基于深度学习的低空小型无人机障碍物视觉感知

寇展^{1,2}, 吴健发^{1,2,3}, 王宏伦^{1,2*}, 阮文阳^{1,2}

1. 北京航空航天大学自动化科学与电气工程学院, 北京 100191

2. 北京航空航天大学飞行器控制一体化技术重点实验室, 北京 100191

3. 北京航空航天大学高等理工学院, 北京 100191

* 通信作者. E-mail: hl_wang_2002@126.com

收稿日期: 2019-02-15; 修回日期: 2019-04-17; 接受日期: 2019-05-05; 网络出版日期: 2020-04-16

国家自然科学基金 (批准号: 61175084, 61673042) 资助项目

摘要 本文提出了一种基于深度学习与目标跟踪方法, 综合单目视觉和双目立体视觉特点的无人机障碍物实时感知方法. 首先, 采用深度学习的方法对相机采集的首帧图片进行障碍物检测与识别, 并采用目标跟踪算法对首帧检测结果进行实时跟踪, 以提高检测系统实时性; 同时, 使用双目立体视觉技术对当前帧图像进行全图的三维重构, 得到环境的空间信息; 之后, 通过在检测结果的感兴趣区域内进行点云聚类提取等策略并结合上述感知到的信息进行信息融合, 得到障碍物的种类、空间位置及轮廓大小. 最后, 开发实物样机对方法进行验证, 结果表明通过采用该方法, 无人机在搭载一个双目相机的情况下即可完成对障碍物的实时感知.

关键词 深度学习, 目标跟踪, 无人机, 障碍感知, 双目视觉

1 引言

近年来, 随着航空技术的飞速发展, 无人机的作业空间逐渐从中高空向低空环境拓展, 而复杂非结构化的低空环境也对无人机的自主感知与避障技术带来了新的挑战. 目前, 无人机在低空环境中的障碍感知主要依靠以毫米波雷达、激光雷达、超声波等为代表的主动式传感器和以光电、红外等视觉探测为代表的被动式传感器. 由于视觉传感器具有重量轻、体积小和获得信息丰富等优点, 因此日益受到青睐, 例如, 意大利宇航研究中心 (CIRA-Italian Aerospace Research Center, CIRA)^[1] 开发出一套全自动多传感器防撞系统, 该系统的感知模块由脉冲 Ka 波段雷达、两个视觉相机、两个红外摄像机, 以及一个专用的图像处理计算机组成, 通过数据融合实现对障碍目标的识别与跟踪; Fränken 等^[2] 提出一种对数极坐标系扩展/无迹卡尔曼 (Kalman) 滤波算法 (LPC-EKF/UKF), 用于融合由雷达、视觉传感器、光电传感器所探测到的入侵飞机的运动学参数.

引用格式: 寇展, 吴健发, 王宏伦, 等. 基于深度学习的低空小型无人机障碍物视觉感知. 中国科学: 信息科学, 2020, 50: 692–703, doi: 10.1360/N112019-00034
Kou Z, Wu J F, Wang H L, et al. Obstacle visual sensing based on deep learning for low-altitude small unmanned aerial vehicles (in Chinese). Sci Sin Inform, 2020, 50: 692–703, doi: 10.1360/N112019-00034

上述研究的共同特点是将视觉传感器与其他传感器进行数据融合,从而实现对环境和威胁的高精度感知。然而,对于低空环境中最广泛应用的小型无人机来说,上述多传感器融合方案在性能、成本、功耗与有效载荷之间难以取得平衡,因此有必要考虑设计一种仅依靠视觉传感器的感知方法。对此,国内外学者也进行了一定的探索,例如 Mueller 等^[3]在四轴飞行器上安装双目相机,通过双目 SLAM 完成定位与建图,成功实现了四旋翼无人机在未知环境下的自主避障。但这种采用双目 SLAM 进行避障的方法没有进行稠密的构建工作,而且受内存限制只能在小范围内建图避障。Ross 等^[4]采用单目视觉帮助无人机穿过森林,使用模仿学习算法 (imitation learning) 进行训练学习使得无人机可以避开树木,但该方法还不够稳定且对无人机飞行速度限制较大。Shuai 等^[5]利用双目视觉技术标定电力线和塔架的位置/形态信息,并将该信息作为基于扰动流体动态系统的航路规划算法^[6]的决策依据,一定程度上实现了无人机低空电力巡检的安全避障。这些研究存在一些共同的缺点,比如只适用于特定环境,难以获得更加丰富全面的障碍物信息等。

随着人工智能的快速发展,深度学习方法^[7~10]在无人机避障的感知过程中有着巨大的应用前景,其强大的逼近能力使其能够提取到更多的障碍特征,并具有更强的环境适应性,从而使得无人机在搭载一个双目相机的情况下能够完成对障碍物的实时感知。基于以上分析,本文将深度学习方法与双目视觉技术相结合,提出一种障碍物视觉感知方法,以实现无人机对障碍物的种类、大小、相对位置的实时综合感知。

2 感知方法总体结构

本文所提出的感知方法总体思路如下,其总体框图见图 1。

(1) 使用相机采集的障碍物图像对深度神经网络模型进行离线训练,使得训练后的模型能够在在线使用的相机所采的图像中识别出障碍物的种类并检测出障碍物在图片中的像素位置与像素大小,之后转入如下的在线感知过程;

(2) 使用双目相机的左相机通过深度学习模型在相机图像中检测识别出障碍物并进行目标跟踪,实时获得障碍物的种类和在每帧图片中的像素位置与大小;

(3) 通过对双目相机进行标定、矫正、双目匹配与 3D 距离恢复的方法对相机左右图进行三维重建,得到无人机飞行环境中障碍物相对于左相机的三维位置信息;

(4) 将深度学习模型检测出的障碍物信息与双目相机获得的三维位置信息进行信息融合,以获得环境中障碍物的种类、外形、轮廓尺寸,以及相对相机的空间位置信息,从而实现对低空小型无人机飞行环境的实时感知。

3 基于单目视觉的障碍物实时检测和跟踪

3.1 基于改进 YOLOv2 算法的障碍物检测与识别

YOLOv2 检测算法^[9,11]是一种强实时性的深度学习目标检测算法。该算法将候选框提取和目标检测统一到了同一个网络中,无需单独提取候选框,并采用整幅图像来进行目标检测与识别,实现了真正的端对端检测,具有准确率高、实时性好和泛化能力强的特点。

图 1 基于视觉的障碍物实时感知方法总体框图

Figure 1 Block diagram of the vision-based real-time obstacle perception method

3.1.1 改进 YOLOv2 算法介绍

在低空小型无人机环境感知任务中, 单目视觉障碍物检测与识别更重要的是为后续的三维重建提供一个感兴趣区域并标识出障碍物类别, 因此为了在该任务中取得更好的效果, 本文通过为 YOLOv2 算法的损失函数引入权重修正系数 γ_c , 使得改进后的算法更关注分类误差, 而不是定位精度. 改进的 YOLOv2 算法训练损失函数如下:

$$\begin{aligned}
 & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
 & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
 & + \gamma_c \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \gamma_c \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\
 & + \sum_{i=0}^{S^2} I_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2, \quad (1)
 \end{aligned}$$

其中第 1 项和第 2 项表示的是定位惩罚项, 即对预测出的边界框位置和长宽与实际标定值的误差的惩罚项, 式中 I_{ij}^{obj} 在第 i 个单元格存在目标, 且该单元格中的第 j 个边界框负责预测该目标时才为 1, 因此只有在目标的中心落在了第 i 个单元格的第 j 个边框中时才计算定位误差; x_i, y_i, w_i, h_i 与 $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$ 分别表示矩形框的中心位置像素坐标和矩形框像素长宽的实际值与预测值; S^2 表示图像划分的区域数; B 表示每个区域预测的包围框个数. 第 3 项与第 4 项表示的是边界框置信度惩罚项, 其中 I_{ij}^{noobj} 表示第 i 个单元格的第 j 个边框内没有目标的中心点, 因此在目标的中心落在了第 i 个单元格的第 j 个边框中时计算第 3 项, 否则计算第 4 项, 区别在于系数的不同, 一般 λ_{noobj} 小于 λ_{coord} , 因为图片分成 $S \times S$ 网格, 其中大部分网格是不包含目标的, 因此为了权衡正负样本对损失函数值的影响, 设置权重系数 λ_{noobj} 小于 λ_{coord} ; C_i 与 $\hat{C}_i$ 分别表示第 i 个单元格包含目标置信度的实际值与预测值; $\gamma_c > 1$ 为本文引入的权重修正系数, 用于增大边界框置信度惩罚项, 使得训练出来的模型在适当降低检测框定位精度的条件下, 提高分类正确率. 第 5 项为类别预测惩罚项, 即预测出的目标类别概率与实际概率误差的惩罚项, 其中, $p_i(c)$ 和 $\hat{p}_i(c)$ 分别表示目标类别的实际概率值和预测概率值, 如果所训练的图

图 2 (网络版彩图) 数据集部分图片展示
Figure 2 (Color online) Samples in the data set

片中确定存在该类目标, 则 $p_i(c) = 1$, 反之, $p_i(c) = 0$, 而 $\hat{p}_i(c) \in [0, 1]$.

3.1.2 数据集生成

模型的训练过程中, 训练数据集的好坏直接决定了 YOLOv2 算法应用到具体场景中的表现, 因此训练数据集的制作十分重要. 数据集须包含多种典型的无人机飞行作业情况, 包括不同的天气环境、光照环境和障碍物的不同情况等. 对于深度学习中的目标检测问题来说, 数据集一般由被检测目标的实际图片及其标注所组成. 然而, 对于无人机低空障碍感知这一具体问题, 受限于采集成本, 机载相机所能够采集到的障碍物实景图片数量相对有限, 且拍摄视角和拍摄距离不够丰富, 无法较为全面地反映障碍物的特征信息. 因此, 如果仅使用有限的障碍物实景图片作为数据集, 所训练得出的模型可能难以满足较高精度的检测要求. 为了扩充障碍物的数据集, 本文采用 Unity3D 软件搭建逼真的飞行模拟场景作为障碍物实景图片的补充. 采用本文的数据集制作方法, 不仅简单便捷且成本低, 并且训练出的模型泛化效果较好, 数据集部分图片展示如图 2 所示.

3.1.3 检测与识别结果展示

将训练完成的模型用于障碍物目标检测与识别, 选取了不同环境不同障碍物的 300 张图片 (包含实际场景和仿真场景) 进行模型的验证, 最终的部分效果展示如图 3 所示.

结果表明, 训练出的模型仅对其中的 1 张图片检测出现了漏检情况, 分类正确率高达 99% 以上, 能够胜任无人机的避障任务. 在 PC 端能够达到每秒 67 帧, 表明算法实时性较好. 在测试集中同时包含了实际场景图片和视景仿真图片的情况下, 仍能取得 99% 以上的分类正确率, 表明训练出来的模型具有较好的泛化性.

图 3 (网络版彩图) YOLOv2 目标检测与识别算法效果展示
Figure 3 (Color online) Target detection and recognition results using YOLOv2

3.2 障碍物目标实时跟踪

尽管基于 YOLOv2 的深度学习检测算法速度不慢, 但由于无人机避障系统中包括基于深度学习的目标检测、双目视觉三维重构和动态航路规划等多个耗时的任务, 为了提高系统实时性, 采用下述的长周期深度学习检测与短周期 KCF 目标跟踪算法^[12] 结合的方法对障碍物进行实时跟踪; YOLOv2 算法在一个慢的线程中运行 (每隔数秒运行一次), 目标跟踪算法在一个快的线程中运行 (每采样周期运行一次), 每次 YOLOv2 算法的检测结果对目标跟踪算法进行修正, 这样既可保证在每个时刻完成障碍物的检测与识别, 提高避障系统实时性, 又能够提高整体的目标检测与识别精度。

4 基于双目立体视觉的三维重建和障碍物信息提取

4.1 基于双目立体视觉的三维重建

通过单目视觉的障碍物实时检测方法实时得到图像中的感兴趣区域, 还需要得到图像中障碍物的空间信息, 因此本文采用双目立体视觉技术^[13,14] 进行环境的三维重构. 双目立体视觉三维重建的一般步骤如图 4 所示, 包括相机标定、图像矫正、双目匹配和三维重构 4 个部分。

其中, 相机标定采用离线标定, 得到相机的内外参数, 然后根据相机内外参数进行图像矫正, 接着采用全局稠密匹配算法进行双目匹配得到稠密视差图, 最后根据相机内外参数与视差图, 采用最小二

图4 (网络版彩图) 双目立体视觉三维重建流程图

Figure 4 (Color online) 3D reconstruction based on binocular stereo vision

乘法^[14], 即可得到全图的三维空间信息, 包括空间点在左相机坐标系下的 XYZ 坐标, 其计算公式为

$$[X, Y, Z]^T = (\mathbf{K}^T \mathbf{K}^{-1})^T \mathbf{U},$$

$$\mathbf{K} = \begin{bmatrix} u_1 m_{31}^1 - m_{11}^1 & u_1 m_{32}^1 - m_{12}^1 & u_1 m_{33}^1 - m_{13}^1 \\ v_1 m_{31}^1 - m_{21}^1 & v_1 m_{32}^1 - m_{22}^1 & v_1 m_{33}^1 - m_{23}^1 \\ u_2 m_{31}^2 - m_{11}^2 & u_2 m_{32}^2 - m_{12}^2 & u_2 m_{33}^2 - m_{13}^2 \\ v_2 m_{31}^2 - m_{21}^2 & v_2 m_{32}^2 - m_{22}^2 & v_2 m_{33}^2 - m_{23}^2 \end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix} m_{14}^1 - u_1 m_{34}^1 \\ m_{24}^1 - v_1 m_{34}^1 \\ m_{14}^2 - u_2 m_{34}^2 \\ m_{24}^2 - v_2 m_{34}^2 \end{bmatrix},$$
(2)

其中 $[X, Y, Z]^T$ 为空间点在左相机坐标系下的 XYZ 坐标; $[u_1, v_1, 1]^T$ 和 $[u_2, v_2, 1]^T$ 分别为空间点在左右相机图像中的齐次像素坐标; m_{ij}^1 和 m_{ij}^2 ($i = 1, 2, 3; j = 1, 2, 3, 4$) 分别为左右相机投影矩阵 $\mathbf{M}_1$ 和 $\mathbf{M}_2$ 的第 i 行第 j 列元素.

4.2 障碍物信息提取

在 3.1 与 4.1 小节中, 已经通过基于单目视觉的障碍物实时检测与识别得到了障碍物在图像中的像素位置及障碍物种类, 并且通过基于双目立体视觉的三维重构技术获得了障碍物的空间三维坐标信息, 因此, 本节将前面获得的障碍物信息进行信息融合, 从而得到障碍物的空间三维位置、物理轮廓, 以及障碍物种类信息, 完成对障碍物的实时感知。

通过信息融合手段^[15]提取障碍物信息的整个过程步骤如下:

(1) 通过 3.1 小节基于单目视觉的障碍物实时检测与识别技术, 能够实时检测出无人机前方飞行环境中每个障碍物在图像中的矩形框位置及对应的障碍物种类。

(2) 通过 4.1 小节基于双目立体视觉的三维重建技术, 能够实时得到无人机前方飞行环境的全图三维稠密点云信息。

(3) 通过第 (1) 步中得到的矩形框, 可以在第 (2) 步得到的全图三维稠密点云信息中提取感兴趣区域内的三维点云信息。

(4) 对第 (3) 步得到的局部点云信息进行无效点剔除, 在本文中, 滤除掉由于立体匹配出错造成的距离为负值的点以及位于小型无人机避障预警距离之外 (该距离由无人机的机动能力和当前飞行速度所决定, 当目标处于该距离以外时, 可以认为其不会对无人机飞行安全造成威胁) 的点。

(5) 对前一步得到的剔除无效点云之后的局部点云图进行无监督聚类。本文采用 K -means 聚类算法, 其中聚类类别数 K 可以通过计算当前矩形框与其他障碍物矩形框的重叠情况来决定, 例如有两个其他矩形框与当前矩形框发生重叠, 那么 K 值将取为 3。

(6) 通过 K -means 聚类算法对当前矩形框内点云进行聚类后, 选取属于当前矩形框的那一类点云簇, 选取标准为点云簇在矩形框内分布范围广且点云数量多。然后从该簇三维点云集中分别找出相机坐标系下 X 的最小和最大值, 分别为 $X_{\min}$ 和 $X_{\max}$, 以及 Y 的最小和最大值, 分别为 $Y_{\min}$ 和 $Y_{\max}$, 然后按下式可得障碍物的物理宽高:

$$\begin{cases} W = X_{\max} - X_{\min}, \\ H = Y_{\max} - Y_{\min}, \end{cases} \quad (3)$$

其中 W 和 H 表示障碍物轮廓的实际宽和高。

(7) 对每一个矩形框重复 (2)~(6) 步, 从而获得每个障碍物的种类、表面中心位置在左相机坐标系的空间三维坐标, 以及实际物理轮廓, 完成障碍物的实时信息感知。

进行信息融合提取障碍物信息整个流程如图 5 所示。

5 实验验证与分析

本文开发了感知系统实物样机, 在室内复杂环境下对本文所提方法的性能进行了验证。实物样机使用 NVIDIA Jetson TX2 开发板, 处理器由双核 Denver CPU 和四核 A57 CPU 组成, GPU 采用 Pascal 架构, 搭载 8 GB 内存。在该样机上运行前文所述的经过训练的算法软件。采用 ZED 双目相机, 其基线为 12 cm, 视场角均为 110° , 可感知的最大距离为 20 m。感知效果如图 6 所示。

其中, 图 6(a) 为感知系统实物样机中基于单目视觉的实时障碍物检测与识别算法对室内环境的检测结果图, 可见, 在复杂的环境下, 算法能够准确识别出障碍无人机, 而未受复杂环境中其他物体的干扰, 同时还能够获取障碍无人机的方位与高低信息以及轮廓尺寸大小; 图 6(b) 为感知系统实物样机

图 5 信息融合进行障碍物信息提取流程图

Figure 5 Obstacle extraction using information fusion method

中双目立体视觉算法对室内环境的重构效果图, 通过该图, 能够获得环境的深度信息. 二者融合, 则可获得障碍物的种类、三维位置和物理轮廓尺寸.

利用所搭建的感知系统测量室内距离双目相机 5 m 范围内的障碍无人机位置及轮廓, 测量结果如表 1 所示.

从表 1 可以看出, 在室内 5 m 环境内进行感知, 得到的障碍物位置及轮廓误差均可控制在 5 cm 以内, 感知精确度较高; 同时感知算法的检测速率为 42 FPS, 实时性较强.

为了进一步验证算法定位的稳定性, 对表 1 中的 5 组实际位置进行多次测量 (每隔 30 帧, 每组测量 10 次), 其误差均值和标准差见表 2, 可以看出目标的方位和高低 (X , Y 向) 误差均值和标准差分别能够控制在 1.4~3.2 cm 和 0.4~1.0 cm 范围内, 而深度 (Z 向) 误差均值和标准差分别仅占实际深度的 0.73%~1.35% 和 0.04%~0.225%, 体现出较高的定位精度和稳定性.

为了进一步验证算法的跟踪性能, 令障碍无人机在 ZED 相机的视角内处于连续移动状态, 利用

图 6 实物样机感知效果

Figure 6 Perception of the physical prototype. (a) Real-time detection result using YOLOv2 and KCF tracking algorithm; (b) point cloud for 3D environment reconstruction using binocular stereo vision

表 1 室内环境下实物样机感知结果

Table 1 Perception results of the physical prototype in indoor environments

		Center coordinate X (m)	Center coordinate Y (m)	Center coordinate Z (m)	Width (m)	Height (m)
Group 1	Real value	0.000	0.000	1.500	0.370	0.370
	Measured value	-0.012	0.021	1.493	0.374	0.374
	Error	-0.012	0.021	-0.007	0.004	0.004
Group 2	Real value	0.000	0.400	1.980	0.370	0.370
	Measured value	0.009	0.443	1.981	0.374	0.374
	Error	0.009	0.043	0.001	0.004	0.004
Group 3	Real value	-0.400	0.400	2.500	0.370	0.370
	Measured value	-0.378	0.433	2.550	0.374	0.382
	Error	0.022	0.033	0.050	0.004	0.012
Group 4	Real value	-0.400	0.400	3.000	0.370	0.370
	Measured value	-0.447	0.447	2.987	0.402	0.382
	Error	-0.047	0.047	0.013	0.032	0.012
Group 5	Real value	0.000	0.400	4.000	0.370	0.370
	Measured value	0.037	0.442	3.951	0.397	0.420
	Error	0.037	0.042	0.049	0.027	0.050

实物样机实时采集其图像并实现目标的检测与跟踪, 结果如图 7 所示.

从图 7 中可以看出, 算法在第 1 帧检测到了目标, 并将其框选出. 算法运行过程中边界框红黑交替闪烁, 随着目标的移动, 其边界框也随之移动, 并始终保持准确跟踪状态, 且 KCF 算法跟踪速率可达 107 FPS, 同样具有较好的实时性. 因此, 所提出的算法具有良好的跟踪性能.

6 结论

本文提出了一种深度学习与目标跟踪相结合的飞行环境障碍物实时感知方法. 该方法综合了单目

表 2 位置误差均值和标准差统计结果

Table 2 Statistical results of position error mean values and standard deviations^{a)}

	X-EM (cm)	X-SD (cm)	Y-EM (cm)	Y-SD (cm)	Z-EM (cm)	Z-SD (cm)
Group 1	1.4	0.5	1.5	0.4	1.1	0.3
Group 2	1.4	0.4	1.3	0.4	1.1	0.4
Group 3	1.8	0.6	2.2	0.4	2.2	0.1
Group 4	2.3	0.6	3.2	1.0	3.8	0.5
Group 5	2.6	0.6	2.8	0.6	5.4	0.9

a) $X/Y/Z$ -EM represents $X/Y/Z$ -axis position error mean value and $X/Y/Z$ -SD represents $X/Y/Z$ -axis standard deviation.

图 7 (网络版彩图) 算法对实物目标的检测与跟踪结果

Figure 7 (Color online) Detection and tracking results for the physical object using the proposed algorithm. (a) Image at the 1st frame; (b) image at the 60th frame; (c) image at the 120th frame

视觉和双目立体视觉的特点, 在单目视觉上利用深度学习实现对障碍物的实时检测与跟踪, 同时利用双目立体视觉技术进行全图的三维重构. 最后, 通过在检测结果的感兴趣区域内进行点云聚类提取等策略并结合上述感知到的信息进行信息融合, 可以得到障碍物的种类、空间位置及轮廓大小. 通过搭建实物样机系统对本文所提出的感知算法进行了验证, 结果表明本文所采用的感知方法具有较好的实时性和较高的精度. 在未来工作中, 一方面, 鉴于深度学习的快速发展, 未来可采用更适合嵌入式平台的深度学习网络, 例如 MobileNet^[16] 和 ShuffleNet^[17], 并将该视觉感知方法应用到成本更低的嵌入式平台上; 另一方面, 将对所提出的方法进行实际飞行验证.

参考文献

- 1 Fasano G, Accardo D, Moccia A, et al. Multi-sensor-based fully autonomous non-cooperative collision avoidance system for unmanned air vehicles. *J Aeronaut Comput Inf Commun*, 2008, 5: 338–360
- 2 Fränken D, Hupper A. Unified tracking and fusion for airborne collision avoidance using log-polar coordinates. In: *Proceedings of the 15th IEEE International Conference on Information Fusion*, Singapore, 2012. 1246–1253
- 3 Mueller M W, D'Andrea R. Critical subsystem failure mitigation in an indoor UAV testbed. In: *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura, 2012. 780–785
- 4 Ross S, Melik-Barkhudarov N, Shankar K S, et al. Learning monocular reactive UAV control in cluttered natural environments. In: *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, Karlsruhe, 2013. 1765–1772
- 5 Shuai C, Wang H, Zhang W, et al. Binocular vision perception and obstacle avoidance of visual simulation system for power lines inspection with UAV. In: *Proceedings of the 36th Chinese Control Conference*, Dalian, 2017. 10480–10485
- 6 Wang H L, Wu J F, Yao P. UAV three-dimensional path planning based on interfered fluid dynamical system: methodology and application. *Unmanned Syst Technol*, 2018, 1: 72–82 [王宏伦, 吴健发, 姚鹏. 基于扰动流体动态系统的无人机三维航路规划: 方法与应用. *无人系统技术*, 2018, 1: 72–82]

- 7 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 580–587
- 8 Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 1137–1149
- 9 Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 2016. 779–788
- 10 Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: Proceedings of the 2016 European Conference on Computer Vision. Berlin: Springer, 2016. 21–37
- 11 Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 7263–7271
- 12 Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. IEEE Trans Pattern Anal Mach Intell, 2015, 37: 583–596
- 13 Tao C B, Qiao L, Sun Y F, et al. Stereo matching algorithm of six rotor UAV based on binocular vision. Laser Infrared, 2018, 48: 1181–1187 [陶重犇, 乔荔, 孙云飞, 等. 基于双目视觉的六旋翼无人机立体匹配算法. 激光与红外, 2018, 48: 1181–1187]
- 14 Gao H W. Computer Binocular Stereo Vision. Beijing: Electronic Industry Press, 2012. 132–135 [高宏伟. 计算机双目立体视觉. 北京: 电子工业出版社, 2012. 132–135]
- 15 Wang Z R, Guo X K, Zhao G. Depth image information fusion and three-dimensional reconstruction method of double binocular stereo vision. Laser Infrared, 2019, 49: 246–250 [王中任, 郭晓康, 赵刚. 两组双目立体深度图像信息融合与三维重建方法. 激光与红外, 2019, 49: 246–250]
- 16 Jacob B, Kligys S, Chen B, et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 2704–2713
- 17 Zhang X, Zhou X, Lin M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 6848–6856

Obstacle visual sensing based on deep learning for low-altitude small unmanned aerial vehicles

Zhan KOU^{1,2}, Jianfa WU^{1,2,3}, Honglun WANG^{1,2*} & Wenyang RUAN^{1,2}

1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;

2. Science and Technology on Aircraft Control Laboratory, Beihang University, Beijing 100191, China;

3. Shenyuan Honors College of Beihang University, Beijing 100191, China

* Corresponding author. E-mail: hl.wang_2002@126.com

Abstract An obstacle real-time sensing method, which is based on deep learning and target tracking method and integrated with monocular vision and binocular vision, is proposed for unmanned aerial vehicles (UAVs) in this paper. Firstly, it uses the deep learning method to detect and recognize the first-frame figure collected by cameras. Then, it uses the target tracking algorithm to track the detection results for the first-frame figure in real time to improve the real-time performance of the detection system. Meanwhile, it uses the binocular vision technology to execute the three-dimensional reconstruction for the current frame of the entire figure to obtain the environmental spatial information. Subsequently, combined with the points clustering strategy and the information fusion method, it can resolve the types, spatial locations, and outlines of obstacles. Finally, to verify the proposed method, we developed a physical prototype, and the results showed that the real-time sensing for obstacles can be realized under the condition that UAVs are equipped with one binocular camera.

Keywords deep learning, target tracking, UAVs, obstacle sensing, binocular vision

Zhan KOU was born in 1995. He received his master's degree from Beihang University, Beijing, in 2019. Currently, he is an assistant engineer at No.717 Research Institute of China Shipbuilding Industry Corporation. His research interest includes UAV sense-and-avoid technology.

Jianfa WU was born in 1993. He received his bachelor's degree from Harbin Engineering University, Harbin, in 2016. Currently, he is a Ph.D. candidate at Beihang University. His research interests include UAV sense-and-avoid technology and multi-agent cooperative path planning.

Honglun WANG was born in 1970. He received his Ph.D. degree from Northwestern Polytechnical University, Xi'an, in 1998. Currently, he is a professor at Beihang University. His research interests include autonomous flight control for UAVs, sense-and-avoid technology for UAVs, and theory and method of path planning in complex environments for UAVs.

Wenyang RUAN was born in 1994. He received his bachelor's degree from Shandong University of Science and Technology, Qingdao, in 2017. Currently, he is a graduate student at Beihang University. His research interest includes UAV sense-and-avoid technology.