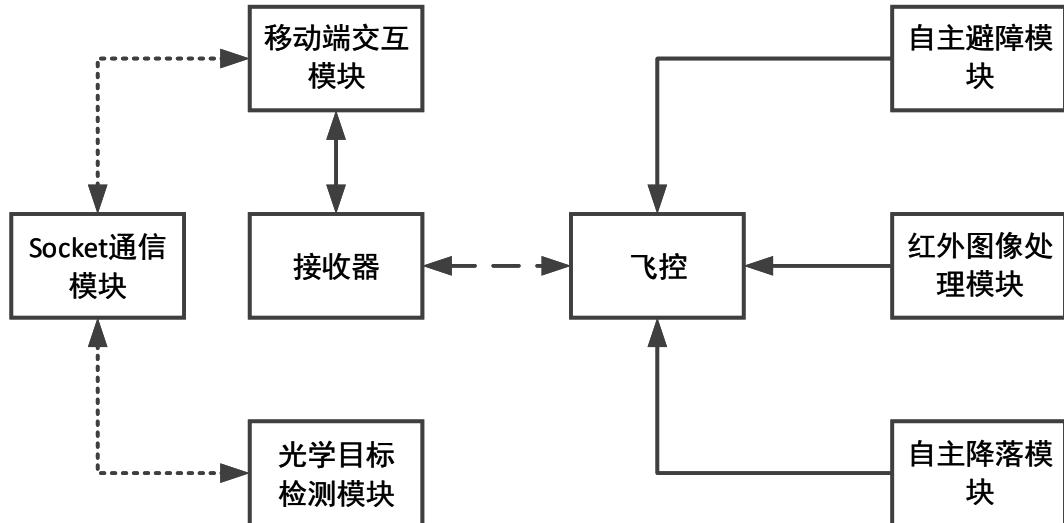


图 2-1 系统模块划分

由各个模块负责的功能以及数据输入输出接口，可以得到系统的工作数据链路图如图2-2所示。

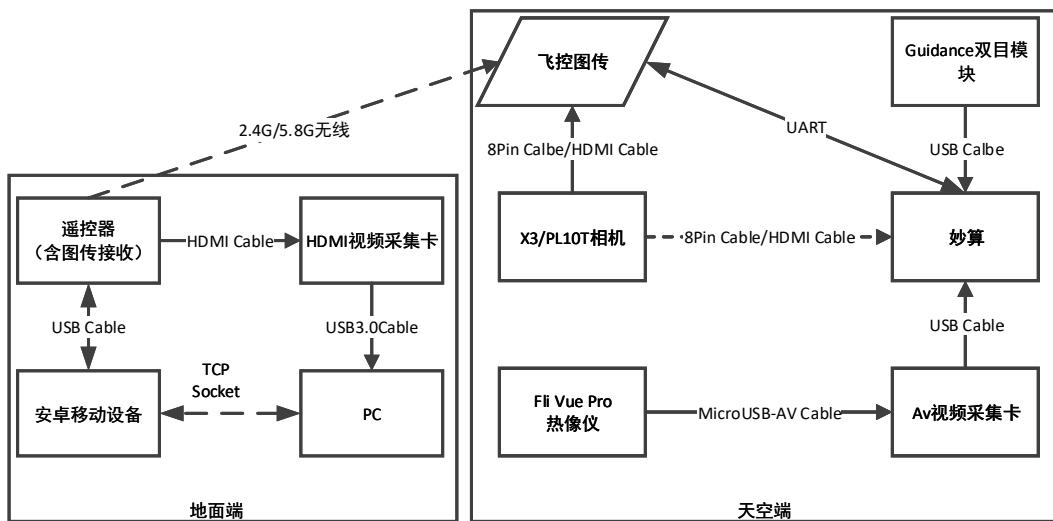


图 2-2 系统工作数据链路图

首先是无人机端。对于 DJI M100 无人机，云台相机的视频流通过 8Pin 电缆连接至无人机的高清图传的视频输入端。高清图传往地面站实时传送无人机视角画面。热成像仪 FLIR VUE Pro 的模拟视频接口通过 USB-AV 电缆传送至 USB 视

频采集卡并连接至机载计算设备妙算 Manifold 进行处理。

Guidance 避障模块的左右视图数据通过 USB 电缆连接至妙算 Manifold 进行分析处理。Manifold 则通过 UART 串口连接线和 DJI M100 飞控进行通信。

地面端的数据接收主要依靠 DJI M100 配套的遥控器。值得注意的是，由于现在的无人机平台基本上都会配备高清图传模块，地面端的图传接收器通常和遥控器集成在一起，且通常都会有相应的视频输出接口以方便外接显示，如 DJI M100 的配套遥控器。如果使用自己组装的无人机平台，遥控器和图传接收分离开来，只需要从图传的视频输入端口采集视频数据。通过一个 HDMI 高清视频采集卡，地面端的笔记本电脑可以采集到实时回传的视频画面并进行处理。

遥控器和移动设备通过 Mini-USB 线缆连接。此外，移动设备和笔记本电脑将被设置到同一个网络，组建局域网，两者之间通过 socket 进行通信。

## 第3章 系统硬件设计平台

### 3.1 无人机平台

由于国内的无人机制造技术已经相当成熟，而本论文关注点在于完整搜救系统的集成化，因此并没有必要从零开始自己组建无人机平台。本论文采用的无人机平台主要是大疆创新出品的 DJI M100 和 DJI M600。



图 3-1 DJI M100 和 M600

DJI M100 预留了两个 UART 接口，可供飞控和机载设备进行串口通信。通过 UART，机载设备可以获取无人机飞行状态信息以及接收来自地面的指令。同时，为了方便开发者进行开发，DJI 推出了专门的 Onboard SDK 开发工具包，用于提供底层的控制接口，让开发者可以专注于逻辑功能的实现。

对于本系统，机载设备完成数据处理任务，并通过 UART 接口将控制指令传至飞控使无人机执行相应的动作。

### 3.2 机载计算设备

机载计算设备采用的是 Manifold。Manifold 功能强大，功耗低，体积小，因此十分适合用作无人机机载计算设备。在本系统中，Manifold 将运行整个无人野外搜救系统的所有逻辑功能，主要包括双目视觉点云图获取、红外图像处理、自主避障以及自主移动降落。



图 3-2 Manifold 妙算及其接口

### 3.3 无人机云台相机

无人机云台相机主要作用是获取光学航拍图像数据，用于目标搜索任务。本系统使用的是 Zenmuse X3 相机，X3 相机可以向地面站提供 1080P 高清视频数据，图传距离可达 2,000m。由于相机拍摄的图像分辨率较高，使用机载计算设备 Manifold 对其进行图像处理比较吃力，因此光学图像处理将在地面端进行。



图 3-3 DJI Zenmuse X3 相机

### 3.4 热成像仪

本系统使用的是美国 FLIR 公司出品的 FLIR VUE Pro 热成像仪，其镜头大小为 19mm，可以提供  $640 \times 512$  分辨率的红外热成像图像，其适合用于小体积、远距离和温差较小的场景，尤其是在搜索、检测等需要快速识别目标的任务中。由于此热成像仪的分辨率较小，机载计算设备 Manifold 可以轻易地对其进行图像处理。

### 3.5 双目视觉模块

系统自主避障所使用到的双目视觉模块来自于大疆公司发布的 Guidance 套件。



图 3-4 Guidance 双目视觉模块

Guidance 的有效观测范围为 0.2m 到 30m，分辨率为  $640 \times 480$ ，能够满足本系统的自主避障的需求。

### 3.6 地面站

地面站主要包含遥控器，安卓手机和笔记本电脑。

遥控器是 DJI M100 配备的，其具有 HDMI 高清视频输出端口，可输出 1080P\_30FPS 高清视频。笔记本电脑通过 1080P 高速 HDMI 视频采集卡读取视频流进行图像数据处理，一方面实时显示回传的无人机第一视角画面，另一方面实现地面端的目标搜索任务，与空中端的红外目标搜索组成双视频流分析，提高目标搜索的准确度和速度。

笔记本电脑性能参数如下：处理器为 Intel I7-7700HQ，主频为 2.8GHz，操作系统为 Windows，内存为 8GB，显卡为 NVIDIA Geforce GTX 1050 显卡。

移动设备主要用于进行任务规划和显示无人机状态，任务规划包括设置搜索起始点，飞行速度，飞行高度等。显示无人机状况包括无人机当前飞行速度，飞行高度，GPS 位置信息以及电池电量信息等。

同时，在本系统中，移动设备还充当通信中介，即手机将天空端利用红外图像处理得到的目标候选区域位置转发给笔记本电脑；以及笔记本电脑根据目标搜索的结果发送相应的指令到天空端，比如变焦相机增大减小焦距，无人机悬停拍照等。

## 第4章 软件设计

### 4.1 无人机地目标搜索

在本系统中，目标搜索将从空中和地面两端同时进行数据处理。天空端使用热成像分析的方法，提取目标候选区域，然后再由地面端的光学图像处理程序进一步分析判别是否找到目标。这种红外和光学图像结合的分析方法可以有效提高搜索效率。

#### 4.1.1 红外图像目标检测

红外热成像技术是通过红外探测器来接收被测物体的红外辐射，然后由信号处理系统转变为目标的热图像的一种技术。本系统使用的 FLIR VUE Pro 红外热像仪可以对 100m 范围内的热能进行高分辨率的探测。

##### (1) 红外热像仪的成像特点

- 红外热像仪所成图像的每个像素的亮度代表该像素对应点的温度大小。
- 红外热像仪分辨率通常较低。在本系统中，一般情况下，目标在红外图像上只能显示模糊的轮廓，通常为  $30 \times 30$  像素左右大小。

基于这些特性，将通过以下手段进行红外图像处理，以提取目标候选包围框 (Bounding Box)。

##### (2) 红外图像预处理

FLIR VUE Pro 采集到的红外视频分辨率为  $640 \times 512$ 。由于本系统工作于野外环境，需要考虑强光对成像质量的影响，在强光条件下，成像画面如同蒙上了一层白雾，如图4-1所示。

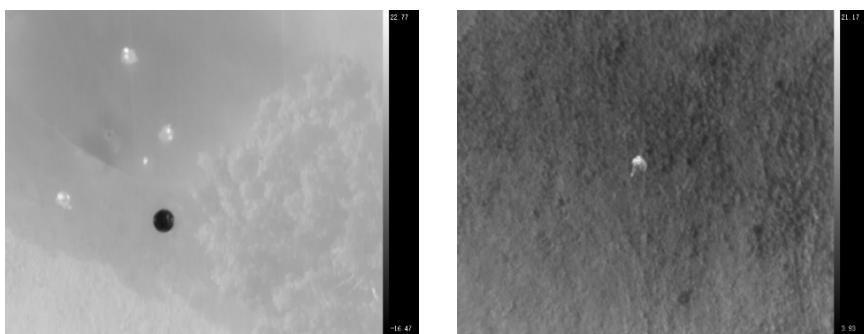


图 4-1 阳光照射下热成像(左);无阳光照射热成像(右)

经过研究后发现，这层白雾向外所辐射的红外能量服从均匀分布。因此，可以对图像进行均值平移，滤除这层白雾。

设图像 I 的长度为 W, 宽度为 H,  $P(x, y)$  为位置  $(x, y)$  的灰度,  $P'(x, y)$  为均值平移后  $(x, y)$  处的灰度值, 则

$$P'(x, y) = P(x, y) - \frac{1}{WH} \sum_{i,j \in W,H} P(i, j) \quad (4-1)$$

经过零均值化之后, 有些像素点的灰度值会变为负数, 因此需要进行归一化。这里采用灰度线性变换进行归一化。为了充分保留图像的细节, 将归一化后每一帧红外图像的灰度都归一化到 0 ~ 255 的范围内。经过这样的处理, 使得在不同的光照条件下, 红外图像的灰度值都是基本相同的, 为后面的进一步处理提供了方便。图4-2为上述处理的实验结果, 可以看出, 尽管光照强烈, 但是经过处理之后, 温度异常区域依然能够被有效的被提取出来。

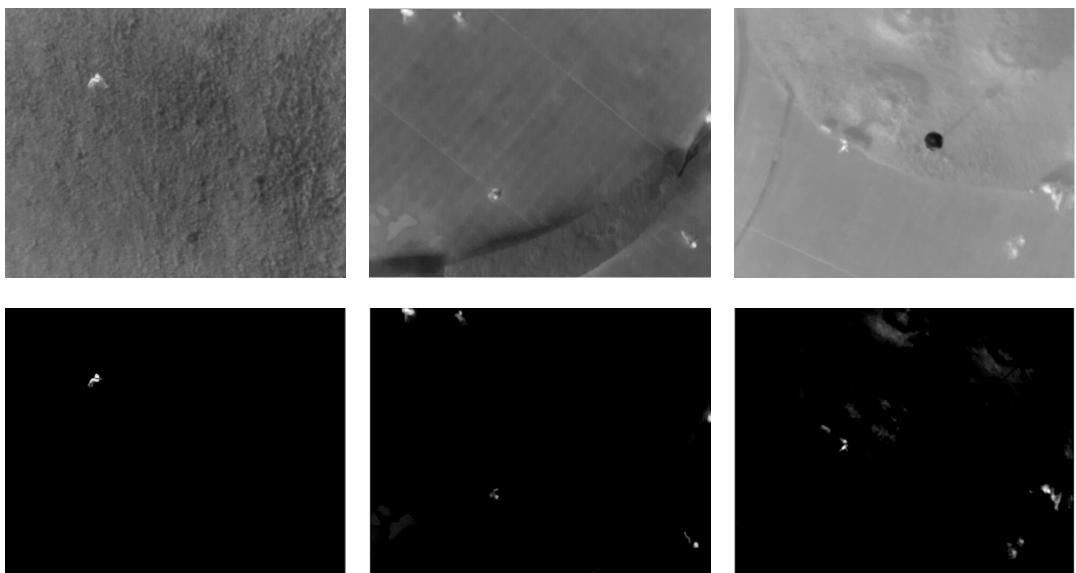


图 4-2 红外图像预处理结果(上为原图)

### (3) 目标候选区域提取

由前面步骤已经可以得到噪点比较少的提取结果。然而, 野外环境背景实在太复杂, 并没有哪种算法能够完全准确的搜寻到目标。但是, 可以通过先验知识, 比如在飞行高度为 20m 的时候, 目标的红外成像大小大约为  $30 \times 30$  个像素大小。过大或者过小的亮斑都极有可能是其他物体。因此通过判定连通域的面积大小, 可以进一步剔除虚警。无人机搜救系统工作时, 天空端的红外图像处理程序根据设置好的阈值, 如果经过上述图像分析处理得到的目标框个数大于 0, 则无人机向地面发送报告指令, 通知地面端采用光学检测手段进行光学搜索, 同时将得到的目标候选包围框的坐标发送回地面端的计算设备。考虑到实际情况, 即待搜救人员如果为多个, 也不会分布的很分散, 因此如果有多个目标, 则候选目标区域为多个目标的最小外接矩形区域, 如图4-3中的红色框所示区域。

地面端会在光学图像中选取对应的图像区域进行进一步的光学目标检测, 同

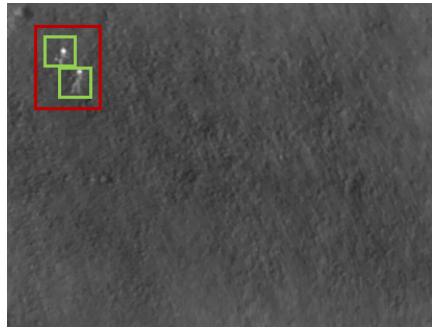


图 4-3 红外图像目标候选区域（红色）

时，对应的光学区域将被高亮显示，方便地面工作人员进行人工同步的筛选。

#### 4.1.2 光学图像目标检测

本系统面向的应用场景为野外搜救，野外环境复杂多变，同时，在无人机视角下，检测目标除了尺度会发生变化，形状，姿态也会发生较大的变化，如图 4-4 所示。经过大量相关实验，在这种场景下，传统的检测算法检测效果不是很理想，但是基于卷积神经网络 DCNN(Deep Convolutional Neural Network) 的检测算法能得到很好的检测结果。

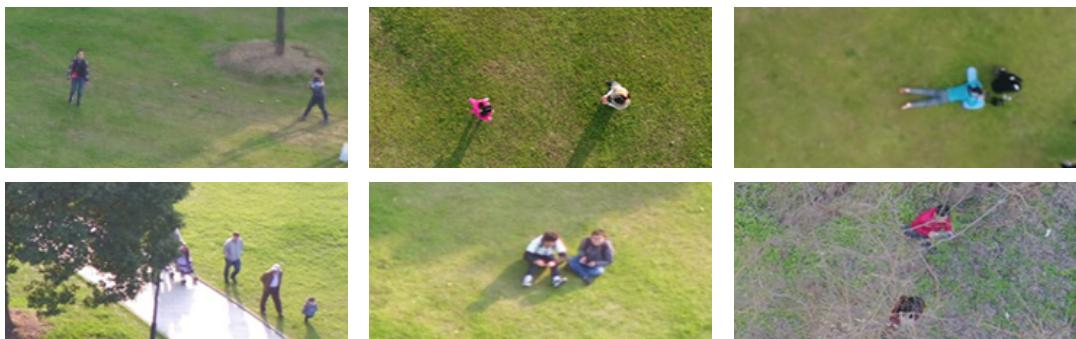


图 4-4 各种姿态下, 不同相机视角下以及遮挡情况下的目标示例样本

#### (1) UAV-PP 无人机数据集

无人机视角下的人会存在各种姿态，经调查，目前比较流行的无人机数据集和通用数据集比如 VOC、ImageNet 都没有专门收录无人机视角下的人体姿态数据。针对基于 DCNN 的检测算法，为了得到更好的检测结果，需要先建立一个无人机视角下人的数据集。

本数据集采取 VOC 数据集格式，所建立的数据集称为 (UAV perspective person dataset, UAV-PP)。UAV-PP 数据集文件目录如图4-5所示。

其中，JPEGImages 存放的是航拍图像。Annotation 文件夹中存放的是 JPEGImages 中每一张图片对应的.xml 格式标注文件，标注格式采用 VOC 格式，如图4-5右图所示。

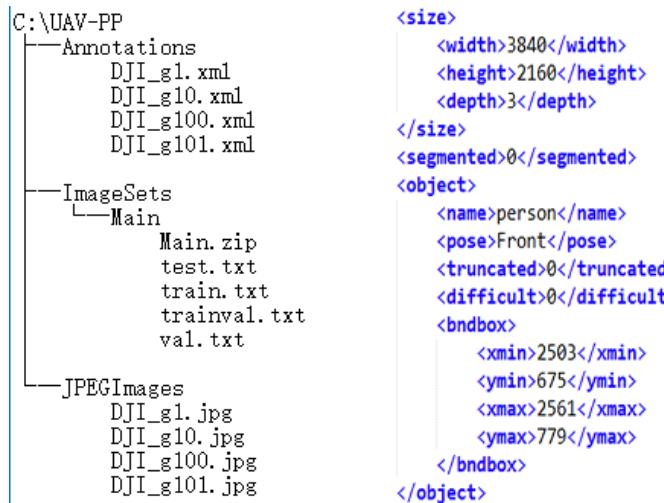


图 4-5 UAV-PP 目录结构 (左) 以及标注示例 (右)

## (2) SSD 检测算法

由于本系统需要实时地对无人机拍摄的图像进行处理，权衡检测速度和检测精度，基于 DCNN 的 SSD(Single Shot MultiBox Detector) 算法能满足本系统的需求。同时，为了更好的应用于本系统，我们使用上述自建数据集对原始 SSD 算法进行了微调，并应用于地面端进行光学图像处理。

SSD 算法组合了多个不同层级不同分辨率下的 feature map 的预测结果来对抗目标的尺度变化。每个对象类别在每个默认框中的存在并且产生调整尺度变化大的物体检测。对于  $300 \times 300$  的输入图像，SSD 可以在 VOC2007 数据集上以 59FPS 的速度达到 74.3 平均精度均值 (Mean Average Precision, mAP)，更精确版本使用  $512 \times 512$  大小的输入图像，可以达到 76.9 的 mAP(均使用 Nvidia TitanX 显卡)。

SSD 模型主要包含可以分为以下四个方面：

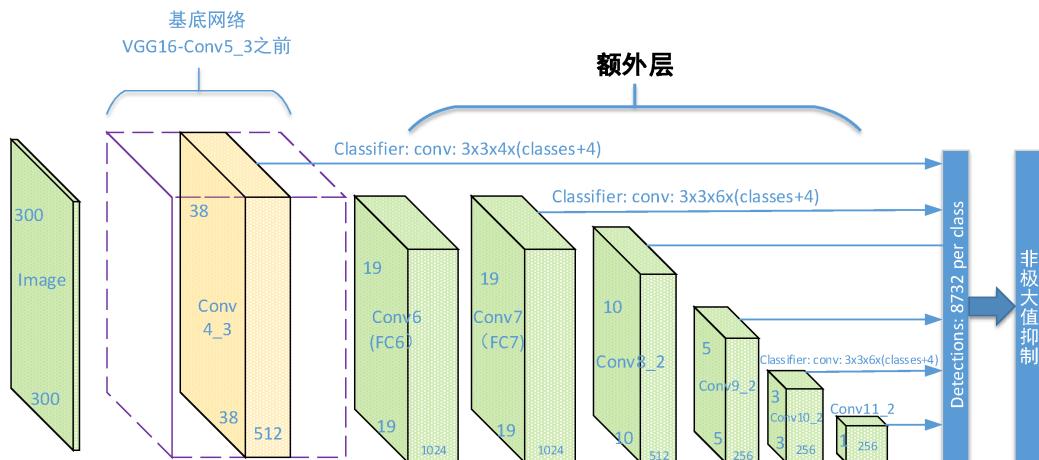


图 4-6 SSD 网络结构图

- 1) **基底网络。**截取当前流行的目标检测网络前面几个包含卷积核池化层用于做基底。
- 2) **多尺度特征图。**在基底网络的基础上，SSD 算法在后面连接了多个卷积特征层。这些层的大小依次递减，这样就可以允许进行多尺度的检测。
- 3) **卷积检测预测器。**在基底网络后添加的每一个层可以利用卷积滤波器来生成一系列固定大小的预测结果。对于一个大小为  $m \times n$  的 P 通道的特征层，使用的是  $3 \times 3$  大小的核 (kernel)，产生的为该位置归属于某一个类别的得分，亦或是相对于默认 bounding box 的形状偏移量。这样，在每一个  $m \times n$  大小的特征图位置上，kernel 都会生成一个值。其中，形状偏移量是相对 default bounding box 与此时特征图上的位置的相对距离。
- 4) **默认 bounding box 和长宽比。**如图4-7所示，在整个 SSD 网络的前面的多个特征图上，每一个特征图单元 (cell) 都会关联一系列默认的 bounding box。每一个默认 bounding box 和 cell 之间的位置是固定的。对于每一个 cell，给出它相对于 cell 中的默认 bounding box 的偏移量，以及这些默认框包含各类别物体的得分。具体来说，如果默认框为  $K$  个，则对于每个 bounding box，在每个位置，要产生  $C$  个类别的分类得分以及 4 个相对偏移量 (相对 bounding box 的 4 个顶点)，所以一共需要  $(C + 4) \cdot K$  个卷积核，对于  $m \times n$  大小的特征图输入，产生  $Kmn \cdot (C + 4)$  大小输出特征图。

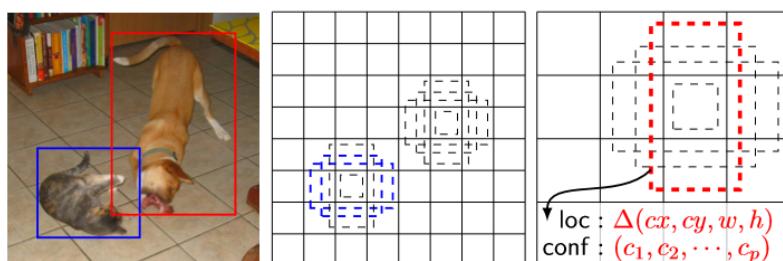


图 4-7 SSD 默认 boundingbox

SSD 框架虽然在 VOC、COCO 等数据集上获得了良好的表现，但是对于本系统的实际应用场景，仍需做进一步的调整与改进。针对自建数据集，SSD 需要做一些参数的调整，主要是一些训练参数，比如学习率，冲量的调整，还包括 Extra Network 网络结构的微调，比如使用的卷积核的大小个数，以及网络最后输出层的输出类别数目。由 SSD 模型的介绍可知，使用自建数据集实施检测任务的时候，可以从设计默认 bounding box 的分布入手。由于无人机视角的特殊性，目标的尺度变化不大，比较有限，因此，我们更改了尺度划分。此外，因为目标实际尺寸比较小，因此训练使用的 batch 大小从 128 适当增大到 512。

另一方面，原始的 SSD 的基底网络是使用 VGG16 前面的预测层，一个  $32 \times 32$  的目标经过 VGG 后变成  $2 \times 2$ ，这样就容易丢失后几层的语义信息。因此原始 SSD 对于小尺寸目标（如无人机视角下的人）的检测效果受到了影响。所以对于小目标检测，需要增加上下文的语义信息。经过研究，在神经网络中引入残差网络 (Residual Network, ResNet) 能够很好的保留目标语义信息。图4-8是 ResNet

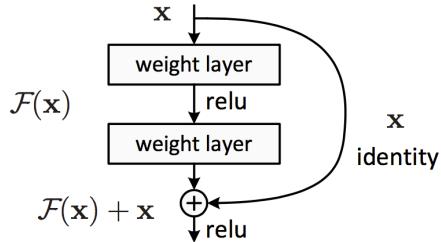


图 4-8 ResNet 基础结构

的基础结构。ResNet 在输出和输入之间引入一个 shortcut connection，这样可以解决网络由于很深出现语义丢失的问题，从而可以把网络做的很深。综合考虑，本论文将使用深度适中的 ResNet-101 结构作为 SSD 的基底网络。

## 4.2 无人机自主避障

野外搜救无人机系统的工作环境是复杂多变的，这对于无人机的飞行安全是一个很大的考验。因此实现野外搜救无人机系统需要关注这一关键技术。本系统使用能得到图像的深度信息的双目相机来实现避障技术。

### 4.2.1 使用双目视觉获取深度信息

使用双目相机的一个好处就是可以根据左右两个相机的视差计算出物体的距离信息。使用两个相机可以得到左右视图，两个视图必定产生差异，称之为“左右视差”，从左右视差 (disparity) 到“深度” (depth) 的转换原理如图4-9所示。

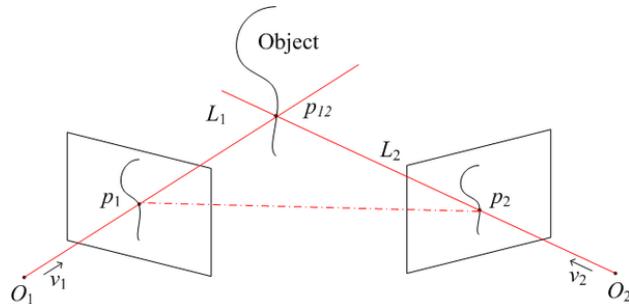


图 4-9 双目视图测距原理

设前方物体上的点  $p_{12}$  对应于左右视图上的点  $p_1$  和点  $p_2$ 。则  $p_1, p_2, p_{12}$  构成一

个三角形，通过求解这个三角形的各边长，顶点位置，就可以算出点  $p_{12}$  的坐标了，再进一步也就可以得到  $p_{12}$  的深度了。由双目视图可以计算差异图，差异图重投影到 3D 就可以得到点云，如图4-10所示。而点云数据恰好是对障碍信息一个最通用的描述。本系统采用的机器人操作系统 ROS 下，拥有的 `stero_image_pro` 节点，可以实现点云的获取。

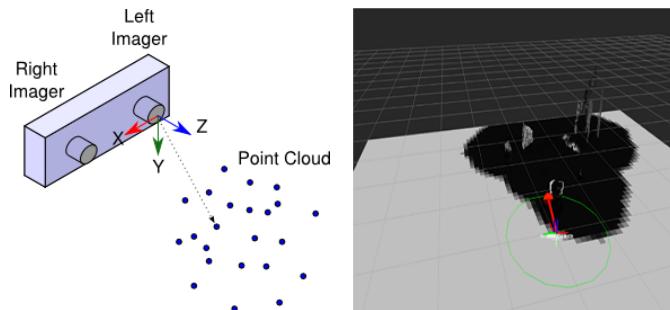


图 4-10 3D 点云 (左); Guidance 生成的点云可视化 (右)

图4-10(右)是本系统根据 Guidance 双目视觉模块生成的 3D 点云图，得到的 3D 点云可以让无人机了解前方一定距离内的障碍信息，为避障提供条件。

#### 4.2.2 使用代价地图

点云数据得到后，就可以得到障碍信息。首先，建立一个空的网格地图。在当前位置点，根据点云信息，有点云出现的位置设为 Occupied，没有出现的位置设为 Free，点云密度大于某个值的地方就认定为有固体障碍。通常使用基于代价地图的路径重规划躲避这些障碍。

代价地图分为全局代价地图和局部代价地图。由于本系统的应用场景是未知地图的巡航飞行，因此只能使用局部代价地图，这个是根据不断收到的传感器数据动态更新的地图。实际的避障中，首先需要初始化一个全局的规划器。这个全局规划器负责生成一个从指定的“起点”到指定的“目标点”之间的最短路径，这个生成的最短路径为最初状态的最佳路径。然后根据局部代价地图，局部规划器将根据最新的障碍信息不断调整更新最佳路径，并产生各个移动方向上的速度命令以便无人机避开障碍物。这些速度命令一般采用动态窗口方法来生成：

- 1) 在机器人控制空间  $(dx, dy, d\theta)$  中离散采样。
- 2) 对于每个采样速度，从机器人的当前状态执行正向仿真，以预测如果使用采样速度一段时间的运动会发生什么。
- 3) 使用包括诸如以下特征的度量来评估（打分）从正向模拟产生的每个轨迹：接近障碍物，接近目标，接近全局路径和速度。然后丢弃非法轨迹（与障碍物碰撞的轨迹）。

4) 选择最高得分的轨迹，并将相关速度发送到移动基地。

5) 清除得分并重复。

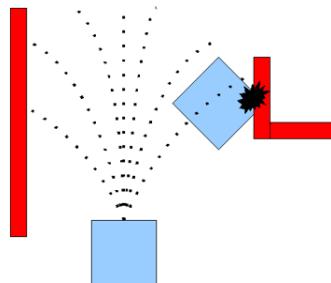


图 4-11 局部规划器对正向模拟路径打分

在机器人操作系统 ROS 中，有专门的导航工具集成这些算法，只需要设法提供它所订阅的几个主题信息，它就会调用内部的全局规划器和局部规划器来生成系列的运动速度指令。然后无人机就可以按照传来的运动速度指令进行障碍躲避了。

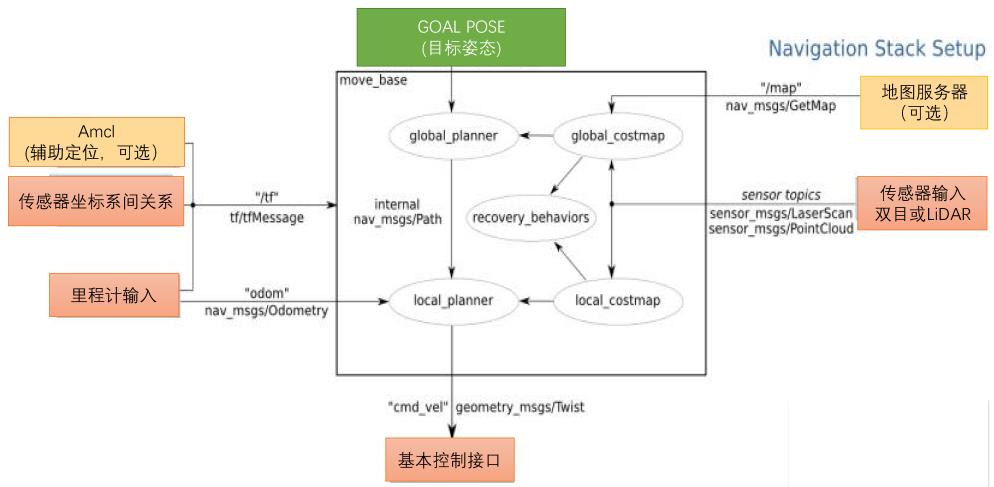


图 4-12 ROS 导航框图

### 4.3 无人机自主移动降落

本部分采用 AprilTag 视觉基准系统，视觉基准系统通常指的是易于识别和彼此区分的人为标志。通过检测 AprilTag，无人机将锁定降落位置点，并根据检测得到的 AprilTag 与无人机机身的相对 6DOF(six degree of freedom) 位姿来调整无人机的速度和姿态，最后实现精准移动降落。对于野外救援，并不能保证有平坦开阔的降落平台，这时可以在随行的汽车车顶上布置该视觉系统，引导无人机安全降落。

#### 4.3.1 AprilTag 视觉基准系统

本系统使用的 AprilTag 由两部分组成：标志检测器 (Tag Detector) 和编码系统 (Coding System)。

因为视觉基准系统的意义主要是提供可靠的基准检测测试，因此需要精心设计。为了提高检测效率，AprilTag 使用黑框将信息荷载包围住，以辅助精确定位。AprilTag 内部的信息荷载只使用黑白块表示，且无任何对称性。AprilTag 包含三个族，包括 36h11, 25h9, 16h5。本系统使用 25h9 族。

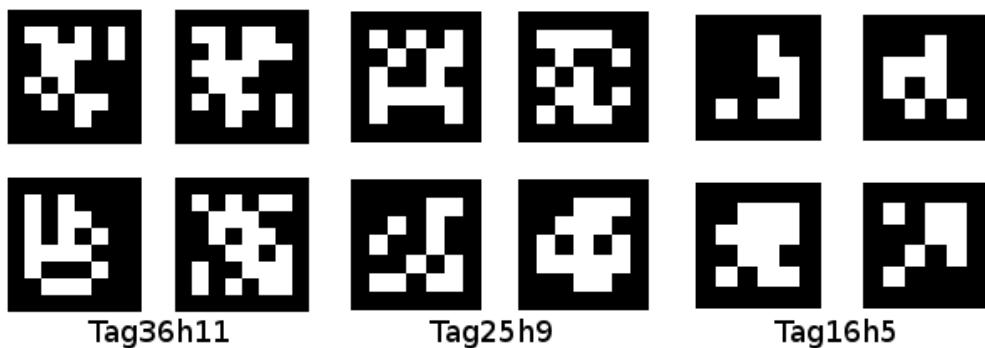


图 4-13 各类 AprilTag

#### 4.3.2 AprilTag 检测

AprilTag 检测的第一步就是四边形的检测，因为已经使用了外围一圈黑框作定位辅助，因此四边形检测可以由以下几个步骤组成：

- 1) 将图像转换为灰度图像。
- 2) 图像滤波。

本部分使用高斯核大小为  $3 \times 3$ ，标准差为  $\sigma = 0.8$  的高斯滤波器。

- 3) 将灰度图二值化。

由于外围黑框的存在，因此使用正向二值化即可。

正向二值化的表达式如下：

$$dst(x, y) = \begin{cases} \max Val & \text{if } src(x, y) > T; \\ 0 & \text{其他;} \end{cases} \quad (4-2)$$

其中  $\max Val$  为设定的规定值， $T$  为二值化的阈值。

关于二值化的阈值，本部分使用的是局部自适应阈值算法，该方法根据局部像素块的像素值分布来确定该像素位置的二值化阈值  $T$ ，从而使不同亮度、对比度以及纹理区域对应不同的二值化阈值，具有很好的鲁棒性。

#### 4) 轮廓检测。

本部分采用的是 OpenCV 中的 *findContours* 函数。将上述得到的二值图像作为输入，可以得到一幅图像中存在的所有轮廓。由于我们需要的仅仅是四边形的轮廓，因此需要对得到的轮廓进行多边形拟合，使用 OpenCV 中的 *approxPolyDP* 函数即可完成。该函数将对输入的一组轮廓进行多边形拟合，通过判断拟合得到的多边形的顶点个数是否为 4 就可以筛选出候选轮廓。

#### 5) 候选区域筛选

由于标志内部仍可能被检测出四边形轮廓，因此，需要判定得到的四边形是否为最外围的边界框。由于实际操作中，标志的尺寸是已知的，无人机飞行的高度和相机成像分辨率也是已知的，这样就可以得到标志占据画面的比例大小来进一步筛选。对于每个候选的拟合得到的四边形，计算其相邻顶点之间的距离：

$$dist_{ij} = (x_i - x_j)^2 + (y_i - y_j)^2 \quad (4-3)$$

其中， $i, j$  为相邻的两个顶点，若满足  $\min dist_{ij} > \varepsilon$  则保留该四边形， $\varepsilon$  为设定的阈值，此处为 5000。

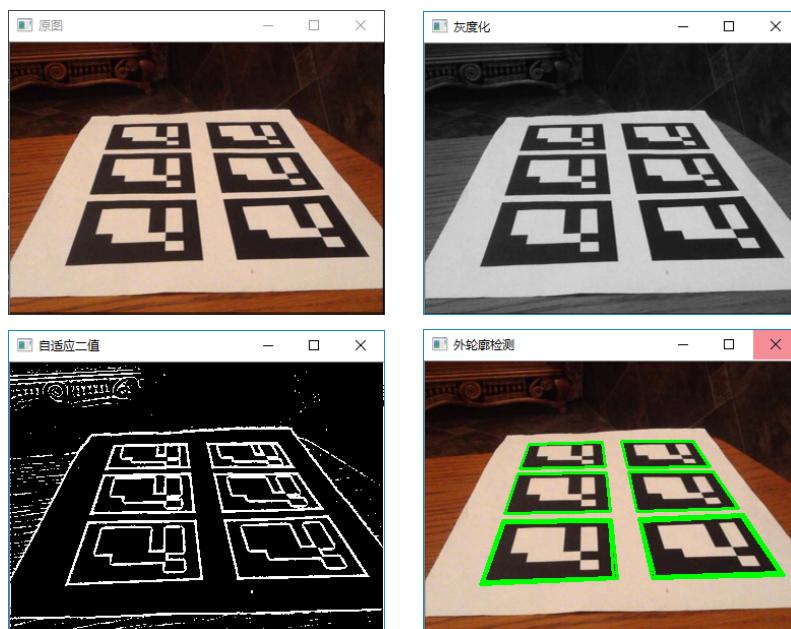


图 4-14 检测结果：原图(左上) 灰度图(右上) 二值化(左下) 外轮廓检测(右下)

#### 4.3.3 AprilTag 识别

AprilTag 系统中，每个标志可以划分为  $7 \times 7$  个细胞单元，黑色格子代表 0，白色格子代表 1。则标志内部可以由 5 个 5bit 的 01 字符串表示（除去外围的 2 个

格子)。接下来将采用海明码编码。5bit 中的 3bit 用于校验, 2bit 用于存放数据, 则每个标志可以表示  $4^5 = 1024$  个数据。引入编码系统, 是为了提高纠错能力以及进一步筛选虚警。

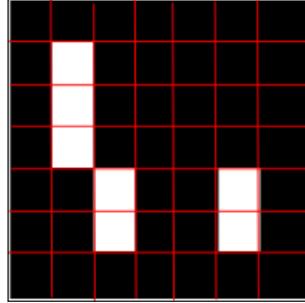


图 4-15 AprilTag 划分 cell 单元

对于海明编码, 如果待检测的二进制码为  $n$ , 若要求纠错能力, 则需设置  $k$  位校验码, 并且满足  $2^k n + k + 1$ , 即  $k$  个校验码需要涵盖  $n$  位数据码以及自身  $k$  位校验码任意一位出错或者全部正确 +1 的情况。因此, 对于 AprilTag 系统, 需要  $k = 3$ 。校验码设为  $c_1, c_2, c_3$ , 将其从左往右间插在第  $2^i$  的位置, 所以一个 AprilTag 可以编码表示为  $c_1c_2b_1c_3b_2$ ,  $b_1, b_2$  为信息码。

根据海明编码规则,  $c_1$  负责校验码字(包括信息码和校验码)的第 1, 3, 5 位;  $c_2$  负责校验码字的第 2, 3 位,  $c_3$  负责校验第 4, 5 位。若采用偶校验(即被校验的位加上自身一共有偶数个“1”), 则  $c_3 = b_2$ 。

有了海明编码, 在检测到候选目标区域四边形后, 就可以通过解码的方式来判定其是否为定义的标志了, 这样则大大降低了虚警, 同时由于有纠错能力, 也有着较好的鲁棒性。

根据海明编码, 下面将确定标志 ID。

由于 AprilTag 可以看出一个编码系统, 因此每个标志可以给它编定一个 ID 号码。主要有下面三步:

- 1) 根据得到的四边形顶点进行反透视变换, 将四边形还原为正方形。
- 2) 对候选区域使用大津阈值二值化, 并归一化大小, 比如  $100 \times 100$ 。
- 3) 解码识别标志。

将  $100 \text{ 像素} \times 100 \text{ 像素}$  划分为  $7 \times 7$  的 cell, 计算轮廓 cell(外围一圈)中非零像素个数, 若大于 cell 中的所有像素个数一半, 则认为是“白”格, 该轮廓被认为不完整, 舍弃之。

然后, 识别将剩余的  $5 \times 5$  个 cell 大小的编码区域并根据海明编码规则进行编码, 存入矩阵 bitMatrix。

由于得到的 bitMatrix 有可能是旋转后的结果，因此还需要找到具体旋转了多少角度。具体来说，因为每一行 5bit 码中有两个 bit 的信息码，因此共有 4 种编码可能（每一行）。对 bitMatrix 的每一个旋转状态（每旋转 90° 为一个状态），对 bitMatrix 中的每一行，寻找其与这四种可能编码的海明距离（即不同值个数）中最小值，四个最小值之和为该旋转状态下的海明距离。可以通过求四个状态下海明距离的最小值来得到标志的旋转状态。最后，一个标志的编码，和顶点顺序都确定下来了。

#### 4.3.4 相机三维姿态解算

为了引导无人机进行姿态调整，最后降落在标志上，在识别了标志之后，需要解算出标志相对于相机的姿态。在相机坐标系中的点  $P_c(x_c, y_c, z_c)$ 。可以由对应的世界坐标系统中的点  $P_w(x_w, y_w, z_w)$  通过一个旋转变换  $R(3 \times 3)$  和一个平移变换  $T(3 \times 1)$  得到：

$$P_c = RP_w + T \quad (4-4)$$

写成齐次式为：

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0_3^T & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (4-5)$$

根据透视投影模型：

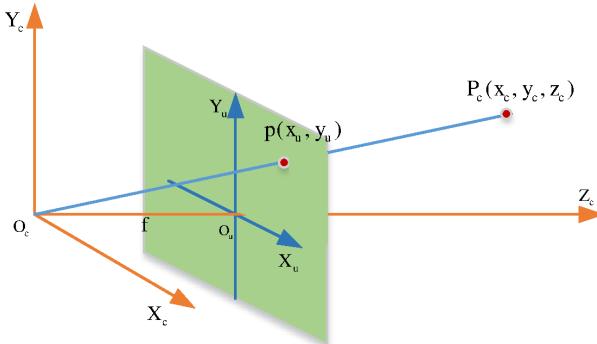


图 4-16 透视投影模型

$$x_u = f \frac{x_c}{z_c} \quad y_u = f \frac{y_c}{z_c} \quad (4-6)$$

$x_u, y_u$  为成像系统感光器件（图像的实际物理）坐标系。则公式4-5化为：

$$z_c \begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} \quad (4-7)$$

又因为相机物理尺寸坐标系和图像坐标系之间有如下关系：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} \quad (4-8)$$

其中  $d_x, d_y$  为图像中每个像素分别在  $u, v$  中上对应的实际物理尺寸，单位为  $m/pixel$ 。由此可以得到图像坐标系到世界坐标系的转换关系：

$$\begin{aligned} z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \begin{pmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \\ &= M_1 M_2 \vec{X}_w \end{aligned} \quad (4-9)$$

其中  $M_1$  称为相机的内参数，是每个相机出厂之时就已经决定好的。可以通过各种相机标定算法标定得到。而  $M_2$  就是需要的外参数，其包含了旋转矩阵和平移矩阵。 $M_1, M_2$  都得到之后，便可求出相机相对物体（世界坐标系）的姿态了。求解  $M_2$  是一个 Perspective-n-points(PnP) 问题。OpenCV 中可以由 solvePnP 函数计算得到。因为 AprilTag 为平面标志，因此可令  $z_c = 0$ ，并以 AprilTag 标志的左上角为世界坐标系原点，则可以唯一确定相机相对于标志的位姿了。

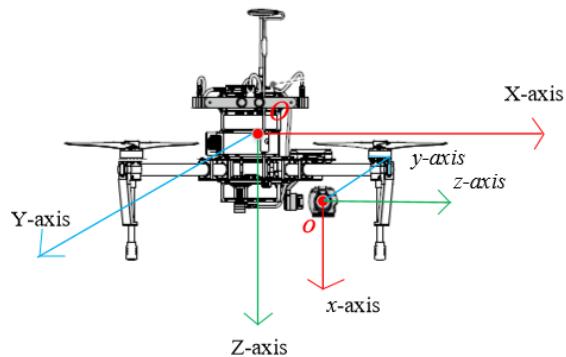


图 4-17 无人机坐标系与相机坐标系

图4-18为解算出来的相机的位姿。

```
3 tags detected:
Id: 1 (Hamming: 0) distance=1.49803m, x=1.35275, y=-0.643546, z=0.00383683, yaw=-0.0645363, pitch
=-0.483491, roll=-0.242449
Id: 0 (Hamming: 0) distance=1.33176m, x=1.25925, y=-0.433072, z=0.0175224, yaw=-0.0621219, pitch=
-0.483176, roll=-0.111067
Id: 3 (Hamming: 0) distance=1.3056m, x=1.20278, y=-0.455977, z=-0.223599, yaw=-0.0169165, pitch=
0.339641, roll=-0.289621
```

图 4-18 相机姿态估计结果

## 第5章 系统测试

### 5.1 红外图像候选区域提取实验结果

本系统采用两路视频源进行图像处理，包含红外图像处理以及光学图像处理。为了得到正确的候选区域坐标，需要求解两个相机坐标系之间的转换关系，如图5-1所示。

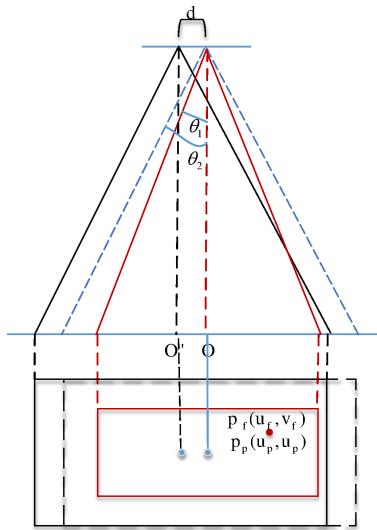


图 5-1 红外与光学相机坐标系之间关系

$\theta_1$  为红外热像仪的视场角， $\theta_2$  为光学相机的视场角大小（以  $x$  轴为例）， $d$  为两个成像系统轴心点的位置差。则红外相机图像坐标系中的点  $p_f(u_f, v_f)$  将通过下列变换映射到光学相机的图像坐标  $p_p(u_p, v_p)$ 。先考虑共轴时（蓝色虚线与红色实线）的情况，有以下关系：

$$u_p = u_{p0} + \text{Width}_P \cdot (u_f - u_{f0}) / \text{Width}_F \quad (5-1)$$

$$v_p = v_{p0} + \text{Height}_P \cdot (v_f - v_{f0}) / \text{Height}_F \quad (5-2)$$

式中， $\text{Width}_P(\text{Width}_F)$  和  $\text{Height}_P(\text{Height}_F)$  分别为光学（红外）图像的宽度和高度。实际上两个相机成像系统不共轴，安装位置有一个水平的偏差  $d$ ，此时上式中的  $u_p$  变为：

$$u_p = u_{p0} + \text{Width}_P \cdot (u_f - u_{f0}) / \text{Width}_F + d / fx \quad (5-3)$$

$v_p$  不变。其中  $fx$  是光学相机成像系统中每个像素代表的实际物理尺寸长度，在飞行高度很大的时候， $d / fx$  是一个很小的数值，由于红外图像处理的结果是提

取目标候选区域，因此也可以忽略该项。此外，系统使用的光学相机畸变很小，不加考虑。

根据第三章介绍的方法，机载计算设备通过 AV 视频采集卡，获取热成像视频数据，并进行处理，处理的结果为一系列坐标位置，代表目标候选区域。同时将红外图像进行伪彩化处理，输出到二进一出视频选择器，以供地面指令切换回传红外视频：计算设备接收飞机推送的切换指令之后，经由 USB 红外遥控功能令切换器进行视频源切换，再将视频通过高清图传传回地面。红外视频回传到地面站对于本系统扩展到夜间救援提供了可能。天空端的坐标回传和指令接收是通过 DJI Onboard SDK 数据透传功能实现。数据处理流程框图如图5-2所示。

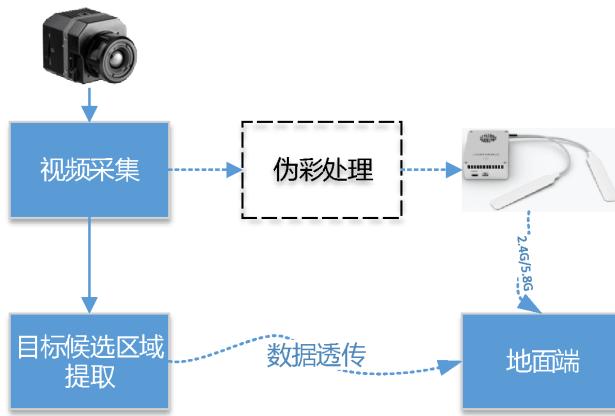


图 5-2 红外视频处理框图

由第三章介绍的处理算法，得到红外图像目标候选区域提取效果如图5-3所示。

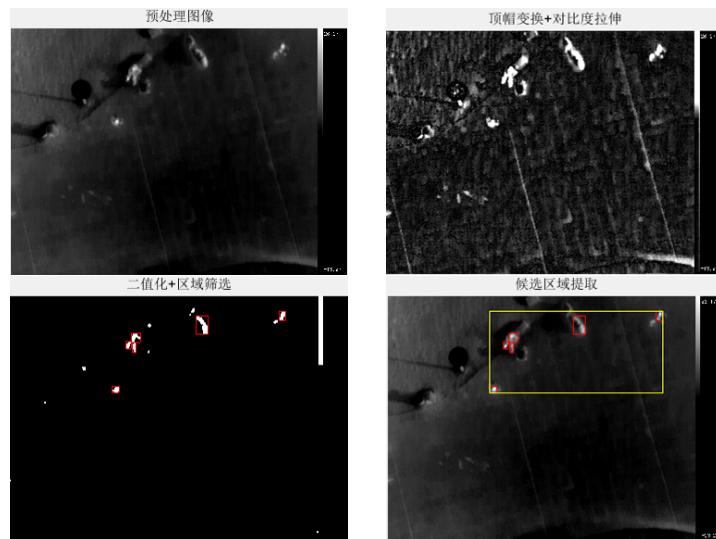


图 5-3 红外目标候选区域提取结果

考虑到实际搜救场景，搜救目标不会分布太分散。因此，在处理红外图像得

到所有目标候选区域框之后，计算它们的最小外接矩形框，将此矩形的四个顶点坐标回传，即上图中黄色矩形框的四个顶点。

地面端根据两个成像系统之间的图像坐标转换关系，得到目标候选区域在光学图像上的对应位置，将其设为光学图像的 ROI(Region Of Interest)，并基于此 ROI 进行光学目标检测。这样一来，可以大大减少不必要的计算，进一步提高目标检测的性能。因为使用原图 1080P 大小进行处理，势必会增加计算时间。使用红外检测结果设置的 ROI 不仅可以提高搜索效率，还可以方便地面工作人员进行观察。

## 5.2 光学目标检测结果

为了便于对改进后的 SSD 算法进行训练，建立了自己的无人机视角下人体数据集，目前，数据集只标注了“人”(person)这一类，仿照了著名的行人数据库 INRIAPerson (2416 个训练正样本， $90 \times 160$  大小)，包含目标的训练图片为 2680 张，大小为  $1280 \times 720$ (720P)。其中每一张中至少包含一个(1 ~ 10)目标样本，每个正样本都有对应的 Ground Truth (包围框)。

为了便于 Ground Truth 的标注，本系统还配套编写开发了一个标注工具(voc-annotation-tool)。其采用 MATLAB 语言开发，具有批量重命名图片，导入图片，以及以 VOC 格式标注图像并生成相应的文件结构等功能。软件运行界面如图5-4所示。



图 5-4 数据集标注工具运行界面

为了提高模型对视角变化的鲁棒性，主要使用的是斜视以及下视视野下的样本。经过标注之后，斜视  $85^\circ$ ，斜视  $45^\circ$  以及下视的各种姿态下的人体训练图片数

目比例大约为 1:1:1。图5-5为标注的样本示例。

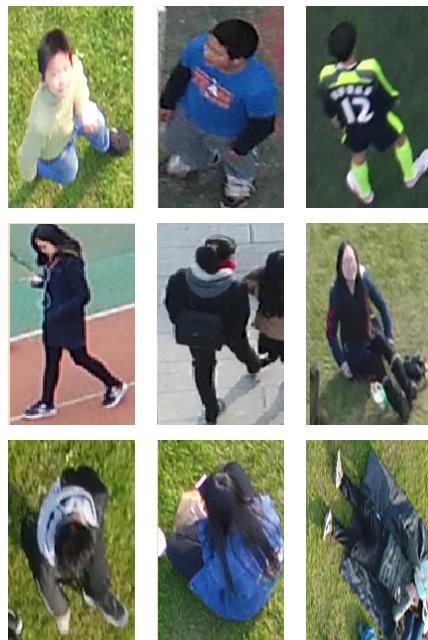


图 5-5 标注样本示例。  
上: 斜视 85°; 中: 斜视 45°; 下: 下视 (斜视 90°)

对于训练，本论文使用的机器配置如下：CPU 为 32 核心 i7-6700K，GPU 为 4 路 8GB 内存的 GTX1080 显卡，内存为 128GB。

将原始 SSD 算法的基础网络更改为残差网络，针对自建数据集训练模型参数，得到修改后的 SSD 目标检测模型。SSD 实际目标检测结果如图5-6所示。



图 5-6 改进 SSD 无人机视角目标检测结果

平均耗时：0.06s/每张图。

由实验结果即图5-6可以看出，改进后的 SSD 算法对于各种姿态以及视角，

如蹲坐(左上), 斜视(右上), 下视以及遮挡(左下)都有着较好的检测效果。

同时, 我们还进行了一系列的对比实验, 对比了传统目标检测算法 DPM、基于深度学习的检测算法 Faster-RCNN、改进前的 SSD 算法以及改进后的 SSD 算法在本数据集上的检测性能。实验对比了四种算法的 P-R 曲线图, 检测结果如图5-7所示。

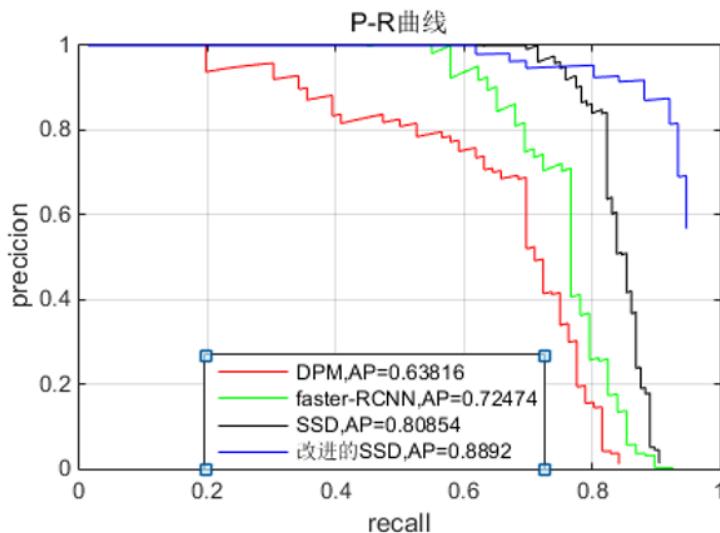


图 5-7 四种流行的检测算法的 P-R 曲线图

从图5-7中可以看出, 在本数据集上, 改进后的 SSD 算法表现最优, 平均检测精度 AP 达到 0.8892。

### 5.3 自主避障实验结果

对于第四章中介绍的基于双目视觉的自主避障技术, 在实际环境中进行了多次实验。由于实验场地的限制, 测试的场景比较有限, 主要测试了无人机系统对于墙壁、树木等静止障碍的躲避, 以及移动障碍(行人)的躲避。

本部分, 根据无人机实际尺寸设置一个“膨胀系数”以及“安全距离”, 分别用于设置实际避障使用到的无人机尺寸以及距离障碍物的最小距离, 实验中, M100 对角线(包含保护罩)的长度为 1m, 因此设置膨胀系数为 1.2, 安全距离为 0.5m, 这样, 实际上当无人机旋翼末端距离障碍 70cm 的时候为极限距离。实际飞行试验中, 设置无人机前进方向的最大飞行速度为 2m/s 进行实验。

实际障碍点云图以及规划的避障路线如图5-8所示。规划的避障路线是根据障碍物信息不断更新的。

实际自主避障实验结果如图5-9。

经过测试, 发现系统不仅对墙壁这种简单的平面障碍躲避灵敏, 还可以在错落复杂的大树之间穿梭以及躲避移动的障碍(行人)。因此有理由相信, 对于实际

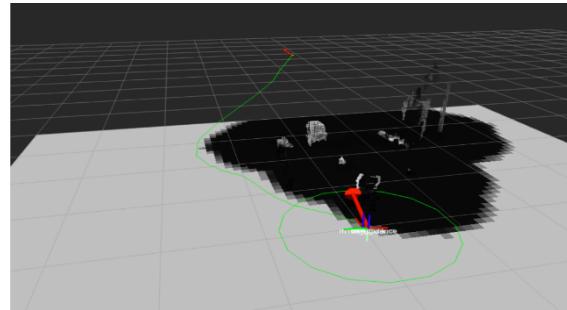


图 5-8 双目视觉生成的 3D 障碍点云以及规划的避障路径



图 5-9 躲避树木; 躲避行人

野外搜救中可能遇到的电线杆，铁塔等高大建筑，也可以达到较好的效果，限于实验环境，尚未进行相应的试验。

#### 5.4 移动降落实验结果

对于无人机移动降落实验，设置如下的试验方案。

打印 A0 大小的系列 AprilTag 标志，标志如下排列。采用大小标志组合主要基于两个考虑：

1. 无人机高空快速锁定降落点

无人机在空中飞行时 ( $10 \sim 20m$ )，地面标志成像大小需要足够大以完成标志检测识别，此时只需检测大尺寸标志。

2. 垂直降落过程中，接近降落平面时，无人机光学相机无法获取大标志的全貌，因此改为检测小标志，做进一步的姿态调整，从而实现精准降落。

实验中，将打印的标志图案贴于长宽分别为： $1.7m, 1.2m$  的白板上，大小模

拟普通敞篷皮卡后备车厢大小。分别放置在武汉大学工学部风雨操场上不同位置，设置无人机根据 GPS(位置精度 5m) 飞行至标志上方，然后系统自行检测降落标志，根据解算得到的相机姿态，飞行到标志正上方，为了保证在靠近过程中标志始终不移出无人机视野，系统对相机的姿态进行动态调整，使得标志始终处于视野中心。



图 5-10 降落标志保持在视野中心 (无人机相机视角视图)

当无人机飞行至标志正上方之后，固定相机为 90° 朝下，以  $0.5m/s$  的速度垂直下降，下降到距离降落平面  $1.3m$  之后切换为检测小标志 (主要是与大标志共线的小标志，取平均值) 并做最后的降落引导。

进行 20 次实验，以降落平面中心为基准点，测得降落点 (飞机质心投影) 距离其的位置偏差标准差为: X 方向:  $+11cm$ , Y 方向:  $+9cm$ 。此降落精度下，该 20 次试验，无人机均成功落在白板上，没有超出界限，如图5-11左图。



图 5-11 降落在平板上 (左); 降落在皮卡后部 (右)

此外，本系统进行了一次成功的降落在真实皮卡上的实验，证明了本系统移动降落的可行性。

## 5.5 配套演示系统

由于本系统是多模块集成系统，飞机端的数据处理以及指令传输主要是在机器人操作系统 ROS 上进行，无需图形化界面。对于地面站端，出于实际需求考虑，本系统包含一个配套的演示系统。其中包括一个移动端 App，负责以下功能：

1. 充当机载计算设备与地面计算设备的中介。通过建立移动设备和无线笔记本的 TCP 通信链路，由此地面计算设备（笔记本电脑）将能够获取到飞行端回传的候选区域系列坐标，以便进行下一步的光学检测。
2. 提供必要的交互，包括系统初始化，设定起始搜索点（GPS 坐标），拍照，录像以及返航等功能。
3. 飞行器状态信息显示，主要是一些重要的飞行器状态信息，如电池剩余量，飞行高度，实时 GPS 等信息等。

移动端 APP 运行界面如图5-12所示。



图 5-12 配套移动应用运行界面

系统运行时，首先点击初始化，获取无人机的控制权，然后输入起始搜索点的经纬度坐标。点击开始任务，则无人机开始起飞，飞向指定搜索起始点，并以之字形扫描搜索区域。飞行过程中无人机进行红外图像处理以及自主避障，直到搜索检测到目标之后，悬停，等待地面的指令：是否继续搜索。点击结束任务，则无人机返回起飞点上方，并启动自动移动降落程序，准确着陆在降落平台上。

为了方便地面工作人员观察评判目标检测结果，在地面笔记本电脑上开发了一个 GUI 界面程序。程序采用 Qt5 跨平台工具，在 Linux 系统上进行开发。采用多线程编程，将 UI 和图像处理分开处理，使得处理结果显示流畅。由于预先得

到了候选区域(红外检测给出), 实际应用于光学检测时, 1080P 分辨率的图片可以达到 22FPS 的处理速度, 基本达到实时。

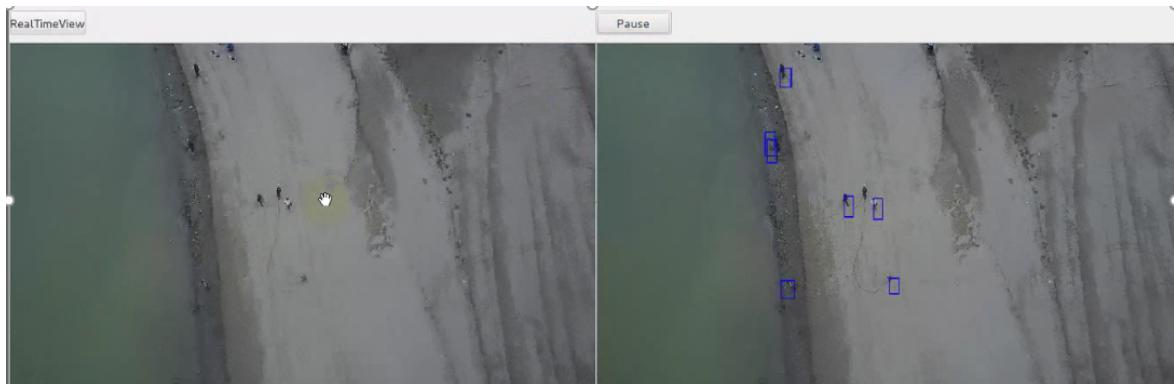


图 5-13 光学处理 GUI 程序运行界面

图5-13是 PC 端软件运行界面, 左边为实时预览界面, 右边是检测结果, 均支持局部放大和拖动。点击”开始”按钮就可进行光学检测。检测结果将使用蓝色包围框框住。下方显示目标当前所在 GPS 经纬度。

经过实际测试, 本系统各功能运行正常, 各模块响应速度快, 飞行试验 30 余次均未出现故障。

## 第 6 章 总结与展望

本文面向野外搜救应用，设计并实现了一套野外无人机搜救系统。针对搜救无人机系统的核心任务和实际需求，设计了系统的软硬件架构。整体硬件架构包括无人机系统平台、双目视觉模块，云台相机模块（包括光学相机和红外相机）、核心计算设备，安卓远程设备，地面工作站的设计与实现，并成功搭建了系统原型，针对野外搜救任务中的三个重要功能：目标搜索、自主避障和移动降落，进行了软件算法设计和验证实验。在目标检测方面，采用了红外图像辅助的基于深度学习目标检测方法，并构建了一个无人机多视角的人体检测数据集，实验结果表明，深度卷积神经网络目标检测方法可以从数据中学习到相比手工设计更能表达目标的特征，也因此获得更高的准确率。在无人机自主避障和路径规划方面，系统通过双目点云估计障碍深度信息，障碍信息通过更新代价地图的方式更新避障路径，从而指导无人机避开障碍。在无人机移动降落方面，利用增强现实技术中的 AprilTag 视觉基准系统，通过对降落标志的荷载进行编码处理来最大限度的减少虚警。并通过求解 PnP 问题，获取无人机相机相对于标志的位姿，并以此作为无人机位姿调整的依据，精确引导无人机降落在标志上。实验结果表明利用移动降落可以达到 10cm 以内的降落精度，移动可以降落在 2 ~ 3m/s 的移动平台上。在系统搭建上，本系统采用天空端和地面站两线工作模式。自主避障，红外目标候选区域提取以及自主返航移动降落均由机载计算设备 Manifold 上完成。地面端主要由遥控器、安卓手机和高性能笔记本电脑组成。为了完成必要的交互以及充当天空端机载计算设备与地面笔记本的通信中介，开发了一个移动应用 (APP)。为了可视化搜索结果，在笔记本电脑上开发了基于 Qt5 的图形界面程序。天空端和地面端程序均使用多线程以及模块化编程，各个模块之间不存在耦合，十分方便单独测试。多次飞行试验表明搭建的系统原型稳定性较高，可以成功完成搜救试验。

本论文主要将无人机搜救工作的三个关键技术进行研究并将其以功能模块的形式集成到一个搜救系统原型中，因此后续仍有很大的发展空间。随着传感器技术的发展，传感器的种类将会越来越多，因此，无人机搜救系统可以考虑搭载更多的传感器，进行多元信息融合从而提高搜救效率。另外，目前目标检测算法是基于深度学习的框架，但是为了充分发掘深度网络模型在特定应用场景下的潜力，仍需要大量的自建数据。目前本系统使用的自建数据集样本数量偏少，后续工作应该包括扩大数据集。此外，单就“人”这一类别而言，姿态为平躺或俯卧，蜷缩等姿态的样本仍然十分有限，因此可以借助机器学习中生成式对抗网络 (Generative Adversarial Nets, GAN) 进行此类样本的自动生成，进一步扩充数据集。

第三点就是增加更多地面站功能。现在的无人机飞控系统通常会支持 MAVLink 协议，这样就使得地面站功能（比如航点飞行，热点跟随）可以很方便实现。实际实施搜救时，使用到更多地面站功能，就能更方便地规划整个搜救任务，进一步提高搜救效率。

## 参考文献

- [1] DJI.DJI-Onboard-SDK[CP/DK].<https://developer.dji.com/cn/onboard-sdk/>,2014-2017, 2017-04-26.
- [2] 高敏, 李怀胜, 周玉龙, 方丹, 张宝全. 背景约束的红外复杂背景下坦克目标分割方法 [J]. 自动化学报, 2016, 42(3):416-430.
- [3] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. European Conference on Computer Vision. Springer International Publishing, 2014:346–361.
- [4] Chan C-S, Chang C-C. An efficient image authentication method based on Hamming code[J]. Pattern Recognition, 2007, 40(2): 681–690.
- [5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. Advances in neural information processing systems, 2015:91-99.
- [6] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. Springer International Publishing, 2016:21–37.
- [7] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. arXiv:1612.08242 [cs], 2016.
- [8] Harris C, Stephens M. A combined corner and edge detector[A]. In Proc. of Fourth Alvey Vision Conference[C]. 1988: 147–151.
- [9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[A]. Proceedings of the 25th International Conference on Neural Information Processing Systems[C]. USA: Curran Associates Inc., 2012: 1097–1105.
- [10] Erhan D, Szegedy C, Toshev A, et al. Scalable Object Detection using Deep Neural Networks[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014:2147-2154
- [11] Zhang L, Lin L, Liang X, et al. Is Faster R-CNN Doing Well for Pedestrian Detection?[J]. Springer International Publishing, 2016:443-457.

- [12] Sniedovich M. Dijkstra's algorithm revisited: the dynamic programming connection[J]. *Journal of Control and Cybernetics*. 2006, 35 (3): 599–620.
- [13] Heng L, Gotovos A, Krause A, et al. Efficient visual exploration and coverage with a micro aerial vehicle in unknown environments[A]. 2015 IEEE International Conference on Robotics and Automation (ICRA)[C]. 2015: 1071–1078.
- [14] Heng L, Honegger D, Lee G H 等. Autonomous Visual Mapping and Exploration With a Micro Aerial Vehicle[J]. *Journal of Field Robotics*, 2014, 31(4): 654–675.
- [15] Olson E. AprilTag: A robust and flexible visual fiducial system[C]. IEEE International Conference on Robotics & Automation, 2011 19(6), pp.3400-3407.
- [16] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision[M]. 2nd ed. New York, NY, USA: Cambridge University Press, 2003.
- [17] Thrun S, Leonard J J. Simultaneous Localization and Mapping[A]. see: B.S. Prof, O.K. Prof. Springer Handbook of Robotics[M]. Springer Berlin Heidelberg, 2008: 871–889.
- [18] Huang R, Tan P, Chen B M. Monocular vision-based autonomous navigation system on a toy quadcopter in unknown environments[A]. 2015 International Conference on Unmanned Aircraft Systems (ICUAS)[C]. 2015: 1260–1269.
- [19] Wendel A. Scalable Visual Navigation for Micro Aerial Vehicles using Geometric Prior Knowledge[D]. PhD thesis, 2013.
- [20] Mei C, Sibley G, Cummins M, et al. A Constant-Time Efficient Stereo SLAM System[M]. 2009.
- [21] Shen S, Mulgaonkar Y, Michael N, et al. Vision-based state estimation for autonomous rotorcraft MAVs in complex environments[A]. 2013 IEEE International Conference on Robotics and Automation[C]. 2013: 1758–1764.
- [22] Ferguson D, Stentz A. Using interpolation to improve path planning: The Field D\* algorithm[J]. *Journal of Field Robotics*, 2006, 23(2): 79–101.
- [23] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147–1163.

- [24] Kohlbrecher S, Stryk O von, Meyer J, et al. A flexible and scalable SLAM system with full 3D motion estimation[A]. 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics[C]. 2011: 155–160.
- [25] Fraundorfer F, Heng L, Honegger D, et al. Vision-based autonomous mapping and exploration using a quadrotor MAV[A]. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems[C]. 2012: 4557–4564.
- [26] Olson E. AprilTag: A robust and flexible visual fiducial system[A]. 2011 IEEE International Conference on Robotics and Automation[C]. 2011: 3400–3407.
- [27] AprilTag: A robust and flexible visual fiducial system[EB/OL]. ResearchGate. /2017-06-08. [https://www.researchgate.net/publication/224252184\\_AprilTag\\_A\\_robust\\_and\\_flexible\\_visual\\_fiducial\\_system](https://www.researchgate.net/publication/224252184_AprilTag_A_robust_and_flexible_visual_fiducial_system).