基于多阶运动参量的四旋翼无人机识别方法

刘孙相与 [1, 2] 李贵涛 [1, 2] 詹亚锋 [1, 2] 高 鹏 [3]

摘 要 以小型多轴无人机为代表的 “低慢小”目标, 通常难以被常规手段探测, 而此类目标又会严重威胁某些重要设施. 因此对该类目标的识别已经成为一个亟待解决的重要问题. 本文基于目标运动特征, 提出了一种无人机目标识别方法, 并揭示了二阶运动参量以及重力方向运动参量是无人机识别过程中的关键参数. 该方法首先提取候选目标的多阶运动参量, 建立梯度提升树 (Gradient boosting decision tree, GBDT) 和门控制循环单元 (Gate recurrent unit, GRU) 记忆神经网络分别完成短时和长期识别, 然后融合表观特征识别结果得到最终判别结果. 此外, 本文还建立了一个综合多尺度无人机数据集 (Multi-scale UAV dataset, MUD), 本文所提出的方法在该数据集上相对于传统基于运动特征的方法, 其识别精度 (Average precision, AP) 提升 103%, 融合方法提升 26%.

关键词 四旋翼无人机, 目标识别, 运动特征, 融合方法

引用格式 刘孙相与, 李贵涛, 詹亚锋, 高鹏. 基于多阶运动参量的四旋翼无人机识别方法. 自动化学报, 2022, **48**(6): 1429−1447

DOI 10.16383/j.aas.c200862

Drone Detection Based on Multi-order Kinematic Parameters

LIU Sun-Xiang-Yu[1, 2] LI Gui-Tao[1, 2] ZHAN Ya-Feng[1, 2] GAO Peng[3]

Abstract Due to the features of low, slow and small aircraft, such as quadrotors, it is a challenging and urgent problem to detect UAVs (Unmanned aerial vehicles) in the wild. Different from the past literatures directly using deep learning method, this paper exploits motion features by extracting multi-order kinematic parameters such as velocity, accelerate, angular velocity, angular velocity vectors and it is exposed that 2nd order and gravity direction motion parameters are key motion patterns for UAV detection. By building GBDT (Gradient boosting decision tree) and GRU (Gate recurrent unit) network, it comes out with a short-term and a long-term detection result, respectively. This recognition process integrates appearance detection result into motion detection result and obtains the final determination. The experimental results achieve state-of-the-art result, with a 103% increase on the precision index AP (Average precision) with respect to the previous work and a 26% increase for hybrid method.

Key words Quadrotors, object detection, motion feature, fusion method

Citation Liu Sun-Xiang-Yu, Li Gui-Tao, Zhan Ya-Feng, Gao Peng. Drone detection based on multi-order kinematic parameters. *Acta Automatica Sinica*, 2022, **48**(6): 1429−1447

“低慢小”(飞行高度低、飞行速度慢、目标小) 目标以其难以被探测、便于隐藏、适用场景广泛的特点, 一直以来都是军事以及科研领域中的研究重点[1−4], 其中 “低慢小”目标的探测识别更是相关课题中的核心和基础问题. 近年来, 四旋翼无人机为代表的新兴 “低慢小”飞行器因其成本低廉、操纵简单、难以被发现的特点, 在航拍、探测、检测等多个领域被广泛应用. 但随之而来也带来诸多安全隐患, 如成都机场无人机 “黑飞”逼停客机、默克尔总理竞选会无人机潜入、叙利亚自制 “武装无人机”自杀式袭击等. 这些已有公共安全事件说明无序飞行的 “低慢小”无人机已经严重威胁到社会秩序和公共安全.

近年来, 人工智能和计算机视觉的发展, 使得基于图像/视频的小目标检测与识别方法的性能有了较大的提升, 成为研究此类问题的新手段[5−8]. 相比于以往基于声谱特征[3−4]、光谱特征[5−6]、射频和雷达[1, 9−10] 等方法, 基于机器视觉的方法具备系统简单、硬件体积小、场景普适性强、探测距离远、识别粒度细等优点. 基于机器视觉的 “低慢小”目标识别方法主要包括表观特征方法[11−26]、运动特征方法[27−35] 以及混合方法[15, 36−45].

基于表观特征的方法, 如部件模型 (Discriminatively-trained part model)[11]、Faster RCNN 神经

收稿日期 2020-10-14 录用日期 2021-02-09
Manuscript received October 14, 2020; accepted February 9, 2021
国家重点研发计划 (2018YFD100303) 资助
Supported by National Key Research and Development Program of China (2018YFD100303)
本文责任编委 穆朝絮
Recommended by Associate Editor MU Chao-Xu
1. 清华大学宇航中心 北京 100084 2. 北京信息科学与技术国家研究中心 北京 100084 3. 北京大学工学院 北京 100871
1. Space Center, Tsinghua University, Beijing 100084 2. Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084 3. College of Engineering, Peking University, Beijing 100871

网络[12]、SSD(Single shot multibox detector) 神经网络[26]、积分通道 (Integral channel)[13] 等在许多常见目标以及一些小目标识别任务中显著提升了识别精度. Zahangir 等[24] 改进循环卷积神经网络, 融合 Inception-V4 和残差网络结构, 形成 IRRCNN识别网络完成对输入图像的目标识别, 在多个数据集上, 如 CIFAR-10、CIFAR-100、TinyImageNet-200 以及 CU3D-100, 达到最佳识别精度. 对于无人机目标来说, Schumann 等[17−18] 提出了采用 Faster RCNN 网络进行识别的方法, 并在其建立的数据集上进行训练, 识别鸟类和无人机两类目标, 在 AVSS2017[23] 测试集上取得了最高精度; Saqib 等[25] 测试了不同结构的卷积神经, 得出采用 VGG16结构的 Fasster-RCNN 神经网络具备最高识别精度; Aker 等[14] 提出了将鸟类和无人机在不同背景下合成的数据集生成方法, 用以训练无人机识别神经网络; Wu 等[28] 提出通过将显著性方法引入至卡尔曼滤波器, 完成对运动小目标的跟踪和定位, 该方法对于四旋翼无人机的跟踪也具有较高精度. Carrio 等[20−21] 在深度图中采用神经网络方法完成四旋翼无人机的识别, 并在 Airsim 飞行仿真软件中建立深度图数据集, 用以训练识别方法, 得到了其数据集上的最优识别精度. 但该方法对目标的表观和运动特征均未直接使用, 对于常见的识别场景适用性较差、识别精度相对较低.

基于运动特征的方法, 主要分为两类, 一类是基于背景减除; 另一类是基于流方法. 背景减除类方法的前提是假设相机不动或者仅有很小移动. 通过对背景进行建模, 从而达到仅在图像中留下前景目标的目的, 此类方法[27, 30−31] 计算复杂度低、适用场景广泛, 但仅能在背景简单下具备足够精度; 流方法[6, 32−34] 依赖于流向量的计算, 其适用于多目标场景、在复杂场景中也具备较高召回率, 但对于识别任务来说, 针对小目标或者复杂场景计算精度不足, 计算复杂度和虚景率也较高. 基于深度网络的光流提取方法提高了光流向量的计算精度, Dosovitskiy 等[33−34] 提出 FlowNet、FlowNet2.0 等结构, 采用 U-Net 架构, 并融合多种网络结构, 取得了目前最优光流提取性能.

融合运动以及表观特征的方法, 目前多以深度网络 (Deep neural network, DNN) 为基础框架, 主要包括卷积神经网络 (Convolutional neural network, CNN)[37−41] 和循环神经网络 (Recurrent neural network, RNN)[42−43, 46]. T-CNN (Tublet CNN)[37] 借用 Faster RCNN 中 RPN (Region proposal network) 的高效结构, 提出 Tubelet 结构关联上下文特征, 即通过光流法得到的在连续多帧中同一目标识别矩形框, 并采用 LSTM (Long short-term memory)[46] 网络作为分类器完成分类. 此方法能够抑制虚景目标, 提升正样本的识别概率, 但对于小目标召回率较低. DFF (Deep feature flow)[38] 使用基于深度网络框架的 FlowNet[33] 方法提取光流特征, 通过目标运动过程联系上下帧并筛选关键帧, 节省了对非关键帧特征提取和识别的计算过程. Zhu 等[39] 在像素级 (Pixel-level) 融合通过 FlowNet 计算得到的光流区域的特征图, 融合相邻多个特征图并输入到最终的判别网络中. 与以上两个工作类似, 本文方法也采用了光流法提取上下帧目标的运动过程, 但并非综合运动过程中变化的外观特征, 而是重建目标运动过程中的运动学参数. Bertasius 等[40] 引入可变尺寸卷积 (Deformable convolution) 对上下帧中目标运动引入的额外特征进行融合, 而非采用光流联系上下帧. Luo 等[41] 融合区域级特征 (Proposal-level) 而非像素级, 其考虑候选区域内的语义特征, 并综合相邻两帧语义特征、位置特征以及时间特征完成识别, 取得了 ImageNet VID[47] 数据集中的最优性能. 以上方法主要以 Faster RCNN 或 RPN 为主要框架, 近年来, 以 RNN 为框架的方法[41−44] 在计算效率以及精度上也达到了较高水平, Xiao 等[42] 利用 ConvGRU结构融合时空特征, 在 ImageNet VID 数据集上, 曾取得最优性能. Chen 等[43] 提出的基于 ConvLSTM[44] 和 SSD (Single shot multibox detector)[26] 网络结构, 并融合注意力机制的方法, 综合了多尺度的特征 (像素级和目标级), 是目前综合计算速度与精度的高性能方法. 本文方法也采用了基于 RNN 结构的 GRU (Gated recurrent unit) 网络[43−44] 作为分类器, 但其输入为运动参量, 而非图像.

特别地, 对于 “低慢小” 目标的混合识别方法, Lv 等[29] 通过融合时空两种特征, 完成了对弱小飞行器目标的探测; Shi 等[36] 提出采用改进粒子滤波的方法探测低速飞行小目标, 对于海面背景的飞行器目标来说, 其相较与分型方法 (Fractal-based) 和三特征方法 (Tri-feature-based) 性能更佳. 对于无人机目标来说, Farhadi 等[23] 提出将前景检测结合目标形状进行识别的方法, 在综合指标上, 取得了 AVSS2017[14, 16, 23] 方法中第二高精度的性能. Sapkota 等[19] 提出利用级联检测的思路, 识别无人机后利用混合高斯概率假设密度滤波器跟踪无人机飞行轨迹, 实现了两架无人机的实时跟踪. Rozantsev 等[15] 融合了表观特征以及运动特征, 利用目标运动补偿来提高识别精度, 即通过决策树和卷积神经网

络估计目标在像平面的运动, 进而采用卷积神经网络识别获得的图像立方体中的目标. 该方法在其提供的测试集中取得了目前最优结果. 但该方法未考虑多干扰目标和多类别的识别, 难以应用在实际场景中.

相较于以往工作, 与文献 [15] 相似, 本文方法也基于融合表观和运动特征的思想, 采用了文献 [6, 32−36] 中所涉及到的光流法进行运动特征提取, 并利用文献 [42−43, 46] 等工作中提及的 GRU 网络完成目标判别. 但不同的是本文从运动学角度直接提取目标的运动特征, 而非仅采用运动特征辅助串联前后帧表观特征的提取. 并且本文采取决策融合的方式而非特征融合, 这样能针对性地充分考虑运动和表观两个不同维度的特征. 从算法适用条件及精度来说, 以往工作都在一定程度上实现了无人机的跟踪和目标的识别, 但基本都要求单一纯净背景下的单目标作为前提条件. 而对于低空干扰目标较多、背景较复杂这一现实约束, 这些方法均无法做到高精度识别. 此外, 以往工作均采用对常见物体识别使用的通用框架, 并未意识到无人机 “低慢小”的特殊之处, 也未对此特点加以利用. 在构建相关实验数据集时, 也未考虑无人机的特征, 涵盖的飞行场景较少.

本文以典型四旋翼无人机探测为目标, 综合其表观和运动特征, 提出了一种基于目标多阶运动参量的识别方法 (Multi-order kinematic parameters based detection method, MoKiP). 本文中, 多阶运动参量是指一个运动参数的集合, 包括零阶运动参量 (表观特征), 一阶运动参量 (速度、角速度), 二阶运动参量 (加速度、角加速度), 以及更高阶的运动参量.

如图 1 所示, 该方法的核心思想如下: 首先提取并跟踪运动候选区域, 并估计候选区域的深度信息, 然后计算出相应的非零阶运动参量, 之后, 采用梯度提升决策树以及记忆神经网络完成基于运动特征的短期和长期识别. 同步地, 采用 Faster RCNN[12] 深度网络对零阶运动参量 (表观特征) 进行识别. 最后, 将零阶和非零阶两部分识别结果, 按照识别概率加权平均融合, 得到最终的判别结果和类别概率.

实验证明, 在目标像素较少、背景复杂以及干扰目标较多的情况下, 相比于以往方法, 本文提出的方法具有更高的识别精度. 此外, 通过灵敏度分析, 本文进一步定量分析了各阶运动参量对识别精度的贡献程度, 并发现二阶参量、重力方向参量是识别过程中影响较大的重要特征.

本文的主要贡献如下:

1) 提出基于多阶运动参量的 “低慢小”识别方法. 较好地处理了低空、复杂背景以及多目标场景下的识别问题.

2) 发现了二阶运动参量以及沿重力方向的运动参量最能反映无人机与其他干扰目标在运动特征上的差异.

3) 建立了多尺度无人机数据集. 包含四旋翼无人机以及行人、车辆、鸟类等干扰目标的相关数据. 并为其它干扰目标进行了数据采集和标定.

1　基于多阶运动参量的无人机识别方法

1.1　总体识别流程

本文在充分挖掘无人机运动信息的基础上, 提出了一种基于多阶运动参量判别融合的无人机识别方法. 其输入为场景的视频片段, 输出为目标的识别矩形框和所属类别概率. 该方法的流程如图 1 所

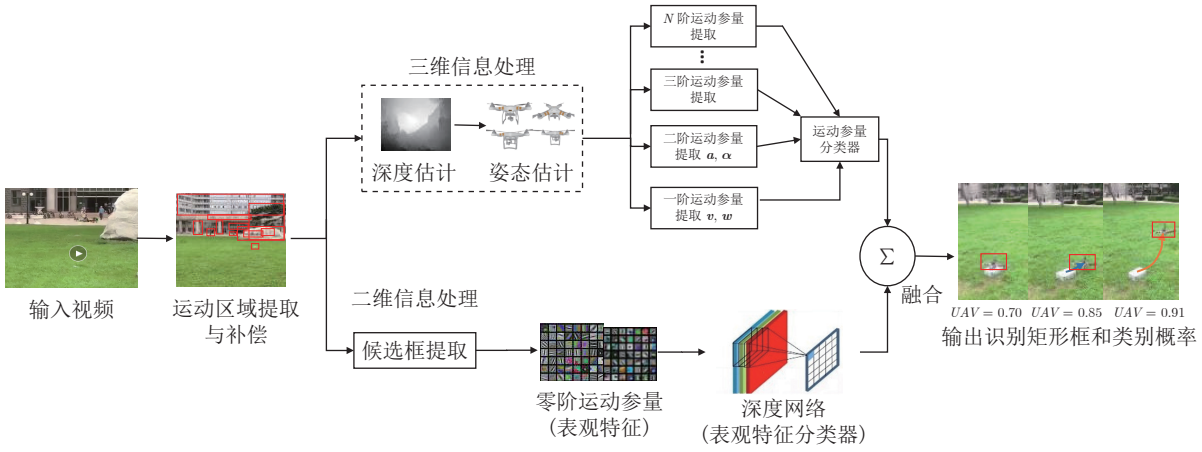


图 1　本方法整体流程图

Fig. 1　The overall flowchart of our method

示: 首先, 利用 ViBe+ (Visual background extractor)[30] 法, 提取候选运动区域. 然后, 分别提取无人机的表观特征和运动特征, 并分别根据这两类特征识别目标类别. 最后, 融合两个识别结果, 给出最终识别的概率.

本文中定义物体的表观特征为零阶运动参量. 其处理流程如图 1 下半分支所示. 利用 Faster RCNN 深度神经网络, 根据输入视频获得目标图像特征的识别矩形框和类别概率. 图 1 上半分支根据目标运动特征, 即非零阶运动参量进行识别. 该方法首先利用 ViBe+法提取运动区域, 其次, 通过单目估计或物理测量等方法获得运动目标区域深度值. 之后, 根据深度图, 估计运动区域内目标的零阶以上运动参量. 然后, 训练得到基于运动参量的 GBDT 决策树 (Gradient boosting decision tree)[48] 和 GRU (Gated recurrent unit)[46] 记忆网络, 分别实现对无人机的短时和长期的识别, 并得出识别矩形框和所属类别概率. 最后, 将零阶和非零阶两部分识别结果, 按照识别概率进行加权平均融合, 得到最终结果和类别概率.

1.2 基于零阶运动参量的特征提取方法

零阶运动参量代表了目标 “不动”时所传递的信息, 也就是其表观特征. 以往工作中已经有了很多成熟有效的算法[11−13, 26, 46, 49] 进行表观特征提取, 本文采用了以提取区域候选网络 (Region proposal network, RPN) 为前端的两阶段 Faster RCNN[12] 神经网络. 其在 Pascal VOC[11]、ImageNet[47] 等公开数据集中, 均取得了最优性能 (State-of-the-art, SOTA). 本文使用基于 Resnet101[49] 框架的 Faster RCNN 网络, 以获得目标识别的矩形框, 以及 5 类目标的识别概率. 所采用的 Resnet101 结构在 ImageNet 数据集中预训练, 并在本文多尺度无人机数据集 (Multi-scale UAV dataset, MUD) 中参数细调 (Fine-tune). 对于 RPN 网络的训练, 尺度参数设置为 5 (2, 4, 8, 16, 32), 3 个矩形框比例分别设为 (0.5, 1, 2), 总共 15 个锚 (Anchors). 在训练时, 使正负样本数比例达到 1:1.

在使用本文融合方法进行识别时, 采用按训练识别概率加权[50] 的方法, 融合基于零阶与下文非零阶的识别结果, 得到最终判别结果. 具体来说, 对于某一候选区域、某一类别的识别概率为分别采用零阶、非零阶运动参量方法识别得到的概率按测试集 (在调参时按训练集) 准确率加权求和的结果. 若某一区域仅被零阶或非零阶中的一种方法所识别, 则另一方法识别概率按零计算.

1.3 基于非零阶运动参量的特征提取方法

图 2 给出了基于非零阶运动参量识别的详细流程, 其输入为运动区域的图像流, 输出为识别得到的识别矩形框与类别概率. 以下各小结将根据运动特征识别的流程, 依次阐述识别过程中的各个环节. 主要包括目标运动区域提取、运动参数辨识、候选目标姿态测量、目标类别与运动参量的条件概率密度函数估计等. 其中, 参数辨识过程包括了相机运动的识别与补偿、水平面估计、深度估计等. 对于条件概率密度函数的估计, 本文利用梯度提升树完成相邻几帧的短时识别; 利用 GRU 记忆网络完成长时识别. 在描述每一步处理的过程中, 本文也将分析每个环节对最终识别效果的影响.

1.3.1 运动目标区域提取

疑似目标区域提取是本文所述识别方法的第

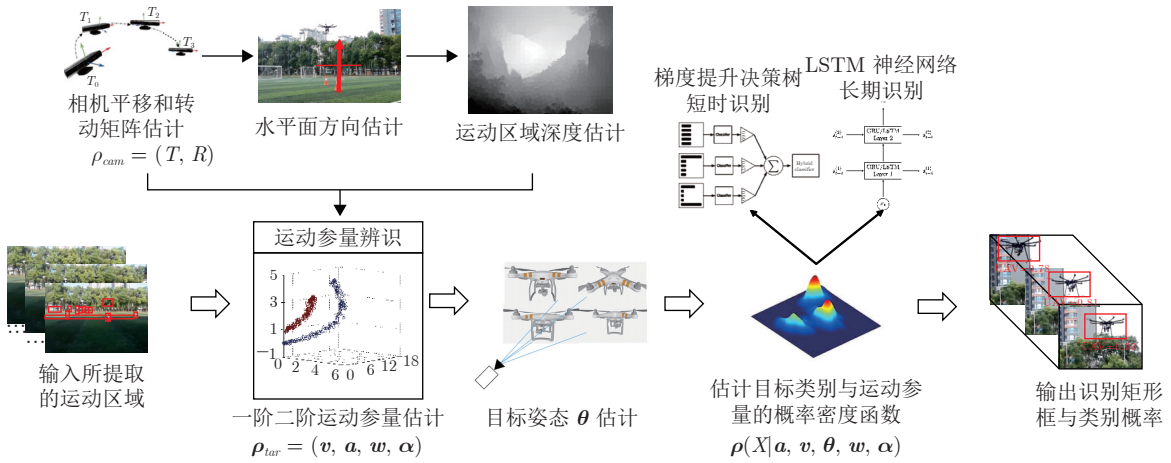


图 2 基于多阶运动参量的目标识别方法流程图 (MoKiP)

Fig. 2 Flowchart of multi-order kinematic parameters based detection method (MoKiP)

1 步. 在无人机识别问题中, 目标所处的环境复杂多样, 反映到图像, 则会导致目标图像具有背景变化剧烈、多目标的特点. 所以本文采用目前在多数常见场景都具备高召回率的 ViBe 改进算法 ViBe+[30] 提取运动区域. 其主要流程为:

1) 背景初始化建模

给每一个像素点建立像素样本集. 一般为从该点邻域以及过去时刻邻域像素中随机选取 20 个点. 邻域点即与该像素点相邻的 8 个像素点.

2) 前景检测

设置闪烁阈值以及更新因子. 对于本时刻某点邻域内, 若邻域点中像素值大于闪烁阈值的点的个数超过更新因子, 则将该点设为前景点.

3) 背景模型更新

某像素点只有被分为背景样本时, 才能被包含在背景模型中, 而前景点不能被用于构成背景模型. 更新过程遵循时间和空间的随机性. 空间随机性是更新的像素随机替代样本中任意像素, 时间随机性是指当一个像素点被判定为背景时, 它有 $1/rate$ 的概率更新背景模型. $rate$ 为更新因子, 根据更新需要设置为 1、5 或者 16 等值, 由于本文所涉及的场景背景有较快速变化, 此处设置 $rate = 5$; 如果某个像素点连续 N 次被检测为前景, 则将其更新为背景点, 一个像素点在本时刻不被更新的概率为 $(N-1)/N$, 在本文设置 $N = 15$.

另外, 加入关于 Ghost 区域的消除、除去不完整目标、自适应阈值等改进, 其余参数详见文献 [30]. 通过 Vibe+方法提取的运动区域如图 3 示意.

Mov_i 表示提取出的运动区域 (目标包络线以内); 如图 3 所示, 采用一个能够包含此区域的最小矩形包围框, 作为待检测目标区域候选矩形框. 则候选窗口可以用式 (1) 表示:

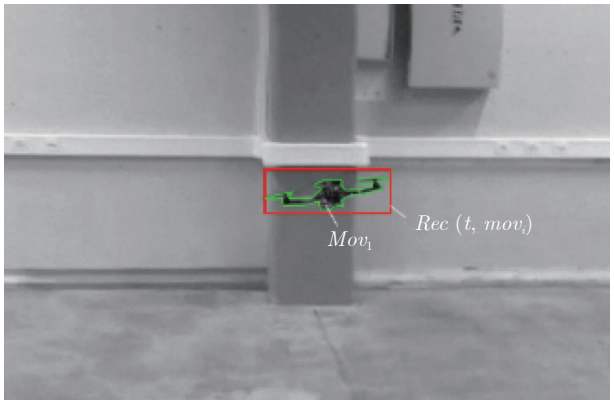


图 3　运动区域提取示意图

Fig. 3　An illustration of the extracted motion ROI (Region of interest)

$$R_m(x,\ y,\ t) = \begin{cases} 1, & (x,\ y) \in Rec(t,\ Mov_i) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

其中, $Rec(t,\ Mov_i)$ 表示在 t 时刻包括运动区域 i 的最小矩形框; R_m 为最小矩形框所形成的二值掩码矩阵.

1.3.2　运动参量辨识

快速且精确的运动参量辨识过程是本文提出方法的核心. 如不加特殊说明, 本文中所有运动参量均以地面坐标系为基准 (X 轴与图像平面坐标系中 u 轴方向保持一致, Z 轴铅垂向上, Y 与 X 和 Z 轴构成的平面垂直, 构成右手坐标系), 并根据其相机坐标系下的测量分量, 通过水平面估计矩阵转换至地面坐标系下对应的分量. 本节所述的一阶和二阶运动参量包括三轴平动速度、三轴角速度, 以及对应的加速度、角加速度. 三阶及三阶以上的参量较为特殊, 将在第 3.6 节中单独阐述.

本文假设相机始终保持静止. 对于场景中的运动目标, 首先获取目标运动区域内的深度值; 然后, 进行水平面的矫正; 最后, 采用差分估计提取得到目标的运动参量.

1) 运动区域深度图

目标在图像平面内的运动和其对应的深度值共同决定了目标在三维空间内的真实运动规律. 因此需要首先获取目标的深度信息.

目前获得深度图的手段有激光测距、立体视觉、图像估计等方法. 根据不同识别场景的需求, 应选取不同方法获得深度图, 在获得的深度图中, 每个像素代表该图像位置的深度值, 此外还可能包含深度测量的置信度或误差等信息. 对于常见的识别场景, 从图像中直接估计深度信息的方法具备更强的适用性, 所以本文选择采用目前单目深度估计方法中具备最佳精度的 DORN (Deep ordinal regression network)[51] 方法.

2) 水平面方向估计

考虑到相机仍存在旋转, 为了更加准确的估计运动参量, 需要补偿相机旋转对参量估计的影响, 修正深度方向至与相机所在世界坐标系保持一致. 本文采用改进隐马尔科夫[52] 方法进行估计, 获得水平面旋转修正矩阵. 该方法能够在多种场景下鲁棒地估计水平面方向, 并具有相比于以往文献较高的精度. 对于本文的识别问题, 当探测场景为低空场景时, 直接采用此方法进行水平面修正. 而当探测场景为对空场景时, 场景中无地平线作为参考, 因此无法获取估计所需的特征. 此情况下, 可近似认为深度方向即为海拔高度方向.

3) 一阶和二阶运动参量提取

根据以上章节获得的候选区域逐点深度信息, 及其在像平面内的轨迹, 本节将从这些信息中提取运动区域内目标的特征点, 然后计算其一阶和二阶运动参量.

首先, 本方法采用具有快速和鲁棒特性的 ORB (Oriented fast and rotated brief)[53] 算法提取 Mov_i 内的目标特征点. 设当前第 t 帧中运动区域 Mov_i 内提取的第 i 个特征点 (ORB 算法具备足够检出速度和特征点鲁棒性) 在图像坐标系下为 $P_t^i=\left(x_t^i,\ y_t^i\right)$, 深度为 d_t^i, 其在 $t-1$ 帧, 对应的匹配点为 $P_{t-1}^i=\left(x_{t-1}^i,\ y_{t-1}^i\right)$, 其深度为 d_{t-1}^i, 其在 $t+1$ 帧对应的匹配点为 $P_{t+1}^i=\left(x_{t+1}^i,\ y_{t+1}^i\right)$, 深度为 d_{t+1}^i, 相机焦距为 f, 帧率为 F, 总共特征点个数为 N, 根据相机几何, 则此时刻第 i 个目标特征点在三维空间中的坐标为 $P_i=\left(X_i^w,\ Y_i^w,\ Z_i^w\right)=\left(d_t^i x_t^i/f,\ d_t^i y_t^i/f,\ d_t^i\right)$, 相机坐标系下坐标为 $P_i^c=\left(X_i^c,\ Y_i^c,\ Z_i^c\right)$.

运动区域 Mov_i 内待提取的运动参数 $\boldsymbol{\rho}=(\boldsymbol{v},\ \boldsymbol{a},\ \boldsymbol{\omega},\ \boldsymbol{\alpha})$, 分别代表目标运动的平动速度, 平动加速度, 角速度, 角加速度矢量, 即目标作为刚体运动时的运动学参数. 为估计这些运动参量, 第 1 步是获得目标的位移向量 $\boldsymbol{T}_t=[T_t^x,\ T_t^y,\ T_t^z]^{\mathrm{T}}$, 以及旋转矩阵

$$R_t=\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}=\begin{bmatrix} R_1^{\mathrm{T}} \\ R_2^{\mathrm{T}} \\ R_3^{\mathrm{T}} \end{bmatrix}$$

若运动区域内相邻帧匹配特征点 $N\geq 4$ 时, 采用 EPnP (Efficient perspective n point)[54] 方法, 则空间中点 $P_i=(X_i^w,\ Y_i^w,\ Z_i^w)$, 其图像中匹配点为 $u_i=(x_i,\ y_i)$, 四个控制点的坐标为 $c_j,\ j=1,\cdots,4$, 世界坐标系下的任意三维点可表示为

$$P_i=\sum_{j=1}^{4}\alpha_{ij}c_j,\quad \sum_{j=1}^{4}\alpha_{ij}=1 \tag{2}$$

通过齐次质心坐标 α_{ij}, 将上式可以写成齐次坐标形式

$$\begin{bmatrix} P_i^w \\ 1 \end{bmatrix}=\begin{bmatrix} c_1 & c_2 & c_3 & c_4 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \alpha_{i3} \\ \alpha_{i4} \end{bmatrix} \tag{3}$$

EPnP 方法引入了控制点 c_j^c, 即任何 1 个 3D 点都可以表示为 4 个控制点的线性组合. 第 1 个控制点为实体所有 3D 点的质心位置, 其余 3 点选在数据的主方. 在相机坐标系下的坐标 c_j^c 为

$$c_j^c=[R\,|T]\begin{bmatrix} c_j \\ 1 \end{bmatrix} \tag{4}$$

则对于任意匹配的空间三维点 P_i, 相机投影模型可重写为如下形式

$$s\begin{bmatrix} \boldsymbol{u}_i' \\ 1 \end{bmatrix}=K\,[R\,|T]\,P_i^c=$$

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}[R\,|T]\sum_{j=1}^{4}\alpha_{ij}\begin{bmatrix} X_j^c \\ Y_j^c \\ Z_j^c \\ 1 \end{bmatrix} \tag{5}$$

根据相机投影模型 (5), 在不考虑外参数矩阵 $[R\,|T]$ 的情况下, 求解控制点在相机坐标系下的坐标, 消去式 (5) 中最后一行.

$$\begin{cases} \displaystyle\sum_{j=1}^{4}\left(\alpha_{ij}f_x X_j^c+\alpha_{ij}\left(c_x-x_i\right)Z_j^c\right)=0 \\ \displaystyle\sum_{j=1}^{4}\left(\alpha_{ij}f_y Y_j^c+\alpha_{ij}\left(c_y-y_i\right)Z_j^c\right)=0 \end{cases} \tag{6}$$

其中除相机内参数 f_x、f_y、c_x、c_y, 以及已知匹配点坐标 $u_i=(x_i,\ y_i)$ 外, 4 个控制点共计 12 个未知参数. 根据中间变量法以及近似线性化的约束, 将所有 N 个特征点作为输入, 可以解得 P_0, 得到质心坐标. 于是, 问题就简化为点匹配的 ICP (Iterative closest point) 问题, 即将质心偏移矩阵 SVD 分解, 得到本时刻的旋转和平移矩阵 $[R_t|T_t]$, 详细计算过程参见文献 [54].

根据所获得的平移和旋转矩阵, 则本运动区域在当前时刻的速度 $\boldsymbol{v}_t$, 沿第 1.3.2 节约定的 X、Y、Z 轴三方向的分量分别为 $(v_t^x,\ v_t^y,\ v_t^z)$, 则

$$\boldsymbol{v}_t=\left(FT_t^x,\ FT_t^y,\ FT_t^z\right) \tag{7}$$

对于已得到的速度向量, 根据中心差分, 设 $\boldsymbol{a}_t$ 三方向分量为 $(a_t^x,\ a_t^y,\ a_t^z)$, 则

$$\boldsymbol{a}_t=\left(F\left(v_{t+1}^x-v_{t-1}^x\right)/2,\ F\left(v_{t+1}^y-v_{t-1}^y\right)/2,\right.$$
$$\left.F\left(v_{t+1}^z-v_{t-1}^z\right)/2\right) \tag{8}$$

由式 (2)~(8), 本文获得了候选目标的平动参数.

对于转动参量, 由罗德里格斯变换, 相邻两帧的旋转角度 φ 为

$$\varphi=\arccos\left(\frac{1}{2}\left[\mathrm{tr}\left(R_t\right)-1\right]\right) \tag{9}$$

其中, $\mathrm{tr}\left(R_t\right)$ 为矩阵 R_t 的迹. 则根据式 (9), 转轴 $\boldsymbol{u}=(u_x,\ u_y,\ u_z)$ (单位向量) 的反对称矩阵 $[\boldsymbol{u}]_x$ 为

$$[\boldsymbol{u}]_x=\left[\begin{array}{ccc}0 & -u_z & u_y\\ u_z & 0 & -u_x\\ -u_y & u_x & 0\end{array}\right]=\frac{1}{2\sin\varphi}\left(R-R^{\mathrm{T}}\right) \tag{10}$$

其中, R^{T} 为 R 的转置. 则角速度 $\boldsymbol{\omega}_t$ 可表示为

$$\boldsymbol{\omega}_t=F\varphi\cdot(u_x,\,u_y,\,u_z) \tag{11}$$

基于此, 采用中心差分, 角加速度可表示为

$$\boldsymbol{\alpha}_t=F^2\left(\varphi_{t+1}-\varphi_{t-1}\right)(u_x,\,u_y,\,u_z)/2 \tag{12}$$

至此就得到了目标所有的一阶与二阶运动参量.

1.3.3　运动参量决策树

至此, 用于描述物体运动特征的一阶与二阶运动参量都已获得. 在本节中, 本文将基于运动参量建立无人机识别模型. 由于参量数量较多且相互关系复杂, 直接估计每一类别关于运动参量的后验概率较为困难 (其中, 参数分别为目标的速度、加速度、角速度以及角加速度). 因此, 本文方法将识别过程分解为短期和长期两个步骤. 短期预测以快速检测为目的, 对于输入视频完成实时处理, 适用于实时性要求较高的场景. 长期识别以高精度检测为目的, 当在视频时长足够的情况下, 确保算法具备较高的识别精度.

其中, 梯度提升树 (Gradient boosting decision tree, GBDT) 完成短时识别并筛选关键运动参量. 而具有更高识别精度的 LSTM 网络则被用来完成长期识别. 该方式能够根据需求选择不同的针对性方法, 在实时性和高精度之间保持较好的平衡.

当前帧以及相邻前两帧的所有运动学参数共计 36 个, 以这些参数作为分量建立描述这些参数的 GBDT 分类树. 选择 CART 树作为弱分类器, 采用交叉熵作为损失函数

$$L\left(X_k,\,f_k\left(\boldsymbol{\rho}\right)\right)=-\sum_{k=1}^{K}X_k\ln p_k\left(\boldsymbol{\rho}\right) \tag{13}$$

其中, $X_k=\{0,\,1\}$ 表示是否属于第 k 类, 1 表示是, 0 表示否, $k=1,\,2,\,\cdots,\,K$ 为总共类别数 (本文类别包含四旋翼无人机、行人、车辆、鸟类、以及其他, 共 5 类, 则 $K=5$), $\boldsymbol{\rho}=(\boldsymbol{v},\,\boldsymbol{a},\,\boldsymbol{\omega},\,\boldsymbol{\alpha})$ 为所采用的运动参数. $p_k\left(\boldsymbol{\rho}\right)$ 为该样本属于每个类别的概率

$$p_k\left(\boldsymbol{\rho}\right)=\exp\left(f_k\left(\boldsymbol{\rho}\right)\right)\Big/\sum_{i=1}^{K}\exp\left(f_i\left(\boldsymbol{\rho}\right)\right) \tag{14}$$

对于输入的训练 $T=\{(\boldsymbol{\rho}_1,\,X_1),\,\cdots,\,(\boldsymbol{\rho}_N,\,X_N)\}$ 的每个样本 $i=1,\,2,\,\cdots,\,N$, 其伪残差为

$$r_{k,\,i}=-\partial L\left(X_k,\,f\left(\boldsymbol{\rho}_i\right)\right)/\partial f\left(\boldsymbol{\rho}_i\right)=X_{k,\,i}-p_k s\left(\boldsymbol{\rho}_i\right) \tag{15}$$

利用 $(\boldsymbol{\rho}_i,\,r_{k,\,i})\,(i=1,\,2,\,\cdots,\,N)$ 拟合一棵分类 CART (Classification and regression tree) 树[48], 得到第 $m=1,\,2,\,\cdots,\,M$ 棵树共 J 个叶子节点权值的最佳负梯度拟合值为

$$c_{m,\,k,\,j}=\left(1-1/K\right)\left(\sum_{\boldsymbol{\rho}_i\in R_{m,\,k,\,j}}r_{k,\,i}\right)\Big/\sum_{\boldsymbol{\rho}_i\in R_{m,\,k,\,j}}\left|r_{k,\,i}\right|\left(1-\left|r_{k,\,i}\right|\right) \tag{16}$$

则更新强分类器为

$$F_{k,\,m}\left(\boldsymbol{\rho}\right)=F_{k,\,m-1}\left(\boldsymbol{\rho}\right)+\sum_{i=1}^{J}c_{m,\,k,\,j}1\left(\boldsymbol{\rho}\in R_{m,\,k,\,j}\right) \tag{17}$$

其中 $F_{k,\,m}\left(\boldsymbol{\rho}\right)$为更新得到的强分类器, $1\left(\boldsymbol{\rho}\in R_{m,\,k,\,j}\right)$ 表示当前参数节点是否属于本次迭代残差构建的 CART 树的叶节点.

当残差满足一定数值或达到迭代次数时, 决策树构建完成. 第 2.5 节将分析设置不同参数对识别性能的影响.

2　实验与结果分析

本文所提出的多阶运动参量识别方法 (Multi-order kinematic parameters based detection method, MoKiP) 需要目标存在足够长的运动行程以提取运动学参数, 并要求输入视频尽可能达到较高的帧率. 因此, 本文采集补充了以往数据集中缺失的若干常见场景数据, 共同形成无人机多阶运动参量数据集 (Multi-scale UAV dataset, MUD). 本文将在该数据集上分析 MoKi 算法的有效性并从识别精度上与以往方法进行对比, 得出本文方法的优缺点.

2.1　实验环境

本文实验所使用数据包括两部分: 1) 公开数据集; 2) 本文采集的近地 UAV 数据集. 本文将获得的视频以是否包含地面分为两大类, 一类是近地场景, 一类是对空场景. 近地场景的视频中包含部分地面以及地面物体, 如建筑、植物等; 对空场景的视频中, 背景完全为天空, 不涉及地面部分.

公开数据集包括 AVSS2017 无人机识别挑战数据集 (Drone-bird dataset, DBD)[23] 以及运动相机飞行器探测数据集 (UAV-aircraft dataset, UAD)[15]. 其中, DBD 包含 6 段对空无人机飞行的视频, 共 2130 帧, 背景简单, 干扰目标为鸟类; UAD 包含 20 段无

人机和飞行器近地飞行的视频, 共 4000 帧, 背景相对复杂, 但无干扰运动目标.

以上数据集中考虑了四旋翼无人机外观和光照的多样性. 但对于无人机常见飞行场景来说, 其他影响识别的重要因素, 如多种干扰目标、目标尺度多样性、不同遮挡程度、不同背景复杂程度等都未得到体现, 因此不能充分反映无人机在日常场景中的飞行特性.

所以, 在此基础上, 本文以目标尺度为依据, 补充了室内场景、城市场景以及部分野外场景的无人机飞行视频. 新加入的数据集不仅包含目标类别和矩形框等简单标注, 而且还标记了目标的深度信息、飞行高度、相机拍摄角度以及运动参数等, 形成多尺度无人机数据集 (Multi-scale UAV dataset, MUD).

干扰目标的数据来自于 KITTI 车辆检测数据集[55]、RGB-D Pedestrian 行人检测数据集[56]、MoveBank[57] 和 NABirds 鸟类飞行探测数据集[58]. 表 1 对比了本文采集的数据、标注情况, 以及常见数据集. 本文所采集数据的部分图片如图 4(c) 所示, 所涉及的主要采集设备以及参数如表 2 所示.

本文所采集数据集相比于以往无人机和常见数据集, 增加了姿态、深度、视角、遮挡以及误差的标注信息, 其中误差信息为采集设备的误差. 本文所涉及的目标以及干扰目标为无人机、行人、车辆、鸟类. 此外, 为了更好的使用这些数据, 本文以是否包含地面为标准, 将这些视频重新组织为两大类, 一类称为近地场景, 即图像中包含部分地面以及地面物体, 如建筑、植物等; 一类是对空场景, 背景完全为天空, 基本不包含可识别的地平线特征.

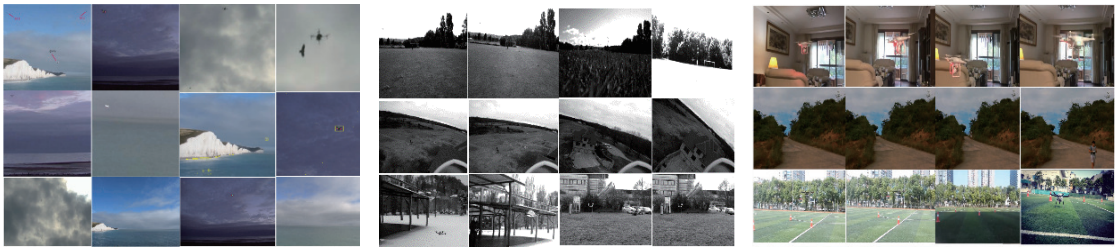
2.2 运动目标区域提取结果

作为本文方法流程的第 1 步, 根据第 1.3.1 节的方法, 采用运动目标提取 Vibe+[30] 法中设置参数前景孔洞最小尺寸为 5 像素, 每像素样本量为 10, 其余参数与文献 [30] 中保持一致. 得到提取运动目标区域以及相机运动补偿结果如图 5.

图 5 为室外/室内两个场景下运动区域提取结果. 其中标出的矩形窗口为获得的待识别目标的区域. 由于目标为获得更高的召回率, 或在召回率相近的情况下, 提取出的运动区域更少. 所以, 为对比不同运动提取方法, 所有采集视频被划分为固定长度片段, 并以召回率、矩形框数量以及单位数量区域下的目标召回率 (单位召回率) 为指标, 对比不同

表 1 本文所采集数据与其他运动目标数据集的对比
Table 1 Comparison of different datasets for moving objects

属性	本文所采数据	Drone-vs-Bird[23]	运动相机数据集[15]	Pascal3D+ 数据集[59]	NYU数据集[60]
目标类别数	5	2	2	12	894
平均每类视频帧数	3000	1500	3000	3000	39
场景	室内/室外	室外	室外	室内/室外	室内
背景单一程度	多背景	单一	单一	多背景	多背景
姿态标注	√	×	×	√	×
深度标注	√	×	×	×	√
多视角覆盖	√	×	√	√	×
遮挡标注	√	×	×	√	√
位置姿态误差	√	×	×	×	×



(a) Drone-vs-bird 数据集
(a) Drone-vs-bird dataset (DBD)

(b) UAV-aircraft 数据集
(b) UAV-aircraft dataset (UAD)

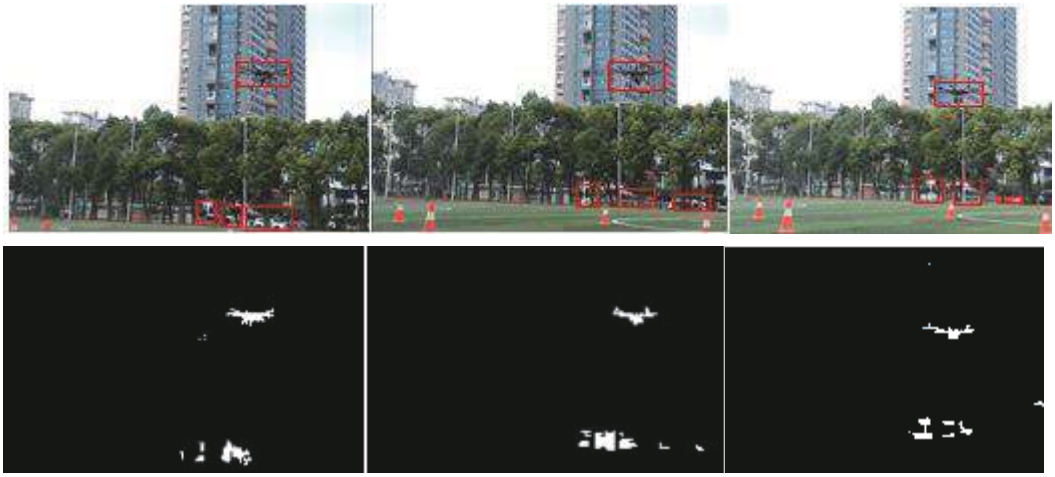
(c) MUD 数据集 (本文)
(c) MUD dataset (ours)

图 4 本文实验所用数据集示意图
Fig. 4 Illustration of parts of MUD used in our work

表 2　MUD 数据集采集设备说明

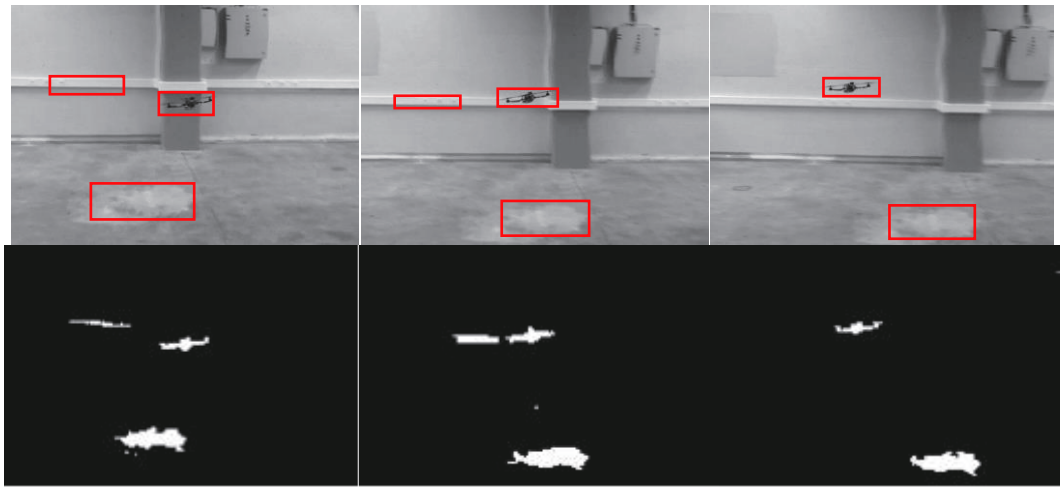
Table 2　Main equipment for acquisition of multi-scale UAV dataset (MUD)

设备	参数	精度
相机	SONY A7 ILCE-7M2, 6 000 × 4 000 像素FE 24 ~ 240 mm, F 3.5 ~ 6.3	—
GPS	GPS/GLONASS双模	垂直 ±0.5 m, 水平 ±1.5 m
激光测距仪	SKIL Xact 0530, 0 ~ 80 m	±0.2 mm



(a) 城市场景

(a) Urban scene



(b) 室内场景

(b) Indoor scene

图 5　运动目标区域提取结果图

Fig. 5　Extraction result of motion ROIs

方法得到下表 3.

从 3 个指标来看, Vibe+法具有更高的召回率. 帧差法和光流法虽召回率也较高, 但召回的假目标较多, 单位召回率较低; 混合高斯法召回率较低. 所以对于随后的处理和识别, 本文采用 Vibe+法作为待识别区域提取的方法, 并采用高斯混合概率假设密度滤波算法 (Gaussian mixture-probability hypothesis density, GM-PHD) 方法对其进行跟踪, 算法实施过程中目标检测概率、目标生存概率、平均杂波数、高斯元门限值等参数与文献 [62] 中保持一致.

2.3　目标深度提取

按第 1.3.2 节所述, 利用 DORN 方法估计场景深度. 对 MUD 数据集中街道场景中某帧计算得到的深度图如图 6 所示. 其中深度真值为数据集中的

表 3 运动目标区域提取算法性能对比
Table 3 Comparison between performance of different motion ROIs

方法	矩形框数量	召回率	单位召回率(每百个)
帧差法[31]	413	0.832	0.201
混合高斯法[27]	315	0.784	0.249
光流法[61]	521	0.853	0.164
Vibe+法[30]	238	0.868	0.365

标注值或通过激光测距的测量值.

图 6 中, (a) 为输入图像, (b) 为深度估计结果, (c) 为目标区域深度真值. 其中白色区域为深度超过 100 米的位置, 可以认为无穷远背景, 不予计算; 对比估计结果与真值图可以看出, 对于远处行人以及空中无人机, 深度的估计结果与真值相符, 以下将具体给出该算法以及其他算法的估计误差.

表 4 以不同估计误差参数对比了目前不同深度估计方法在本文所采用数据中的估计精度, 其中误差参数定义如文献 [51]. 其中误差项参数越小精度越高, 涉及 δ 的误差项越大精度越高. 从表中可以看出, 相比于激光测距的精度, 深度估计的算法误差随对探测距离的增加而明显增大. 多目视觉方法在近距离的深度测量中具备较高精度以及较低的时间复杂度, 但随深度的增加深度测量精度严重下降, 更适合于在室内场景中使用. DORN 方法具有当前方法中最佳的精度和鲁棒性.

深度估计的方法亦可根据场景和需求选择其他方法.

2.4 运动参量提取

根据以上小节所得到的运动区域深度值, 利用第 1.3.2 节中的估计方法, 可以提取出运动区域中的运动参数 $\boldsymbol{\rho} = (\boldsymbol{v},\ \boldsymbol{a},\ \boldsymbol{\omega},\ \boldsymbol{\alpha})$, 对于各个运动参量的定义和坐标设定与第 2.3 节中保持一致. 运动参数以及姿态真值由无人机自带的实时动态差分系统 (Real-time kinematic, RTK) 以及运动捕捉系统 (Motion capture system, MCS) 给出. 其测量精度为: 当 GPS 正常工作时, 垂直方向定位精度为 ±0.5 m, 水平方向定位误差为 ±1.5 m; 若视觉定位也正常工作, 则垂直方向定位精度为 ±0.3 m, 水平方向定位误差为 ±0.5 m.

从图像中提取的参数, 其与标定值误差如图 7 所示. 表 5 中为图 7 各参数的说明. 图 7 中分别为目标速度、加速度、角速度、角加速度的三方向分量的估计误差以及空间定位误差, 其中红色标点为典型异常值, 右表为图中各符号说明. 图中, 采用本文

(a) 输入图像
(a) Input image

(b) 深度估计结果
(b) Depth estimation result

(c) 深度真值
(c) Ground truth

图 6 深度估计结果图
Fig. 6 Result of depth estimation

表 4 不同深度估计方法误差对比
Table 4 Error of different depth estimation methods

方法	探测范围	绝对误差	平方误差	均方根误差	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
DORN[51]	0 ~ 100 m	0.103	0.321	9.014	0.832	0.875	0.922
GeoNet[63]	0 ~ 100 m	0.280	2.813	14.312	0.817	0.849	0.895
双目视觉[64]	0 ~ 100 m	0.062	1.210	0.821	0.573	0.642	0.692
激光测距	0 ~ 200 m	0.041	2.452	1.206	0.875	0.932	0.961
DORN[51]	200 ~ 500 m	0.216	1.152	13.021	0.672	0.711	0.748
GeoNet[63]	200 ~ 500 m	0.398	5.813	18.312	0.617	0.649	0.696
双目视觉[64]	200 ~ 500 m	0.786	5.210	25.821	0.493	0.532	0.562
激光测距	200 ~ 500 m	0.078	3.152	2.611	0.891	0.918	0.935

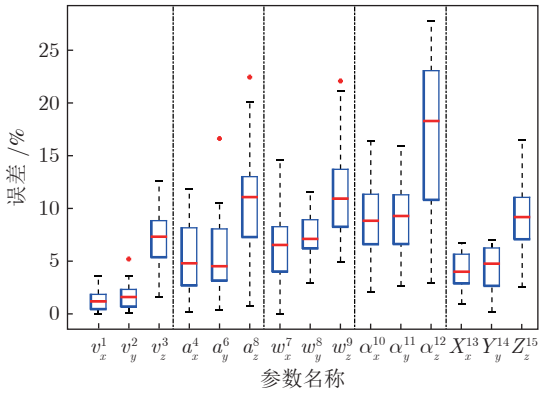


图 7　运动参量估计误差箱图

Fig. 7　Boxplot for motion parameter error estimation

表 5　图 7 中参数对照表

Table 5　Illustrations of parameters in Fig. 7

参数	说明
$\boldsymbol{v}$	速度
$\boldsymbol{a}$	加速度
$\boldsymbol{\omega}$	角速度
$\boldsymbol{\alpha}$	角加速度
X 轴分量方向	与图像平面坐标系中 u 轴方向保持一致
Y 轴分量方向	与图像平面坐标系中 v 轴方向保持一致
Z 轴分量方向	铅垂向上

方法估计得到的速度、加速度、角速度、角加速度的最小误差、最大误差以及平均误差百分数分别为: 0.0, 12.7, 3.4 (速度参量); 0.1, 20.1, 6.9 (加速度参量); 0.0, 21.2, 8.3 (角速度参量); 2.1, 27.8, 12.1 (角加速度参量). 从整体来看, 平动参量 (速度、角速度) 估计误差低于转动参量 (角速度、角加速度), 估计精度更高; 一阶参量 (速度、角速度参量) 的 X、Y、Z 三方向分量估计精度相比于相应的二阶参量具有更高的精度. 运动参量在 X、Y 轴方向分量的估计精度相比于 Z 轴相应参量分量的精度更高, 估计误差的标准差也更小, 所有参量中, 速度参量的 X、Y 轴分量的估计精度最高, 误差在 5% 以下; 角加速度参量的 Z 轴分量误差最大, 约为 20%.

由于平动为无人机运动的主要方式, 反映在图像中, 目标的特征点在帧间产生明显的位移, 定位的偏差相对于目标的位移相对较小, 所以估计误差相对较小. 而因转动产生的特征点位移较小, 对特征点定位精度敏感, 定位误差产生的转动参量估计的偏差会更大.

另外, 二阶参量估计是在一阶参量基础上完成的, 所以一阶参量的估计误差会累积到二阶参量的估计中, 导致二阶参量的估计误差更高.

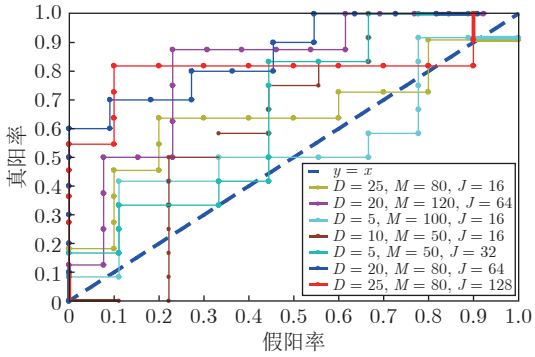
2.5　基于运动参量的无人机识别方法结果与分析

2.5.1　基于运动参量的梯度提升树模型的识别结果

本文采用决策树模型, 利用运动学参数进行目标识别. 根据文献中常用参数搭配[48] 通过网格搜索法 (Grid search) 选择较优的参数组合, 设置不同的决策树深度 D、决策树数量 M (弱学习器最大数量)、叶子结点数量 J, 获得训练集上识别精度最高的参数组合. 在不同参数取值下得到无人机分类器判决接收者操作特征曲线 (Receiver operating characteristic curve, ROC) 如下. 其横坐标为假阳率 (False positive rate), 纵坐标为真阳率 (True positive rate), 用于评价模型的判决能力. ROC 曲线下面积 (Area under curve, AUC) 值在 0 ~ 1 之间. 越大其分类正确率越高.

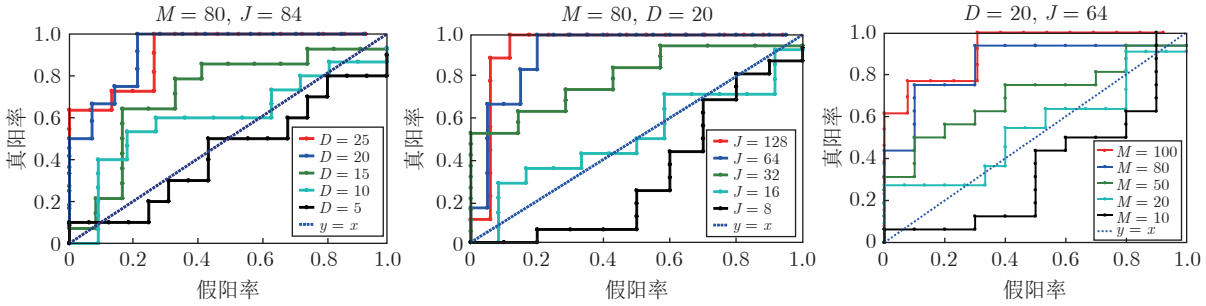
如图 8(a)、(b) 为设置不同的 D、M、J 参数时, 梯度提升树训练后的 ROC 曲线. 图 8(a) 中为按训练后分类器的性能得到若干典型参数组合; 图 8(b) 中为获得固定 D、M、J 参数中的两个时, 另外一个单一变量对 ROC 曲线的影响, 用以筛选出最优参数. 在图 8(b) 中, 当固定 M、J 参数时, 随 D (决策树深度) 的增大, ROC 曲线上移, 说明分类器的准确率上升, 但当增加到 $D = 20$ 后, 再增大 D, ROC 曲线不再上移, 说明该分类器接近性能上限; 当固定 D、J 参数时, 随 M (决策树数量) 的增大, ROC 曲线持续上移, 曲线下面积 (Area under curve, AUC) 从 0.580 上升至 0.825, 但上升的幅度越来越小, $M = 80$ 至 $M = 100$ 相比于之前相邻曲线面积增加的 0.134, 下降至 0.065, 增长率从 16.2% 下降至 6.7%; 当固定 M、D 增加 J (叶子节点数量) 时, ROC 曲线仍上移, 但当增至 $J = 128$ 时, 增长率相比于之前增长率下降至 7.1%. 总结图 8(a)、(b) 中各参数组合的 ROC 曲线, 为尽量保证训练时分类器具备较高精度, 并防止过拟合现象, 最终梯度提升树的参数值为 $D = 20$, $M = 80$, $J = 64$ (其曲线下面积 AUC 值为 0.812). 学习率设置为 0.6, 子采样系数设置为 0.8, 损失函数为对数损失. 以下是在该参数组合下, 不同场景目标的识别结果.

图 9 为室内、对空、以及低空野外 3 个不同场景下的识别结果示意. 场景中除目标外, 同时还包括本文所涉及的主要干扰目标, 包括鸟类、行人、车辆和其他干扰目标. 不同目标以不同颜色予以标识, 并给出识别结果以及类别概率. 从后两段识别结果来看, 即使外观特征不显著的情况下, 本方法也能够在运动过程中动态辨识目标, 类别概率会随着目标的运动而变化. 当出现典型的运动方式时, 符合



(a) 几种较优参数组合的 ROC 曲线

(a) ROC curves of several competitive parameter combinations



(b) 单参数变化时的 ROC 曲线(左中右分别为: 单独变化)

(b) ROC curves of single changing parameter (from left to right: respectively)

图 8 不同参数组合的 ROC 曲线单参数变化时的 ROC 曲线 (左中右分别为 D、M、J 单独变化)

Fig. 8 ROC curves of different GBDT parameter combinations (The subplots from left to right are corresponding to D、M、J respectivly)

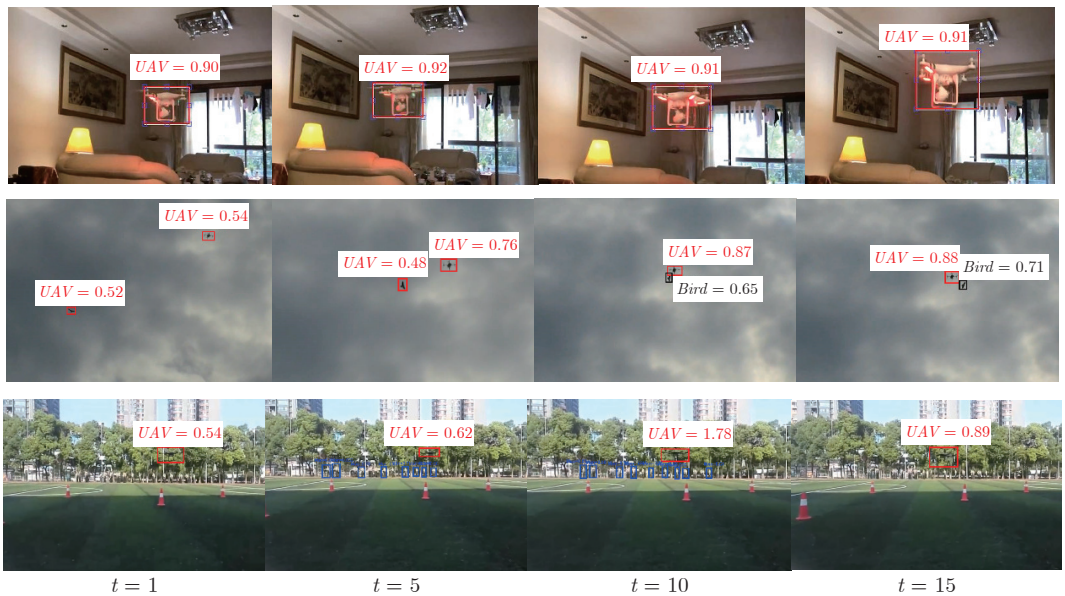


图 9 基于运动参量决策树的无人机识别结果

Fig. 9 Results of MoKiP by using GBDT

该运动方式的目标类别概率会明显上升, 错误的类别概率就会逐渐下降, 当类别概率超过 50% 时, 则框出该目标为此类别. 为消除系统累积误差, 本方法将在每 20 秒初始化 1 次.

训练得到的包括无人机、鸟类、行人、车辆以及其他类别的多分类器, 其混淆矩阵如表 6 所示. 其中数字表示预测正确的样本所占该类样本总数的比例. 从表中可以看出, 无人机、行人、车辆的识别精度较高; 鸟类的识别精度最低, 混淆率较高, 更容易与无人机以及其他物体飞行物体所混淆. 相比于鸟类, 无人机的识别精度更高, 不易被其他飞行物体所干扰, 但其主要干扰目标仍为鸟类. 行人和车辆识别精度最高, 主要由于其运动复杂度低、运动变化少, 运动特性明确. 总的来说, 根据运动参量决策树对本文涉及的类别识别正确率 (对角线数值) 均能达到 0.55 以上.

表 6　运动参量的决策树模型识别结果混淆矩阵

Table 6　Confusion matrix of MokiP by using GDBT

真实值 预测值	旋翼无人机	鸟类	行人	车辆	其他物体
旋翼无人机	**0.67**	0.25	0.02	0.01	0.12
鸟类	0.21	**0.58**	0.01	0.00	0.10
行人	0.01	0.02	**0.75**	0.06	0.09
车辆	0.01	0.00	0.10	**0.80**	0.08
其他物体	0.10	0.15	0.12	0.13	**0.61**

2.5.2　与以往方法对比分析

为对比本方法与以往不同方法的识别性能, 本节以 PR 曲线 (Precision-recall curve) 和识别精度 AP (Average precision) 值为指标[11−17], 给出本方法与目前几种主要 “低慢小”运动目标识别方法的性能对比, 如图 10 所示.

在图 10 绘出了包括本文方法在内的多种目前具备较优性能的 “低慢小”识别方法在所述 MUD 数据集上的识别 PR 曲线, 所涉及方法为: 基于深度光流特征的 FlowNet 2.0[34] 运动特征方法、Xiao 等[42] 采用 ConvGRU(RNN) 结构融合时空特征的混合方法、Schumann 等[17−18] 基于 Faster RCNN[12] 并根据四旋翼无人机训练的改进表观特征方法、Luo 等[41] 引入语义信息关联相邻帧目标框的时空特征混合方法 (ImageNet VID 数据集中目标的最优方法 SOTA)、Rozantsev 等[15] 引入相机补偿的改进 Faster RCNN 方法 (四旋翼无人机目标识别的最优方法 SOTA) 以及本文 MoKiP 方法采用 GBDT、GRU 神经网络的两种实现和融合方法. 图中每条 PR 曲线绘出随某算法召回率上升时, 准确率的变化情况. 每条曲线头部保持平直, 准确率基本保持不变, 保持在高准确率; 当到达转折点时开始下降, 尾部为下降过程.

图 10 中曲线头部为识别方法能够达到的最高准确率. 对于本文所研究的目标和数据集, 以往基于运动特征的最优方法 FlowNet2.0[34] 的最高准确率为 0.68, 基于表观特征的最优方法 Luo[41] 的最高准确率为 0.83, 混合方法中的最优方法, 即目前最优方法 Rozantsev[15] 的最高准确率为 0.88. 本文基于运动特征的方法 (非零阶运动参量) 最高准确率为 0.76, 混合方法最高为 0.92. 通过对运动特征的充分提取和细化, 相比与以往基于运动和混合方法, 本文方法在最高准确率 (曲线头部部分) 分别有 0.06 和 0.04 的提升. 但是基于非零阶运动参量的方法相对于以往表观识别方法[18, 41], 曲线头部准确率有 0.06 左右的下降. 这是由于, 在低召回率的情况下,

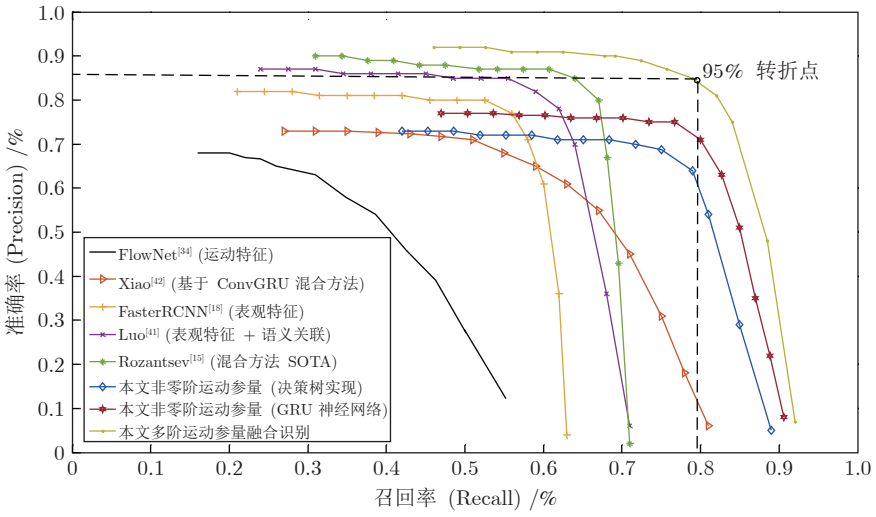


图 10　不同识别方法的性能对比图

Fig. 10　Comparison of performance for different detection methods

最先被召回的目标主要是像素量高、细节丰富的目标, 直接利用深度网络的表观识别方法精度更高,

图 10 表示出本文多阶运动参量融合识别方法 PR 曲线在准确率下降到 95% 时的转折点位置. 结合表 7, 根据 95% 转折点位置可以看出, 以往方法最大值位置为 Recall = 0.57 (方法 [42]), 本文基于非零阶和多阶运动参量的 95% 转折点分别为 Recall = 0.73 和 Recall = 0.82. 以往方法的 PR 曲线 (方法 [15, 18, 34, 41−42]) 都位于本文方法 PR 曲线的左侧, 说明本文方法随召回率升高仍能保持较高的准确率, 鲁棒性强. 在高召回率时主要为以往方法难以识别的困难目标, 其主要表现为微像素量 (目标总像素量少于 150)、外观呈现形式多样、遮挡部分多, 往往出现在多目标干扰的复杂背景中. 但这些在表观特征中的识别困难, 在本文通过运动参量形成的运动特征空间中, 不同目标的运动模式差距较大, 而同类目标即使外观的千差万别, 却具备相似的特征运动模式.

表 7 不同识别方法性能指标对比表

Table 7 Comparison of performance indexes for different detection method

方法	AP精度	95%转折点	曲线尾部梯度	AP50	AP90
FlowNet[33]	32.2	0.30	**2.87**	42.0	10.3
IRRCNN[24]	36.7	0.37	10.18	50.9	7.2
Xiao[42]	50.3	0.55	**2.34**	59.3	19.7
Faster RCNN[18]	47.8	0.57	11.07	62.1	18.5
Luo[41]	57.2	0.59	7.70	71.7	24.4
Rozantsev[15]	62.1	0.65	14.10	**81.3**	37.2
本文非零阶参数方法(GRU)	65.6	0.78	5.34	79.5	39.8
本文多阶运动参量方法	78.5	0.80	6.54	91.2	46.8

PR 曲线 95% 转折点后下降部分为其尾部. 本文采用尾部下降梯度参数衡量识别方法退化速度, 即由 95% 转折点下降至准确率为 0.1 位置连线的斜率, 见表 7. 以往工作中具备较高精度的神经网络方法, 包括文献 [15, 18, 24], 尾部下降梯度大于 10, 当在网络中加入语义以及时空约束后, 方法 [41] 尾部梯度为 7.70, 相比于直接采用神经网络方法梯度下降较小. 以运动特征为基础的方法[34] 以及以 RNN 为基础框架的方法[42], 尾部梯度分别为 2.87 和 2.34, 具备最缓的下降速度, 鲁棒性强. 本文基于非零阶运动参量的方法和多阶运动参量方法尾部梯度分别为 5.34 和 6.54, 为文献中最优方法[15](尾部梯度为 14.10) 的 40% 左右, 下降速度更慢. 这说明随着目标识别困难的增加, 本文识别方法退化速度慢, 鲁棒性强.

以下, 本文从具体指标上, 对比了不同 “低慢小”识别方法, 如表 7 所示.

表 7 比较了本文与以往文献方法在 AP 精度 (AP50、AP90)、95% 转折点、尾部梯度等参数上的性能差异, AP 精度数值皆为百分数. 其中, 文献 [34] 为基于运动特征的识别方法; 方法 [18, 24, 41] 是以深度卷积网络为框架的基于表观特征的识别方法, 文献 [42] 为以 RNN 为框架的混合识别方法; 文献 [15] 为融合运动与表观特征混合的最高精度方法. 从表中可以看出, 本文根据四旋翼无人机非零阶运动参量 (运动特征) 的识别方法相比于当前最优运动特征方法[34]、最优表观特征方法[41] 和最优混合方法[15] 分别提升 33.4 (103%)、8.4 (14%) 和 3.5 (5%). 本文融合表观特征和运动特征的多阶运动参量方法在 AP 识别精度上达到 78.5, 相比于当前具备最高精度的混合方法[15], 提升了 16.4 (26%). 进一步从 AP50、AP90 精度来看, 即当 IoU 阈值分别设为 50% 和 90% 得到的识别 AP 精度值, 当前最优方法[15] 仅在 AP50 指标上具备较高精度, 其余指标皆为本文方法更优. 总的来说, 从各项精度指标来看, 本文提出的多阶运动参量识别方法 (MoKiP) 对于四旋翼无人机目标相比于以往方法具备更高识别精度.

2.6 一阶、二阶以及多阶运动参量的识别显著性

为进一步分析不同运动参量对识别精度的影响, 本节对上文所获得决策树中的不同运动参量进行统计, 并分别使用不同的参量组合对目标进行识别, 最终得到不同运动参量对识别的敏感度分析.

图 11 是根据第 1.3.3 节所述过程以及第 2.5.1 节参数训练得到的一棵决策树, 红色和蓝色分别是根据一阶和二阶运动参量的预测分支.

每个节点包括本分支的样本总量、各类别样本数量、基尼系数以及判别条件.

对训练得到的所有决策树, 按照运动参量的阶数以及性质进行统计, 通过基尼系数衡量每一类运动参数对识别的重要程度, 则参量贡献度 D 被定义为:

$$D = Gini_{\boldsymbol{\rho}} / \sum Gini$$

$Gini_{\boldsymbol{\rho}}$ 为在所有训练得到的决策树中采用某类参量 $\boldsymbol{\rho}$ 为切分变量的节点的基尼系数. 按参量不同性质, 得到参量贡献度占比表, 见表 8、表 9 所示.

表 8 中所涉及的一阶参量包括速度、角速度, 二阶参量包括加速度、角加速度; 平动参量包括速度和加速度, 旋转参量包括角速度和角加速度. 其中一阶平动参量和旋转参量分别占 7.2% 和 20.1% 的贡献度; 二阶平动参参量和旋转参量分别占 34.1%

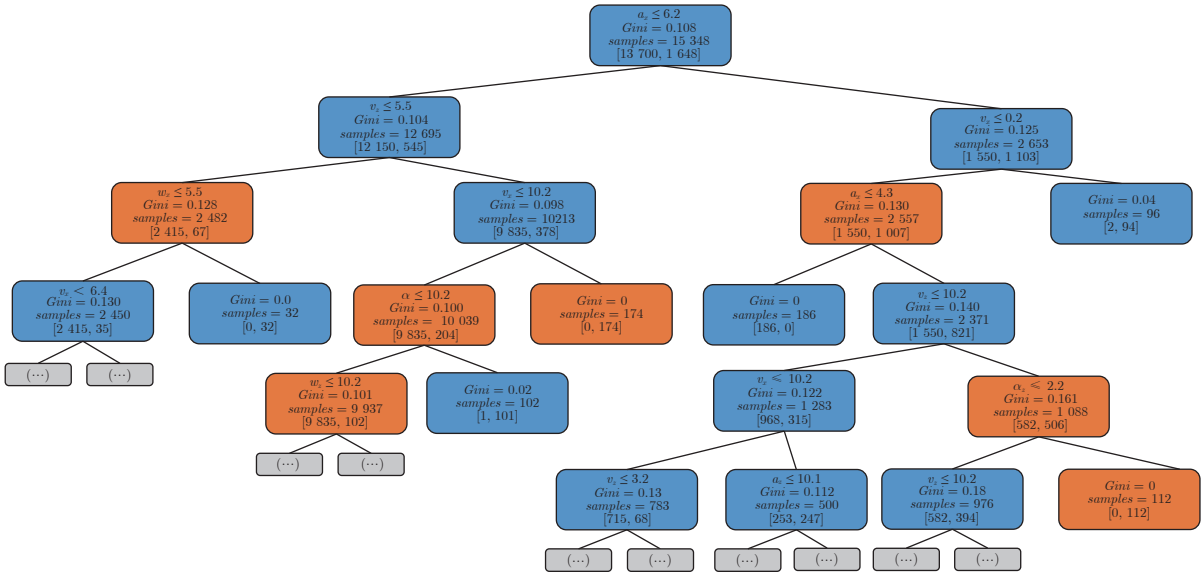


图 11　训练得到的梯度提升树示意图

Fig.11　A single tree from the trained GDBT

表 8　运动参量的性质对无人机识别的影响表

Table 8　Impact of the parameter properties on UAV detection

参量贡献度 D	平动参量	旋转参量	总贡献度
一阶参量	**7.2%**	20.1%	27.3%
二阶参量	34.1%	**38.6%**	**72.7%**
总贡献度	41.3%	58.7%	1

表 9　运动参量的方向对无人机识别的影响表

Table 9　Impact of the parameter direction on UAV detection

参量贡献度 D	沿 X 轴方向	沿 Y 轴方向	沿 Z 轴方向	总贡献度
平动参量	**8.3%**	8.8%	**24.2%**	41.3%
旋转参量	18.7%	18.8%	**22.2%**	58.7%
总贡献度	27.0%	27.6%	**46.4%**	1

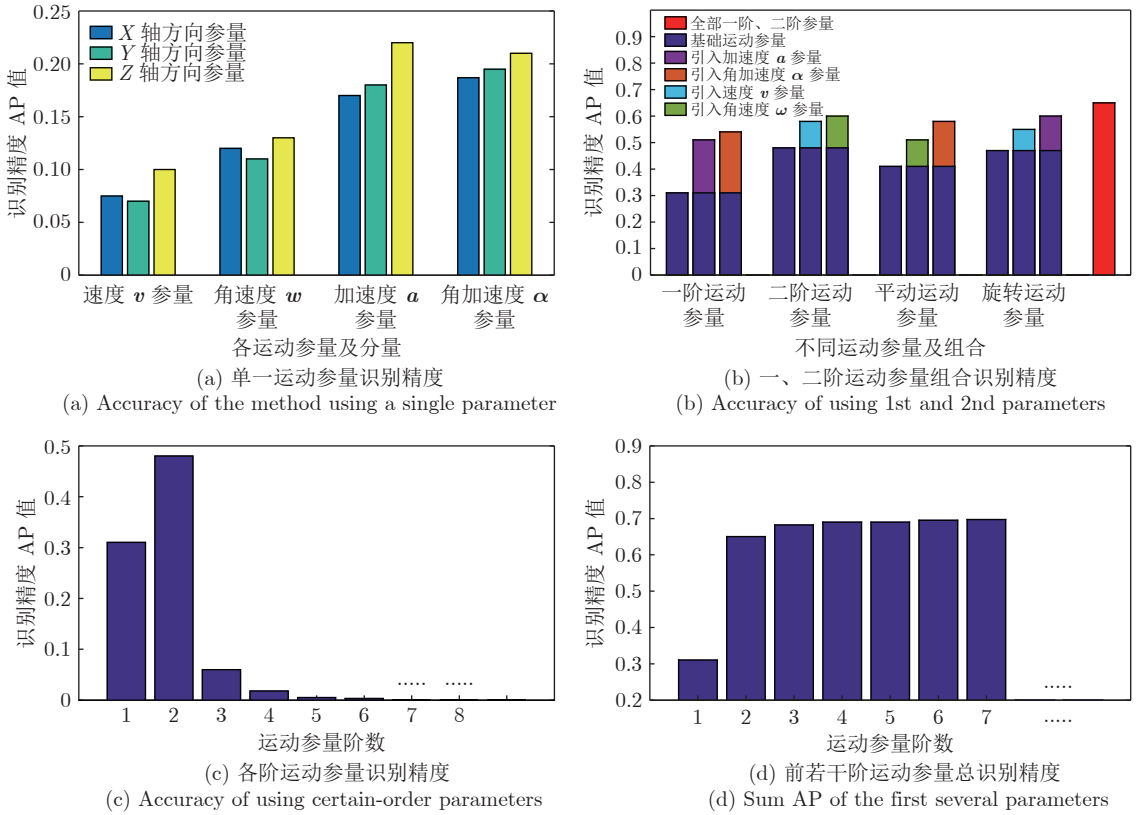
和 38.6% 的贡献度. 在表中所涉及的参量中, 一阶平动参量贡献度最低, 二阶旋转参量贡献度最高, 相同阶数的旋转参量比平动参量贡献度更高. 从总和来说, 二阶运动参量对本文提出模型的贡献度最大, 达到 72.7%; 一阶参数贡献度为 27.3%, 仅为二阶参量的 38%, 说明在识别过程中二阶运动参量起了更重要的作用. 从运动方式来说, 旋转参量相对于平移参量, 能更大程度上反映出目标的特征. 综上, 二阶参量是无人机识别过程中的重要参量, 精确估计二阶旋转参量是本文识别方法的基础.

表 9 显示了运动参量的方向性对识别的影响, 其中所述沿 X、Y、Z 轴方向参量与第 1.3.2 节保持一致. 表中沿 X 轴方向平动参量贡献度最小为 8.3%, 沿 Z 轴 (即重力方向) 方向的平移参量贡献度最大为 24.2%. 沿 X 轴和 Y 轴的运动参量在各贡献度数据上都相近, 并均低于相应 Z 轴方向运动参量, 其中平移参量在数值上低 15.4%, 旋转参量低 3.4%. 从总贡献度来看, Z 轴方向参量总贡献度比 Y 轴总贡献度高 68.1%. 这说明 Z 轴方向运动参量为识别过程中的主要参量, 沿 Z 轴方向的运动是无人机区别于其他目标的主要运动方式.

进一步, 本节通过剥离实验, 采用不同的运动参量组合对四旋翼无人机进行识别, 获得以下识别结果.

图 12 中, 图 12(a)、(b) 为采用本文所涉及的一、二阶运动参量组合进行识别的结果; 图 12(c)、(d) 为采用二阶以上高阶运动参量识别结果. 图中 X、Y、Z 轴方向与第 1.3.2 节中保持一致. 图 12(a) 为单独采用单一运动参量 (分量) 进行识别的结果, 其中蓝绿黄色分别代表沿运动参量 X、Y、Z 轴分量结果. 其中单独使用速度参量 Y 轴分量识别的精度最低, 为 0.07; 单独使用加速度 Z 轴分量识别的精度最高, 为 0.22. 单独使用二阶参量均高于单独使用一阶参量的识别精度; 单独使用 Z 轴方向分量的识别精度高于相应参量在 X、Y 轴分量的精度. 这也从另一侧面印证了表 7、8 的结论, 即二阶运动参量以及运动参量的 Z 轴分量是无人机识别过程中的重要参量.

图 12(b) 为采用本文所涉及的运动参量不同组合进行识别的精度对比. 其中, 基础运动参量为一阶、二阶、平动、旋转等参量组合, 再将不同的其他



(a) 单一运动参量识别精度
(a) Accuracy of the method using a single parameter

(b) 一、二阶运动参量组合识别精度
(b) Accuracy of using 1st and 2nd parameters

(c) 各阶运动参量识别精度
(c) Accuracy of using certain-order parameters

(d) 前若干阶运动参量总识别精度
(d) Sum AP of the first several parameters

图 12 不同参量组合的识别结果图

Fig. 12 Detection results of different parameter combinations

参量加入到基础运动参量中后, 得到不同参量组合的识别结果. 两参量组合的基础运动参量结果体现出与表 7 一致的结论, 使用二阶参量 (0.48) 相比于一阶参量 (0.31) 识别精度更高, 在识别中的贡献度更大, 为更重要的运动参量; 旋转参量 (0.47) 稍高于平动参量 (0.41) 的识别精度. 在三参量组合中, 组合识别精度最高为 0.60, 高单精度最低为 0.51. 在基本参量中, 相比于加入速度和角速度参量的最大提升 0.10 (24.3%), 加入加速度、角加速度参量的提升更大, 最小提升为 0.13 (27.6%). 进一步说明二阶参量在识别过程中为更重要的参量、更显著地影响识别效果.

图 12(c)、(d) 绘出了使用不同阶运动参量识别的结果. 图 12(c) 为单独使用某一阶运动参量识别的精度结果; 图 12(d) 为使用前若干阶运动参量识别的精度结果. 由于在运动学中描述物体运动均采用一阶和二阶运动参量, 所以本文也主要使用二阶及以下运动参量进行识别. 但从参数辨识和运动特征提取的角度来说, 由于采集得到的均是离散的数据, 想要尽可能精确地估计得到运动参量或者恢复目标整个运动过程, 只能基于近似的方法. 根据泰勒展开, 任意运动轨迹上的一点的位置均可由其在选定点多阶导数形成的多项式进行逼近, 而所涉及的多阶导数, 即为本文所涉及的一阶、二阶以及高阶运动参量. 同样, 对于整个转动过程也需要利用多阶旋转参量进行逼近. 另一方面, 越高阶的运动参量越能够反映目标在更长一段时间内运动的整体特征. 所以, 在识别过程中使用高阶运动参量是有必要的.

从图 12(c) 中看出, 单独使用一阶运动参量精度为 0.311, 单独使用二阶运动参量精度为 0.480, 单独使用三阶参量精度下降至 0.059, 四阶参量精度下降 69.5% 至 0.018, 到六阶的识别精度仅为 0.003, 说明三阶以上运动参量识别贡献度显著降低. 再结合图 12(d) 前若干阶参量总识别精度来看, 使用一、二阶参量识别精度为 0.656, 三阶参量引入后识别精度上升至 0.681, 增幅 3.8%; 四阶参量引入后, 增幅仅为 1.3%; 至六阶参量引入, 总识别精度为 0.697, 增长为 0.2%. 总的来说, 三阶以上总识别精度未有显著增长, 一、二阶运动参量能较完整的包含目标的全部运动特征.

3 结论

本文提出了一种基于运动参量建模的 “低慢小”

目标识别方法. 相比于以往方法, 本方法进一步完善了运动特征的描述, 并在所涉及的数据集上, 相比于以往文献中的方法, 显著地提升了四旋翼无人机的识别精度. 在实验中本文也发现, 二阶参量、旋转参量、以及重力方向的运动参量是四旋翼无人机识别过程中的重要参量, 反映出目标在运动模式上的差异.

References

1 Li Bo, Meng Li-Fan, Li Jing, Liu Chun-Mei, Huang Guang-Yan. Research on detecting and locating technology of LSS-UAV. *China Measurement & Test*, 2016, **42**(12): 64−69
(李波, 孟立凡, 李晶, 刘春美, 黄广炎. 低空慢速小目标探测与定位技术研究. 中国测试, 2016, **42**(12): 64−69)

2 Wang Z H, Lin X P, Xiang X Y, Blasch E, Pham K, Chen G S, et al. An airborne low SWaP-C UAS sense and avoid system. In: Proceedings of SPIE 9838, Sensors and Systems for Space Applications IX. Baltimore, USA: SPIE, 2016. 98380C

3 Busset J, Perrodin F, Wellig P, Ott B, Heutschi K, Rühl T, et al. Detection and tracking of drones using advanced acoustic cameras. In: Proceedings of SPIE 9647, Unmanned/Unattended Sensors and Sensor Networks XI; and Advanced Free-Space Optical Communication Techniques and Applications. Toulouse, France: SPIE, 2015. 96470F

4 Mezei J, Fiaska V, Molnár A. Drone sound detection. In: Proceedings of the 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI). Budapest, Hungary: IEEE, 2015. 333−338

5 Zhang Hao-Kui, Li Ying, Jiang Ye-Nan. Deep learning for hyperspectral imagery classification: The state of the art and prospects. *Acta Automatica Sinica*, 2018, **44**(6): 961−977
(张号逵, 李映, 姜晔楠. 深度学习在高光谱图像分类领域的研究现状与展望. 自动化学报, 2018, **44**(6): 961−977)

6 He Lin, Pan Quan, Di Wei, Li Yuan-Qing. Supervised detection for hyperspectral imagery based on high-dimensional multiscale autoregression. *Acta Automatica Sinica*, 2009, **35**(5): 509−518
(贺霖, 潘泉, 邸韡, 李远清. 高光谱图像高维多尺度自回归有监督检测. 自动化学报, 2009, **35**(5): 509−518)

7 Ye Yu, Wang Zheng, Liang Chao, Han Zhen, Chen Jun, Hu Rui-Min. A survey on multi-source person re-identification. *Acta Automatica Sinica*, 2020, **46**(9): 1869−1884
(叶钰, 王正, 梁超, 韩镇, 陈军, 胡瑞敏. 多源数据行人重识别研究综述. 自动化学报, 2020, **46**(9): 1869−1884)

8 Zhao J F, Feng H J, Xu Z H, Li Q, Peng H. Real-time automatic small target detection using saliency extraction and morphological theory. *Optics & Laser Technology*, 2013, **47**: 268−277

9 Nguyen P, Ravindranatha M, Nguyen A, Han R, Vu T. Investigating cost-effective RF-based detection of drones. In: Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use. Singapore: Association for Computing Machinery, 2016. 17−22

10 Drozdowicz J, Wielgo M, Samczynski P, Kulpa K, Krzonkalla J, Mordzonek M, et al. 35 GHz FMCW drone detection system. In: Proceedings of the 17th International Radar Symposium. Krakow, Poland: IEEE, 2016. 1−4

11 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

12 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137−1149

13 Dollar P, Tu Z W, Perona P, Belongie S. Integral channel features. In: Proceedings of the 2009 British Machine Vision Conference. London, UK: BMVA Press, 2009. 91.1−91.11

14 Aker C, Kalkan S. Using deep networks for drone detection. In: Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce, Italy: IEEE, 2017. 1−6

15 Rozantsev A, Lepetit V, Fua P. Flying objects detection from a single moving camera. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 4128−4136

16 Coluccia A, Ghenescu M, Piatrik T, De Cubber G, Schumann A, Sommer L, et al. Drone-vs-Bird detection challenge at IEEE AVSS2017. In: Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce, Italy: IEEE, 2017. 1−6

17 Schumann A, Sommer L, Klatte J, Schuchert T, Beyerer J. Deep cross-domain flying object classification for robust UAV detection. In: Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce, Italy: IEEE, 2017. 1−6

18 Sommer L, Schumann A, Müller T, Schuchert T, Beyerer J. Flying object detection for automatic UAV recognition. In: Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce, Italy: IEEE, 2017. 1−6

19 Sapkota K R, Roelofsen S, Rozantsev A, Lepetit V, Gillet D, Fua P, et al. Vision-based unmanned aerial vehicle detection and tracking for sense and avoid systems. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon, Korea: IEEE, 2016. 1556−1561

20 Carrio A, Vemprala S, Ripoll A, Saripall S, Campoy P. Drone detection using depth maps. In: Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain: IEEE, 2018. 1034−1037

21 Carrio A, Tordesillas J, Vemprala S, Saripalli S, Campoy P, How J P. Onboard detection and localization of drones using depth maps. *IEEE Access*, 2020, **8**: 30480−30490

22 Ganti S R, Kim Y. Implementation of detection and tracking mechanism for small UAS. In: Proceedings of the 2016 International Conference on Unmanned Aircraft Systems (ICUAS). Arlington, USA: IEEE, 2016. 1254−1260

23 Farhadi M, Amandi R. Drone detection using combined motion and shape features. In: IEEE International Workshop on Small-Drone Surveillance Detection and Counteraction Techniques. Lecce, Italy: IEEE, 2017. 1−6

24 Alom M Z, Hasan M, Yakopcic C, Taha T M, Asari V K. Improved inception-residual convolutional neural network for object recognition. *Neural Computing and Applications*, 2020, **32**(1): 279−293

25 Saqib M, Khan S D, Sharma N, Blumenstein M. A study on detecting drones using deep convolutional neural networks. In: Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce, Italy: IEEE, 2017. 1−5

26 Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single shot MultiBox detector. In: Proceedings of the 14th European Conference on Computer Vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016. 21−37.

27 Zhou Wei-Xiang, Sun De-Bao, Peng Jia-Xiong. The study of preprocessing algorithm of small moving target detection in infrared image sequences. *Journal of National University of Defense Technology*, 1999, **21**(5): 60−63
(周卫祥, 孙德宝, 彭嘉雄. 红外图像序列运动小目标检测的预处理算法研究. 国防科技大学学报, 1999, **21**(5): 60−63)

28 Wu Y W, Sui Y, Wang G H. Vision-based real-time aerial ob-

ject localization and tracking for UAV sensing system. *IEEE Access*, 2017, **5**: 23969−23978

29 Lv P Y, Lin C Q, Sun S L. Dim small moving target detection and tracking method based on spatial-temporal joint processing model. *Infrared Physics & Technology*, 2019, **102**: Article No. 102973

30 Van Droogenbroeck M, Paquot O. Background subtraction: Experiments and improvements for ViBe. In: Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, USA: IEEE, 2012. 32−37

31 Zamalieva D, Yilmaz A. Background subtraction for the moving camera: A geometric approach. *Computer Vision and Image Understanding*, 2014, **127**: 73−85

32 Sun Y F, Liu G, Xie L. MaxFlow: A convolutional neural network based optical flow algorithm for large displacement estimation. In: Proceedings of the 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES). Wuxi, China: IEEE, 2018. 119−122

33 Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V, et al. FlowNet: Learning optical flow with convolutional networks. In: Proceedings of the 2015 International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 2758−2766

34 Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 1647−1655

35 Chen Xin, Wei Hai-Jun, Wu Min, Cao Wei-Hua. Tracking learning based on gaussian regression for multi-agent systems in continuous space. *Acta Automatica Sinica*, 2013, **39**(12): 2021−2031
(陈鑫, 魏海军, 吴敏, 曹卫华. 基于高斯回归的连续空间多智能体跟踪学习. 自动化学报, 2013, **39**(12): 2021−2031)

36 Shi S N, Shui P L. Detection of low-velocity and floating small targets in sea clutter via income-reference particle filters. *Signal Processing*, 2018, **148**: 78−90

37 Kang K, Li H S, Yan J J, Zeng X Y, Yang B, Xiao T, et al. T-CNN: Tubelets with convolutional neural networks for object detection from videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, **28**(10): 2896−2907

38 Zhu X Z, Xiong Y W, Dai J F, Yuan L, Wei Y C. Deep feature flow for video recognition. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 4141−4150

39 Zhu X Z, Wang Y J, Dai J F, Yuan L, Wei Y C. Flow-guided feature aggregation for video object detection. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 408−417

40 Bertasius G, Torresani L, Shi J B. Object detection in video with spatiotemporal sampling networks. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 342−357

41 Luo H, Huang L C, Shen H, Li Y, Huang C, Wang X G. Object detection in video with spatial-temporal context aggregation. arXiv: 1907.04988, 2019.

42 Xiao F Y, Lee Y J. Video object detection with an aligned spatial-temporal memory. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 494−510.

43 Chen X Y, Yu J Z, Wu Z X. Temporally identity-aware SSD with attentional LSTM. *IEEE Transactions on Cybernetics*, 2020, **50**(6): 2674−2686

44 Shi X J, Chen Z R, Wang H, Yeung D Y, Wong W K, Woo W C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015. 802−810.

45 Gao Xue-Qin, Liu Gang, Xiao Gang, Bavirisetti D P, Shi Kai-Lei. Fusion Algorithm of Infrared and Visible Images Based on FPDE. *Acta Automatica Sinica*, 2020, **46**(4): 796−804
(高雪琴, 刘刚, 肖刚, Bavirisetti Durga Prasad, 史凯磊. 基于FPDE的红外与可见光图像融合算法. 自动化学报, 2020, **46**(4): 796−804)

46 Bluche T, Messina R. Gated convolutional recurrent neural networks for multilingual handwriting recognition. In: Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, Japan: IEEE, 2017. 646−651

47 Deng J, Dong W, Socher R, Li L J, Li K, Li F F. ImageNet: A large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 248−255

48 Son J, Jung I, Park K, Han B. Tracking-by-segmentation with online gradient boosting decision tree. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3056−3064

49 Wang F, Jiang M Q, Qian C, Yang S, Li C, Zhang H G, et al. Residual attention network for image classification. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 6450−6458

50 Zhang Xiu-Wei, Zhang Yan-Ning, Guo Zhe, Zhao Jing, Tong Xiao-Min. Advances and perspective on motion detection fusion in visual and thermal framework. *Journal of Infrared and Millimeter Waves*, 2011, **30**(4): 354−360
(张秀伟, 张艳宁, 郭哲, 赵静, 仝小敏. 可见光-热红外视频运动目标融合检测的研究进展及展望. 红外与毫米波学报, 2011, **30**(4): 354−360)

51 Fu H, Gong M M, Wang C H, Batmanghelich K, Tao D C. Deep ordinal regression network for monocular depth estimation. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2002−2011

52 Dragon R, van Gool L. Ground plane estimation using a hidden Markov model. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 4026−4033

53 Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. In: Proceedings of the 2011 International Conference on Computer Vision. Barcelona. Spain: IEEE, 2011. 2564−2571

54 Lepetit V, Moreno-Noguer F, Fua P. EPnP: An accurate O(n) solution to the PnP problem. *International Journal of Computer Vision*, 2009, **81**(2): 155−166

55 Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 3354−3361

56 Bagautdinov T, Fleuret F, Fua P. Probability occupancy maps for occluded depth images. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 2829−2837

57 Kranstauber B, Cameron A, Weinzerl R, Fountain T, Tilak S, Wikelski M, et al. The Movebank data model for animal tracking. *Environmental Modelling & Software*, 2011, **26**(6): 834−835

58 Belongie S, Perona P, Van Horn G, Branson S. NABirds dataset: Download it now! [Online], available: https://dl.allaboutbirds.org/nabirds, March 30, 2020

59 Xiang Y, Mottaghi R, Savarese S. Beyond PASCAL: A benchmark for 3D object detection in the wild. In: Proceedings of the

2014 IEEE Winter Conference on Applications of Computer Vision. Steamboat Springs, USA: IEEE, 2014. 75−82

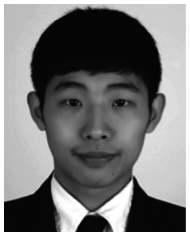
60 Silberman N, Hoiem D, Kohli P, Fergus R. Indoor segmentation and support inference from RGBD images. In: Proceedings of the 12th European conference on Computer Vision. Florence, Italy: Springer, 2012. 746−760

61 Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence. Vancouver, Canada: Morgan Kaufmann, 1981. 674−679

62 Yazdian-Dehkordi M, Rojhani O R, Azimifar Z. Visual target tracking in occlusion condition: A GM-PHD-based approach. In: Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012). Shiraz, Iran: IEEE, 2012. 538−541

63 Yin Z C, Shi J P. GeoNet: Unsupervised learning of dense depth, optical flow and camera pose. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1983−1992

64 Mur-Artal R, Tardós J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 2017, **33**(5): 1255−1262



刘孙相与 清华大学航天航空学院博士研究生. 主要研究方向为目标识别, 目标分割和深度学习.
E-mail: lsxy_qd@126.com
(**LIU Sun-Xiang-Yu** Ph. D. candidate at School of Aerospace Engineering, Tsinghua University. His research interest covers object detection, object segmentation, and deep learning.)

李贵涛 清华大学航天航空学院副教授. 主要研究方向为计算机仿真和图像处理.
E-mail: ligt@tsinghua.edu.cn
(**LI Gui-Tao** Associate professor at School of Aerospace Engineering, Tsinghua University. His research interest covers computer simulation and image processing.)

詹亚锋 清华大学信息国家研究中心教授. 主要研究方向为 TT&C 系统, 信号处理和深空通信.
E-mail: zhanyf@tsinghua.edu.cn
(**ZHAN Ya-Feng** Professor at Beijing National Research Center for Information Science and Technology, Tsinghua University. His research interest covers TT&C systems, communication signal processing, and deep space communications.)

高 鹏 北京大学工学院博士后. 主要研究方向为计算机体系结构, 机器学习和图像处理. 本文通信作者.
E-mail: gaopeng1982@pku.edu.cn
(**GAO Peng** Postdoctoral researcher at College of Engineering, Peking University. His research interest covers computer architecture, machine learning, and image processing. Corresponding author of this paper.)