

农业车辆双目视觉障碍物感知系统设计与试验

魏建胜¹, 潘树国^{1*}, 田光兆², 高旺¹, 孙迎春¹

(1. 东南大学仪器科学与工程学院, 南京 210096; 2. 南京农业大学工学院, 南京 210031)

摘要: 为保证智能化农业机械在自主导航过程中的安全可靠, 该研究将嵌入式 AI 计算机 Jetson TX2 作为运算核心, 设计一套基于双目视觉的农业机械障碍物感知系统。使用深度卷积神经网络对作业环境中的障碍物进行识别, 并提出一种基于改进 YOLOv3 的深度估计方法。将双目相机抓取的左右图像分别输入至改进的 YOLOv3 模型中进行障碍物检测, 并将输出的目标检测框信息进行目标匹配和视差计算, 完成对障碍物的识别、定位和深度估计。试验结果表明, 该系统能够对障碍物进行准确识别, 平均准确率和召回率分别达到 89.54% 和 90.18%; 改进 YOLOv3 模型深度估计误差均值、误差比均值较原始 YOLOv3 模型分别降低 25.69%、25.65%, 比 Hog+SVM 方法分别降低 41.9%、41.73%; 动态场景下系统对障碍物深度估计的平均误差比为 4.66%, 平均耗时 0.573 s, 系统在深度预警时能够及时开启电控液压转向模块进行安全避障。研究结果可为农业机械的自主导航提供有效的环境感知依据。

关键词: 农业机械; 图像处理; 障碍物感知; 深度估计; 视差计算

doi: 10.11975/j.issn.1002-6819.2021.09.007

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2021)-09-0055-09

魏建胜, 潘树国, 田光兆, 等. 农业车辆双目视觉障碍物感知系统设计与试验[J]. 农业工程学报, 2021, 37(9): 55-63.

doi: 10.11975/j.issn.1002-6819.2021.09.007 <http://www.tcsae.org>

Wei Jiansheng, Pan Shuguo, Tian Guangzhao, et al. Design and experiments of the binocular visual obstacle perception system for agricultural vehicles[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(9): 55-63. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.09.007 <http://www.tcsae.org>

0 引言

由于农业环境的复杂多样性, 农业机械为实现其自主导航^[1-5]需要可靠的障碍物感知系统。障碍物感知系统主要包括障碍物检测^[6-7]和深度估计^[8-9]2 部分。传统的障碍物检测是基于人为设计的浅层目标特征^[10-12], 如 SIFT 特征、HOG 特征、局部二值特征等。由于农业环境结构复杂, 光照强度不均匀等, 此类特征的检测效果不够稳定。随着人工智能的发展, 农业机械的自主导航逐渐应用深度卷积神经网络^[13-16](Convolutional Neural Network, CNN)来完成检测任务。相比于原有基于人为设计特征的检测方式, 深度卷积神经网络对环境特征的检测更加丰富和多层次, 且能够从大规模数据集中不断学习当前任务的特征表达, 从而获得更优的检测效果。在障碍物的深度估计中, 普遍采用激光雷达^[17-18]、深度相机^[19-20]和双目相机^[21-22]等作为测距传感器。激光雷达测距范围广、精度高, 但二维激光雷达无法检测扫描线以外的障碍物, 不能适应农田等颠簸路面的情况, 而三维激光雷达造价昂贵, 严重制约了其在农业障碍物检测中的应用^[23]; 深度相机能够获得图像中每个像素的深度信息, 但测量结果受外界环境干扰较大, 难以在室外环境中应用; 双

目相机能够独立完成对目标的深度估计, 但窄基线的双目相机测距范围有限, 需不断提高基线的宽度以适应大规模的场景。

为适应农业环境的复杂多样性, 选用双目相机作为系统的视觉感知传感器, 并采用深度卷积神经网络进行障碍物的检测。由于深度卷积神经网络对图像处理的计算量巨大, 需要图形处理器 (Graphics Processing Unit, GPU) 来加速处理, 仅中央处理器 (Central Processing Unit, CPU) 配置的工控机已不能满足计算要求, 而工作站等高计算量设备体积和质量过大, 严重占用机械空间资源, 因而本文选用 NVIDIA Jetson TX2^[24-27]型嵌入式人工智能 (Intelligent Artificial, AI) 计算机作为深度卷积神经网络的计算载体。

在传统的障碍物感知中, 通常将障碍物检测和深度估计作为 2 个独立的任务进行处理, 2 个任务之间信息不能共享, 造成了计算资源的浪费。基于此, 本文提出一种基于改进 YOLOv3 的深度估计方法, 将传统障碍物感知任务中的障碍物检测和深度估计 2 个任务进行融合, 利用障碍物检测中的部分信息进行深度估计, 实现系统对障碍物检测和深度估计端到端处理。

1 视觉感知系统设计

1.1 硬件组成

系统选用东方红 SG250 型拖拉机作为移动载体, Jetson TX2 作为运算核心, 搭载本文所设计的障碍物视觉感知系统。系统硬件由拖拉机、导航控制模块和视觉感知模块组成。

收稿日期: 2020-07-24 修订日期: 2021-03-31

基金项目: 国家自然科学基金项目 (41774027, 41904022)

作者简介: 魏建胜, 博士生, 研究方向为机器视觉导航。

Email: To_weijiansheng@163.com

*通信作者: 潘树国, 教授, 博士生导师, 研究方向为高精度卫星导航定位和多源泛在导航定位技术。Email: psg@seu.edu.cn

视觉传感器选用 MYNTAI 公司 S1030-120 型双目相机, 基线 120 mm, 焦距 2.1 mm, 分辨率 752×480 像素。主控制器 Jetson Tx2 的 GPU 配有 256 个 NVIDIA CUDA 核心, 其 CPU 为双核 Denver 2 64 位 CPU 和四核 ARM A57 Complex 的组合。视觉感知模块如图 1 所示。

对拖拉机进行并联电控液压油路改造, 控制器选用 STM32, 与模拟量输出模块、电液比例控制器和比例换向阀等共同构成系统的导航控制模块, 如图 2 所示。

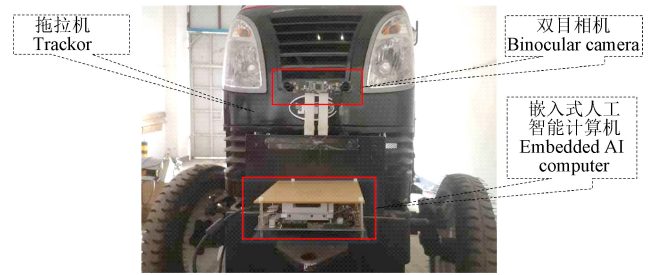


图 1 视觉感知模块
Fig.1 Visual perception module

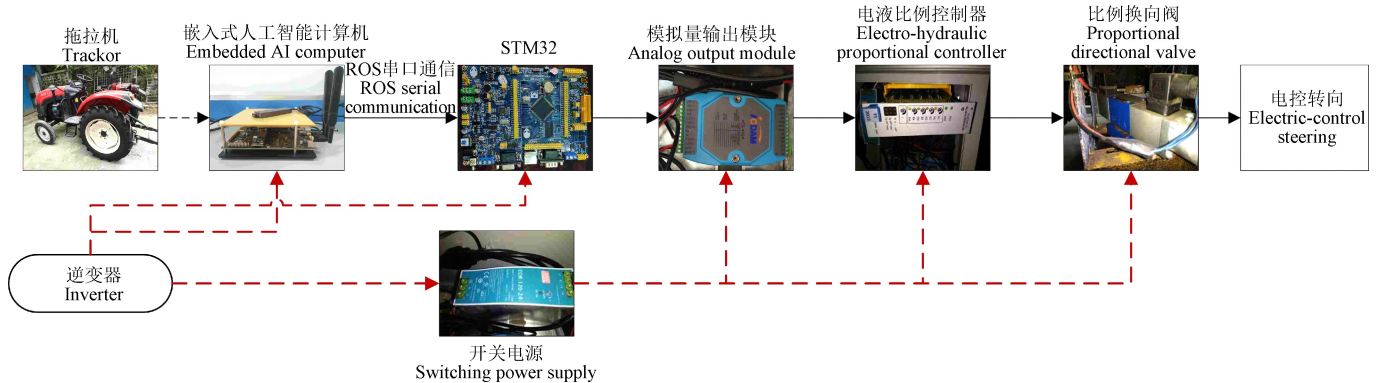


图 2 导航控制模块
Fig.2 Navigation control module

1.2 技术流程

障碍物视觉感知系统的运行环境为 Ubuntu16.04 LTS, 基于 ROS (Robot Operating System) 完成信息的传递和通信, 实现拖拉机对其导航控制路径上障碍物的准确识别、定位和深度估计, 并将结果传给决策中心, 系统的技术流程如图 3 所示。

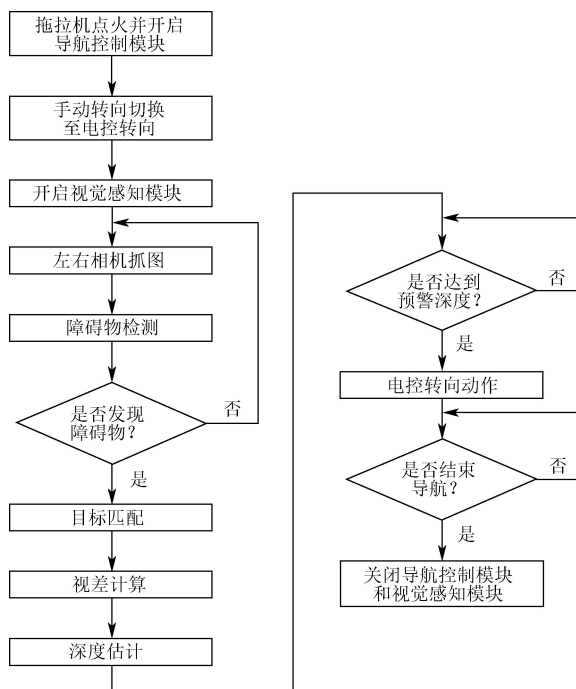


图 3 视觉感知系统技术流程
Fig.3 Technical flow chart of visual perception system

拖拉机点火并切换至电控转向状态, 然后视觉感知模块开启进行左右相机的抓图, 接着将左右图像分别输入至改进的障碍物检测模型进行检测。检测到障碍物类别、定位信息后进行目标匹配和视差计算, 并估计出深度值, 若其超过预警深度值则会触发电控转向。

2 基于改进 YOLOv3 的深度估计

为了准确估计障碍物到拖拉机的实时深度, 本文提出一种基于改进 YOLOv3 的深度估计方法, 将障碍物检测和深度估计进行融合, 一次性完成对障碍物类别、定位和深度信息的全部输出。首先将双目相机抓取的左右图像分别输入改进的 YOLOv3 模型中进行障碍物检测, 输出农业环境下障碍物的类别和定位信息并进行目标匹配, 得到障碍物在左右图像中的对应关系。最后根据障碍物的对应关系计算像素视差, 并输入至双目成像模型进行深度估计, 算法流程如图 4 所示。

2.1 改进 YOLOv3 模型

本文选用 YOLOv3 模型^[28-30]作为深度估计的前端框架, 并针对双目成像模型和农业环境目标的特殊性进行锚框聚类 and 边界框损失函数的改进。

2.1.1 YOLOv3

YOLOv3 是目前为止最先进的目标检测算法之一, 其在 YOLOv2 的基础之上进一步融合多尺度检测和多尺度训练等改进措施, 使用更深的 darkNet53 作为特征提取网络, 并加入残差模块解决深层网络梯度问题。图 5 是 YOLOv3 的网络结构, 它从 3 个不同的尺度 (32×32 、 16×16 和 8×8) 提取特征至 YOLO 层进行检测。

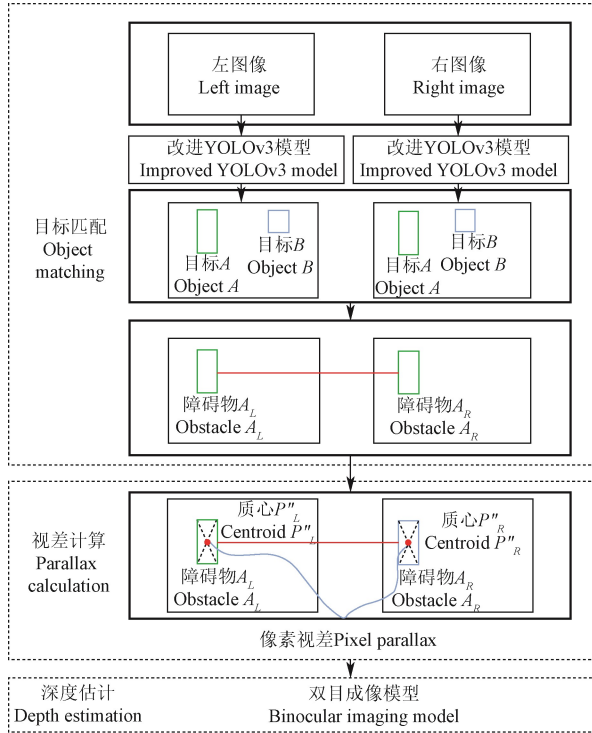


图4 深度估计算法流程

Fig.4 Algorithm flow chat of depth estimation

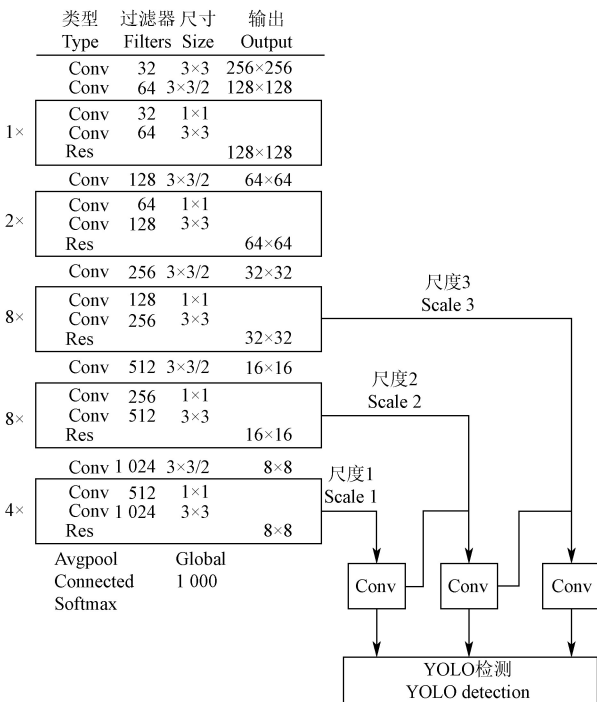


图5 YOLOv3 网络结构

Fig.5 Network structure of YOLOv3

2.1.2 数据集聚类

由于没有公开的农业障碍物数据集，本文在农业环境中设置人、农具和树桩等作为障碍物，从2020年开始通过网络爬虫和相机抓图自建数据集开展研究工作，并利用Labellmg进行类别和检测框标注，包括训练集2400张和测试集270张。由于农业环境中的障碍物形态多变，原COCO数据集中的先验框尺寸难以满足该数据集。为获得合适的先验框，本文采用K-means算法对农业障碍

物数据集进行聚类分析，并针对特征金字塔在52×52、26×26和13×13特征图上分别应用先验框(16,28)、(25,42)、(39,43)、(69,75)、(99,80)、(88,113)、(92,123)、(155,149)、(199,236)。

2.1.3 $loss_D$ 损失函数

原始YOLOv3模型的损失函数中，针对检测框 x 、 y 、 w 和 h 坐标的误差权值是相同的，但在双目成像模型中深度估计的精度取决于视差，即左右图像目标检测框质心的 u 轴坐标差，使得YOLOv3模型对检测框的横向偏差特别敏感。因此，在深度估计的边界框损失函数 $loss_D$ 中独立出 x 坐标的误差权值项并提高，同时降低其他3项的权值。

针对大小目标对目标检测框精度的影响不一致，YOLOv3原文采用反向赋权的方式来控制精度，即目标越大 x 、 y 、 w 和 h 坐标的误差权值越小。针对机械在农业环境下自主导航的需求，要求远处目标深度估计误差偏大而近处误差偏小，在损失函数 $loss_D$ 中对 x 误差项正向赋值，其余项固定赋值。改进后的损失函数如式(1)所示。

$$loss_D = \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} [(3 + w_i \cdot h_i)(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 + \sum_{r=0}^{n-1} ((r == truth_{class}) ? 1 : 0 - predict_{class_r})^2 + (truth_{conf} - predict_{conf})^2] \quad (1)$$

式中 K 为输入层网格数； M 为单个网格预测的锚框数； I_{ij}^{obj} 为目标的判断标志； w 、 h 、 (x, y) 和 \hat{w} 、 \hat{h} 、 (\hat{x}, \hat{y}) 分别为目标检测框宽、高、质心坐标的真值和预测值，像素； n 为目标类别数； r 为当前类别索引； $truth_{class}$ 、 $truth_{conf}$ 和 $predict_{class_r}$ 、 $predict_{conf}$ 分别为类别和置信度的真值和预测值； $(\cdot) ? 1 : 0$ 表示括号内判断条件为真则为1，反之则为0。

2.2 深度估计过程

2.2.1 目标匹配

相机抓取的图像中有些物体是拖拉机的作业目标，如：果实、庄稼和果蔬等；另一些物体是障碍物，如：行人、农具、树桩等。目标匹配是改进模型检测出选定的障碍物后，进一步在左右图像中确定同一障碍物的对应关系，具体过程如下：1)将左右相机图像 P_L 、 P_R 输入改进YOLOv3模型，输出各自的目标检测框 $BBOX_L$ 、 $BBOX_R$ 及其类别 $CLASS_L$ 、 $CLASS_R$ ；2)计算目标检测框的像素面积 S_L 、 S_R 和目标检测框质心的像素坐标 C_L 、 C_R ；3)得到目标检测框的像素面积之差 S_E 和目标检测框质心的 v 轴坐标之差 V_E ；4)判断 $CLASS_L$ 、 $CLASS_R$ 是否相同，且 S_E 、 V_E 是否小于阈值 A 、 B ；若同时满足，则匹配成功，反之匹配失败。

2.2.2 像素视差计算

若左右图像中2个目标检测框匹配成功，即表示其为某一障碍物在左右相机的成像位置，反之则继续抓图进行目标匹配。如图6所示，目标检测框 $BBOX_L$ 、 $BBOX_R$

为某一障碍物在左右相机的像素平面成像位置, 选用目标检测框的质心代表障碍物, 左右质心 P_L'' 、 P_R'' 的坐标分别为 (u_L, v_L) 、 (u_R, v_R) , 则障碍物在左右相机的像素视

差 D 为

$$D = u_L - u_R \quad (2)$$

式中 u_L 、 u_R 分别为像素平面上左右质心的 u 轴坐标。

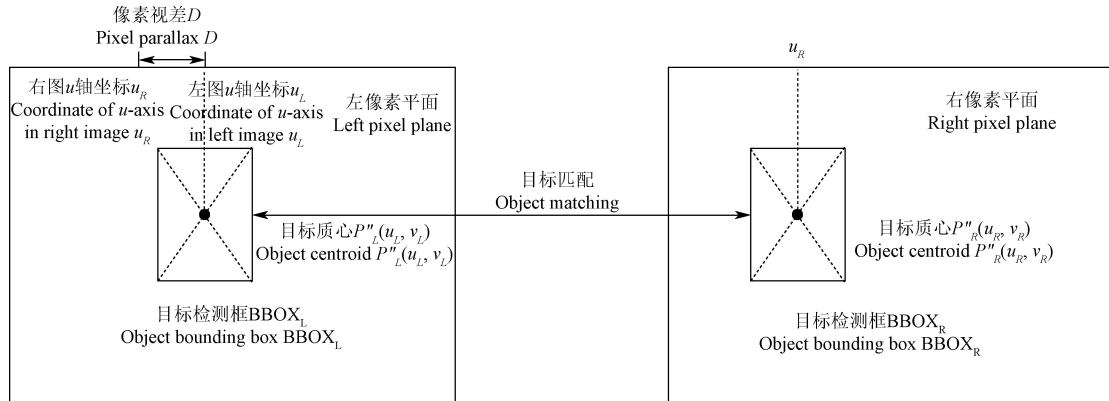


图6 左右相机目标匹配原理

Fig.6 Object matching principle in left and right cameras

2.2.3 深度估计

立体视觉中, 记 O_L 、 O_R 为左右相机光圈中心; X_L 、 X_R 为成像平面的坐标; f 为焦距, m ; z 为深度, m ; b 为左右相机的基线, m ; 空间点 P 在左右相机中各成一像, 记 P_L 、 P_R , 左右图像横坐标之差 $(X_L - X_R)$ 为视差 d , m ; 根据三角相似关系有:

$$z = \frac{fb}{d} = \frac{fb}{X_L - X_R} \quad (3)$$

目标边界框质心在左右视图中的像素坐标 u_L 、 u_R 分别为

$$\begin{cases} u_L = \alpha X_L + c_x \\ u_R = \alpha X_R + c_x \end{cases} \quad (4)$$

式中 α 为物理成像平面坐标系的横向缩放系数, 像素/ m ; c_x 为原点横向平移量, 像素。

进一步得视差 d 为

$$d = X_L - X_R = \frac{D}{\alpha} \quad (5)$$

将式 (5) 代入式 (3) 得深度 z 为

$$z = \frac{f\alpha \cdot b}{D} = \frac{f_x \cdot b}{u_L - u_R} \quad (6)$$

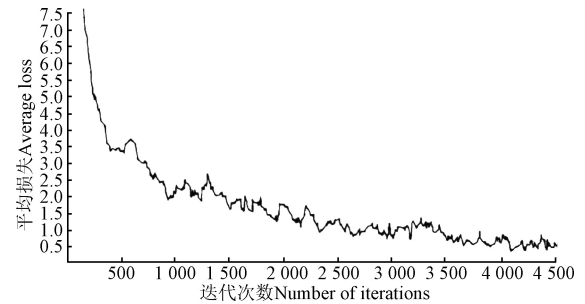
3 视觉感知系统试验

3.1 模型训练与测试试验

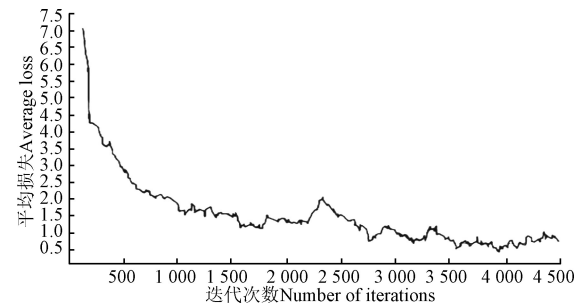
针对训练集, 本文选用 DELL T7920 型图形工作站 (12G 内存 TITAN V 型显卡) 对 YOLOv3 模型和改进 YOLOv3 模型分别进行相同的迭代训练, 2 种模型的训练损失函数如图 7 所示。

表 1 是 YOLOv3 模型和改进 YOLOv3 模型结果。由表 1 可知, 相较于 YOLOv3 模型, 改进 YOLOv3 模型的准确率和召回率分别下降 0.5% 和 0.4%。原因是本文对模型的改进侧重于目标检测框 x 轴上的精度, 而适当降低对 y 轴和宽高的精度, 提高了目标的深度估计精度。同时, 模型改进前后在深度学习工作站和嵌入式终端 TX2 上的

速度测试结果基本一致。



a. YOLOv3 模型
a. YOLOv3 model



b. 改进 YOLOv3 模型
b. Improved YOLOv3 model

图7 模型的训练损失

Fig.7 Training loss of models

表1 训练集测试结果

Table 1 Test results on training set

模型 Models	准确率 Accuracy/%	召回率 Recall/%	DELL T7920 帧率 DELL T7920 frame rate/(帧·s ⁻¹)	TX2 帧率 TX2 frame rate/(帧·s ⁻¹)
原始 YOLOv3 模型 Original YOLOv3 model	89.99	90.54	30.2	3.6
改进 YOLOv3 模型 Improved YOLOv3 model	89.54	90.18	30.4	3.5

3.2 深度估计试验

3.2.1 参数标定与评价指标

试验在南京农业大学农田试验场进行, 分别将Hog+SVM模型、YOLOv3模型和改进YOLOv3模型部署至嵌入式终端并对点火驻车状态下拖拉机前方1.6~3.4 m距离段上的静止障碍物进行深度估计。试验过程中, 保持双目相机与障碍物的图像质心在同一水平面上, 且双目相机左右光心的中点与图像质心的连线垂直于相机基线方向。对障碍物进行多组深度估计试验, 获得障碍物的深度估计值, 并将UT393A型测距仪的测量值(精度 ± 1.5 mm)作为距离真值进行误差分析。试验中采用误差均值 e_m 和误差比均值 e_{rm} 作为深度估计精度的指标, 其定义分别如下:

$$e_m = \frac{\sum_{i=1}^n |z_i - z|}{n} \quad (7)$$

$$e_{rm} = \frac{\sum_{i=1}^n \left| \frac{z_i - z}{z} \right|}{n} \quad (8)$$

式(7)~(8)中 z_i 为深度估计值, n 为测量次数, 本文试验取 $n=3$ 。

利用MYNTAI相机自带的软件开发工具包(Software

Development Kit, SDK)对系统视觉感知模块中的双目相机参数进行标定, 并获得去畸变的双目图像, 结果如表2所示。

表2 双目相机标定结果

相机 Cameras	u轴尺度因子 u-axis scale factor f_x / 像素	v轴尺度因子 v-axis scale factor f_y / 像素	u轴平移量 u-axis translation c_x / 像素	v轴平移量 v-axis translation c_y / 像素
左相机 Left camera	360.91	359.51	370.14	222.87
右相机 Right camera	361.08	361.72	369.17	222.13

3.2.2 深度估计

图8是不同障碍物(人、农具和树桩)在不同1.6~3.4 m距离段上的检测结果, 其中像素面积差 S_E 的阈值 A 和v轴坐标差 V_E 的阈值 B 分别设为经验值60和4。图8a障碍物目标为人, 目标深度真值为2.8 m, 检测框质心视差为15.47像素; 图8b障碍物目标为农具, 目标深度真值为2.6 m, 检测框质心视差为16.66像素; 图8c障碍物目标类别为树桩, 目标深度真值为2.9 m, 检测框质心视差为14.94像素。由图8可知, YOLOv3模型和改进YOLOv3模型的检测精度高且鲁棒性强, 而Hog+SVM模型的检测精度和鲁棒性均较差。

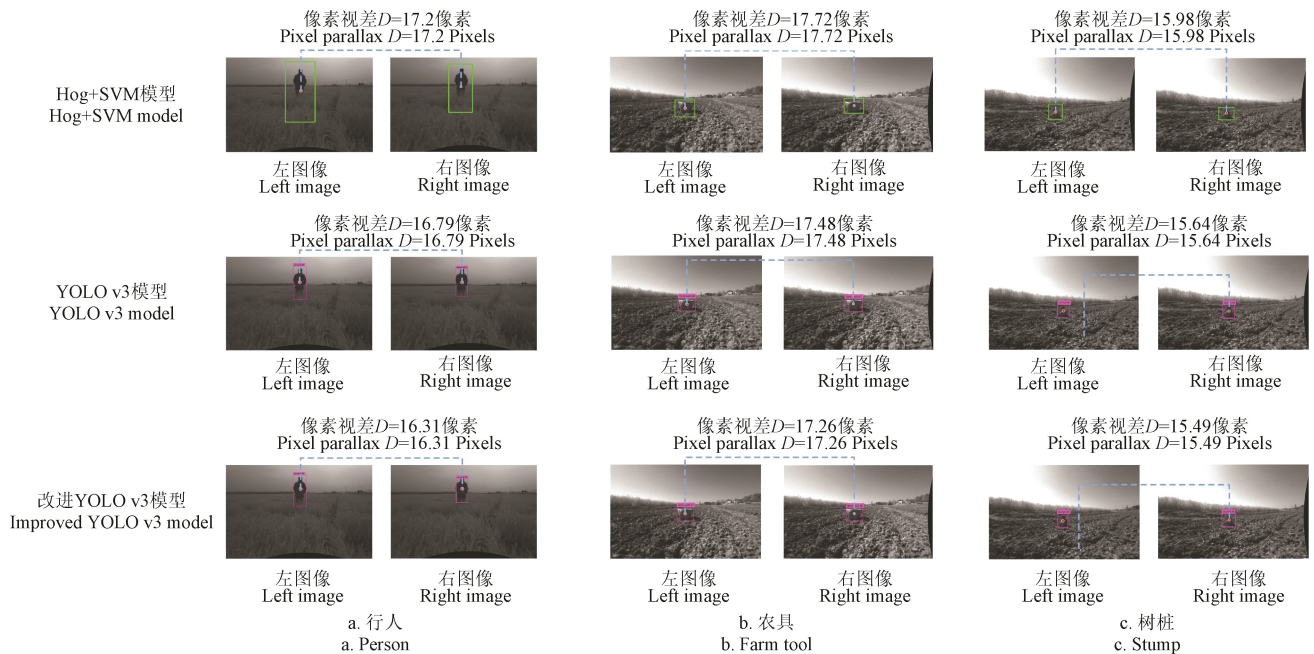


图8 不同模型对不同障碍物的静态深度估计结果对比

Fig.8 Comparison of static depth estimation results of different obstacles by different models

图9为不同障碍物(人、农具和树桩)在其深度估计范围内1.6~3.4 m距离段内的深度误差结果。由图9可知, 3种障碍物应用改进YOLOv3模型的深度估计误差均值 e_m 和误差比均值 e_{rm} 相对于YOLOv3模型和Hog+SVM模型均有降低。改进YOLOv3模型对行人的误差均值 e_m 和误差比均值 e_{rm} 相较于YOLOv3模型分别降低38.92%、37.23%, 比Hog+SVM模型分别降低53.44%、53.14%; 改进YOLOv3模型对农具

的误差均值 e_m 和误差比均值 e_{rm} 相较于YOLOv3模型分别降低26.47%、26.12%, 比Hog+SVM模型分别降低41.9%、41.73%; 改进YOLOv3模型对树桩的误差均值 e_m 和误差比均值 e_{rm} 相较于YOLOv3模型分别降低25.69%、25.65%, 比Hog+SVM模型分别降低43.14%、43.01%; 随着障碍物目标与相机之间距离增大, 3种模型的深度估计误差均值 e_m 和误差比均值 e_{rm} 均无明显变化规律。

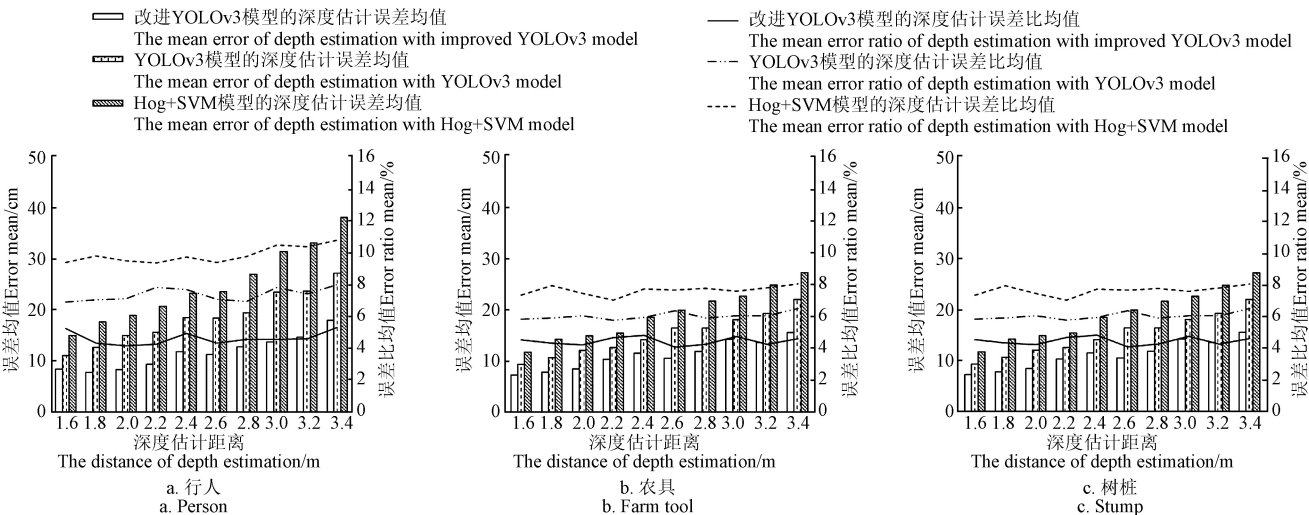


图 9 不同模型对不同障碍物的深度估计误差
Fig.9 The depth estimation errors of different models for different obstacles

3.3 动态避障试验

试验于 2020 年 5 月在南京农业大学农机试验场进行, 对动态场景下视觉感知系统的深度估计精度和避障效果进行测试。由于行驶中的拖拉机与障碍物行人之间的深度真值无法用测距仪实时测得, 试验中在拖拉机车尾部署一个 GPS 接收机, 同时行人手持一个 GPS 接收机, 并用 2 个 GPS 定位坐标之间的距离除去拖拉机 GPS 接收机至车头的距离 (289 cm) 作为实时的深度真值, 自主导航试验装置如图 10 所示。



图 10 试验装置
Fig.10 Testing apparatus

由于计算机处理有效数字长度有限, 将坐标原点从 117°E 与赤道的交点向东平移 159 000 m, 向北平移 3 557 000 m, 并以平移后的点为参考原点。试验中拖拉机按导航基线 AB 由 A 向 B 直线行驶, 起点 A 坐标 (227.198 6, 176.401 6), 平均速度 0.31 m/s, 障碍物行人沿导航基线由 B 向 A 行走, 起点 B 坐标 (227.173 9, 182.225 8), 平均速度 0.18 m/s。为保证行人安全, 将深度预警值设为 2.4 m, 相机抓图的频率设置为 1 Hz, 电控液压转向模块转角为 18°。本次试验从拖拉机点火启动开始, 至拖拉机转向避障后障碍物离开相机视角结束, 共耗时 9 s。拖拉机启动时深度值为 5.22 m, 之后每秒输出 1 次深度估计值, 在行驶 6 s 后深度估计结果为 2.14 m, 小于试验设置的预警值并触发电控液压转向模块以固定右转角 18°进行安全避障, 执行避障动作 3 s 后障碍物离开相机视角同时系统停止深度估计, 试验结果如图 11 所示。

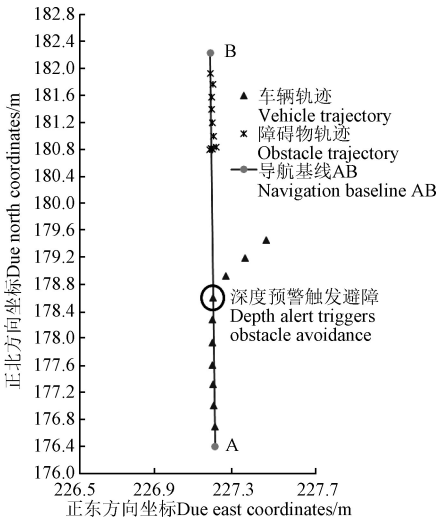


图 11 自主导航与避障试验结果
Fig.11 Autonomous navigation and obstacle avoidance test results

表 3 为动态试验中深度估计精度的误差统计, 深度真值用 GPS 坐标间的欧氏距离除去车载 GPS 到相机的距离 289 cm 进行表示, 深度估计值为本文基于改进 YOLOv3 的深度估计结果。

表 3 动态场景深度估计 Table 3 Depth estimation in dynamic scene					
行驶时间 Travel time/s	深度 Depth/ $\times 10^{-2}$ cm		误差均值 Mean error e_m /cm	误差比均值 Mean error ratio $e_{rm}/\%$	耗时 Time- consuming/s
	估计值 Estimation value	真值 True value			
0	549	522	27	5.17	0.59
1	502	476	26	5.46	0.58
2	444	425	19	4.95	0.56
3	362	379	17	4.49	0.58
4	341	325	16	4.92	0.59
5	284	271	13	4.80	0.58
6	214	223	11	4.93	0.55
7	191	184	7	3.80	0.53
8	160	153	7	4.58	0.58
9	124	129	5	3.88	0.58

由表 3 可知, 随着自主导航时间的增加, 拖拉机与障碍物之间的深度真值和深度估计值不断减小且变化趋势一致; 同时, 误差均值 e_m 从初始的 27 cm 不断减少至 5 cm, 但平均误差比均值 e_{rm} 为 4.66%, 无明显变化规律, 维持在 6% 以下, 比静态深度估计的误差比均值降低 7.19%, 算法平均耗时为 0.573 s。

4 结 论

1) 本文设计了一套基于农业机械的障碍物视觉感知系统, 系统将嵌入式 AI 计算机作为控制核心, 极大节省了机械的空间资源, 能够对障碍物的类别和深度信息进行准确检测。

2) 本文提出一种基于改进 YOLOv3 模型的深度估计方法, 通过增大模型对图像 x 轴的敏感程度来提高深度估计精度, 相对于 YOLOv3 模型和传统检测方法 Hog+SVM 其深度估计的误差均值、误差比均值均有较大改善。动态避障试验结果表明, 随着行驶过程的进行, 误差均值从初始的 27 cm 不断减少至 5 cm, 误差比均值无明显变化规律, 但始终维持在 6% 以下, 比静态深度估计的误差比均值降低 7.19%, 成功避障。

由于 YOLOv3 模型的网络结构深参数量大, 在嵌入式终端的实时推理速度有限。后续研究中将选用更加轻量级的 YOLOv3-tiny 模型和运算力更高的终端 xavier, 进行动态障碍物的深度估计试验, 不断提高平台的实时推理帧率。

[参 考 文 献]

- [1] 姬长英, 周俊. 农业机械导航技术发展分析[J]. 农业机械学报, 2014, 45(9): 44-54.
Ji Changying, Zhou Jun. Current situation of navigation technologies for agricultural machinery[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(9): 44-54. (in Chinese with English abstract)
- [2] 张美娜, 吕晓兰, 陶建平, 等. 农用车辆自主导航控制系统设计与试验[J]. 农业机械学报, 2016, 47(7): 42-47.
Zhang Meina, Lv Xiaolan, Tao Jianping, et al. Design and experiment of automatic guidance control system in agricultural vehicle[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(7): 42-47. (in Chinese with English abstract)
- [3] 赵博, 王猛, 毛恩荣, 等. 农业车辆视觉实际导航环境识别与分类[J]. 农业机械学报, 2009, 40(7): 166-170.
Zhao Bo, Wang Meng, Mao Enrong, et al. Recognition and classification for vision navigation application environment of agricultural vehicle[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(7): 166-170. (in Chinese with English abstract)
- [4] 马兆青, 袁曾任. 基于栅格方法的移动机器人实时导航和避障[J]. 机器人, 1996(6): 25-29.
Ma Zhaoqing, Yuan Cengren. Real-time navigation and obstacle avoidance based on grids method for fast mobile robot[J]. Robot, 1996(6): 25-29. (in Chinese with English abstract)
- [5] 胡静涛, 高雷, 白晓平, 等. 农业机械自动导航技术研究进展[J]. 农业工程学报, 2015, 31(10): 1-10.
Hu Jingtao, Gao Lei, Bai Xiaoping, et al. Review of research on automatic guidance of agricultural vehicles[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(10): 1-10. (in Chinese with English abstract)
- [6] 姬长英, 沈子尧, 顾宝兴, 等. 基于点云图的农业导航中障碍物检测方法[J]. 农业工程学报, 2015, 31(7): 173-179.
Ji Changying, Shen Ziyao, Gu Baoxing, et al. Obstacle detection based on point clouds in application of agricultural navigation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(7): 173-179. (in Chinese with English abstract)
- [7] 薛金林, 董淑娴, 范博文. 基于信息融合的农业自主车辆障碍物检测方法[J]. 农业机械学报, 2018, 49(S1): 29-34.
Xue Jinlin, Dong Shuxian, Fan Bowen. Detection of obstacles based on information fusion for autonomous agricultural vehicles[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(S1): 29-34. (in Chinese with English abstract)
- [8] 李乐, 张茂军, 熊志辉, 等. 基于内容理解的单幅静态街景图像深度估计[J]. 机器人, 2011, 33(2): 174-180.
Li Le, Zhang Maojun, Xiong Zhihui, et al. Depth estimation from a single still image of street scene based on content understanding[J]. Robot, 2011, 33(2): 174-180. (in Chinese with English abstract)
- [9] 丁伟利, 李勇, 王文锋, 等. 基于轮廓特征理解的城市道路图像深度估计[J]. 光学学报, 2014, 34(7): 173-179.
Ding Weili, Li Yong, Wang Wenfeng, et al. Depth estimation of urban road image based on contour understanding[J]. Acta Optica Sinica, 2014, 34(7): 173-179. (in Chinese with English abstract)
- [10] 徐越, 李盈慧, 宋怀波, 等. 基于 Snake 模型与角点检测的双果重叠苹果目标分割方法[J]. 农业工程学报, 2015, 31(1): 196-203.
Xu Yue, Li Yinghui, Song Huaibo, et al. Segmentation method of overlapped double apples based on Snake model and corner detectors[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(1): 196-203. (in Chinese with English abstract)
- [11] 李小红, 谢成明, 贾易臻, 等. 基于 ORB 特征的快速目标检测算法[J]. 电子测量与仪器学报, 2013, 27(5): 455-460.
Li Xiaohong, Xie Chengming, Jia Yizhen, et al. Rapid moving object detection algorithm based on ORB features[J]. Journal of Electronic Measurement and Instrumentation, 2013, 27(5): 455-460. (in Chinese with English abstract)
- [12] 孙锐, 陈军, 高隼. 基于显著性检测与 HOG-NMF 特征的

- 快速行人检测方法[J]. 电子与信息学报, 2013, 35(8): 1921-1926.
- Sun Rui, Chen Jun, Gao Jun. Fast pedestrian detection based on saliency detection and HOG-NMF features[J]. Journal of Electronics & Information Technology, 2013, 35(8): 1921-1926. (in Chinese with English abstract)
- [13] 李思雯, 吕建成, 倪胜巧. 集成的卷积神经网络在智能冰箱果蔬识别中的应用[J]. 数据采集与处理, 2016, 31(1): 205-212.
- Li Siwen, Lv Jiancheng, Ni Shengqiao. Integrated convolutional neural network and its application in fruits and vegetables recognition of intelligent refrigerator[J]. Journal of Data Acquisition and Processing, 2016, 31(1): 205-212. (in Chinese with English abstract)
- [14] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.
- Lu Hongtao, Zhang Qinchuan. Applications of deep convolutional neural network in computer vision[J]. Journal of Data Acquisition and Processing, 2016, 31(1): 1-17. (in Chinese with English abstract)
- [15] Williams H, Jones M H, Nejati M, et al. Robotic kiwifruit harvesting using machine vision, convolutional neural networks and robotic arms[J]. Biosystems Engineering, 2019, 181: 140-156.
- [16] 薛金林, 闫嘉, 范博文. 多类农田障碍物卷积神经网络分类识别方法[J]. 农业机械学报, 2018, 49(S1): 35-41.
- Xue Jinlin, Yan Jia, Fan Bowen. Classification and identification method of multiple kinds of farm obstacles based on convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(S1): 35-41. (in Chinese with English abstract)
- [17] 薛金林, 张顺顺. 基于激光雷达的农业机器人导航控制研究[J]. 农业机械学报, 2014, 45(9): 55-60.
- Xue Jinlin, Zhang Shunshun. Navigation of an agricultural robot based on laser radar[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(9): 55-60. (in Chinese with English abstract)
- [18] 张漫, 苗艳龙, 仇瑞承, 等. 基于车载三维激光雷达的玉米点云数据滤波算法[J]. 农业机械学报, 2019, 50(4): 170-178.
- Zhang Man, Miao Yanlong, Qiu Chengrui, et al. Maize point cloud data filtering algorithm based on vehicle 3D LiDAR[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 170-178. (in Chinese with English abstract)
- [19] 李诗锐, 李琪, 李海洋, 等. 基于 Kinect v2 的实时精确三维重建系统[J]. 软件学报, 2016, 27(10): 2519-2529.
- Li Shirui, Li Qi, Li Haiyang, et al. Real-time accurate 3D reconstruction based on Kinect v2[J]. Journal of Software, 2016, 27(10): 2519-2529. (in Chinese with English abstract)
- [20] 孟明, 杨方波, 余青山, 等. 基于 Kinect 深度图像信息的人体运动检测[J]. 仪器仪表学报, 2015, 36(2): 386-393.
- Meng Ming, Yang Fangbo, She Qingshan, et al. Human motion detection based on the depth image of Kinect[J]. Chinese Journal of Scientific Instrument, 2015, 36(2): 386-393. (in Chinese with English abstract)
- [21] 沈彤, 刘文波, 王京. 基于双目立体视觉的目标测距系统[J]. 电子测量技术, 2015, 38(4): 52-54.
- Shen Tong, Liu Wenbo, Wang Jing. Distance measurement system based on binocular stereo vision[J]. Electronic Measurement Technology, 2015, 38(4): 52-54. (in Chinese with English abstract)
- [22] 杜钊君, 吴怀宇. 基于激光测距与双目视觉信息融合的移动机器人 SLAM 研究[J]. 计算机测量与控制, 2013, 21(1): 180-183.
- Du Zhaojun, Wu Huaiyu. Research of mobile robot SLAM based on laser range finder and binocular vision data fusion[J]. Computer Measurement & Control, 2013, 21(1): 180-183. (in Chinese with English abstract)
- [23] 何勇, 蒋浩, 方慧, 等. 车辆智能障碍物检测方法及其农业应用研究进展[J]. 农业工程学报, 2018, 34(9): 21-32.
- He Yong, Jiang Hao, Fang Hui, et al. Research progress of intelligent obstacle detection methods of vehicles and their application on agriculture[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(9): 21-32. (in Chinese with English abstract)
- [24] Giubilato R, Chiodini S, Pertile M, et al. An evaluation of ROS-compatible stereo visual SLAM methods on a Nvidia Jetson TX2[J]. Measurement, 2019, 140: 161-170.
- [25] Tang J X, Ericson L, Folkesson J, et al. GCNv2: Efficient correspondence prediction for real-time SLAM[J]. IEEE Robotics and Automation Letters, 2019, 4(4): 3505-3512.
- [26] Goyal M, Reeves N D, Riajbhandar S, et al. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 23(4): 1730-1741.
- [27] Fu Y W, Jha D K, Zhang Z Y, et al. Neural network-based learning from demonstration of an autonomous ground robot[J]. Machines, 2019, 7(2): 24.
- [28] Redmon J F. YOLOv3: An incremental improvement[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [29] 戴伟聪, 金龙旭, 李国宁, 等. 遥感图像中飞机的改进 YOLOv3 实时检测算法[J]. 光电工程, 2018, 45(12): 84-92.
- Dai Weicong, Jin Longxu, Li Guoning, et al. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3[J]. Opto-Electronic Engineering, 2018, 45(12): 84-92. (in Chinese with English abstract)
- [30] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报, 2019, 35(3): 164-173.
- Zhao Dean, Wu Rendi, Liu Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 164-173. (in Chinese with English abstract)

Design and experiments of the binocular visual obstacle perception system for agricultural vehicles

Wei Jiansheng¹, Pan Shuguo^{1*}, Tian Guangzhao², Gao Wang¹, Sun Yingchun¹

(1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China;

2. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China)

Abstract: Machine learning was efficiently incorporated to design a visual perception system for obstacle-free path planning in agricultural vehicles. The present system aims to ensure the safety and reliability of intelligent agricultural vehicles in the process of autonomous navigation. Hardware and software were mainly included in the system. The hardware consisted of visual perception and navigation control module. Since the visual perception task needed real-time image processing, the embedded AI computer Jetson TX2 was taken especially as the core of computing to operate. A deep Convolutional Neural Network (CNN) was used to identify agricultural obstacles. The complex structure and uneven illumination were considered in the agricultural environment, thereby enhancing stability in object detection. The CNN performance of environmental features was much better, compared with the traditional detection using artificially designed features. Moreover, better detection was achieved under continuous learning features in the current task from the large-scale dataset. The improved YOLOv3 was utilized to integrate object detection for the simultaneous output of all information, including category, location, and depth estimation. A binocular camera was used to capture the left and right images, all of which were firstly input into the improved YOLOv3 model for object detection. The output of the improved YOLOv3 model was used for object matching to complete obstacle recognition, where the relationship of obstacles was determined in the left and right images. The location of matching objects was then used to calculate the parallax of the obstacle between left and right images. Finally, the parallax of the obstacle was input into the binocular imaging model for depth estimation. The accuracy of depth estimation was improved, with the increase of model sensitivity to the X -axis of images. The mean error, mean error ratio, and mean square error of depth estimation were greatly improved, compared with the original YOLOv3 and HOG+SVM model. The experimental results showed that the embedded AI computer-processed images in real-time, ensuring the detection accuracy of the improved YOLOv3 model. In object detection, a highly accurate identification was achieved in the agricultural obstacles with an average accuracy rate of 89.54%, and a recall rate of 90.18%. In the first kind of obstacle, the mean error and mean error ratio of the improved YOLOv3 model were 38.92% and 37.23% lower than those of the original one, while 53.44% and 53.14% lower than those of the HOG+SVM model, respectively. In the second kind of obstacle, the mean error and mean error ratio of the improved YOLOv3 model were 26.47% and 26.12% lower than those of the original one, while 41.9% and 41.73% lower than those of the HOG+SVM model, respectively. In the third kind of obstacle, the mean error and mean error ratio of the improved YOLOv3 model were 25.69% and 25.65% lower than those of the original one, while 43.14% and 43.01% lower than those of the HOG+SVM model, respectively. In addition, there was no obvious change in the mean error, mean error ratio, and mean square error of the three models, when changing the distance between obstacle and vehicle. The average error ratio was 4.66% in the depth estimation of obstacles under the dynamic scenario, and the average time was 0.573 s. An electrically controlled hydraulic steering was also used in time for obstacle avoidance during depth warning. The findings can provide an effective basis for environment perception for agricultural vehicles in autonomous navigation. In the following research, the more lightweight YOLOv3-tiny model and the terminal processor Xavier with higher computing power can be selected to conduct the depth estimation, aiming to increase the real-time inference speed of visual perception system in modern agriculture.

Keywords: agricultural machinery; image processing; obstacle perception; depth estimation; parallax calculation