

## A 3-D Visual Recognition System Assisted by Radio Frequency Technology

ZHANG Zhi<sup>1\*</sup>, LIU Ziyu<sup>2</sup>, QIU Linglong<sup>2</sup>, WANG Xuanye<sup>3</sup>, DONG Xu<sup>1</sup>

(1. School of Information Engineering, Zhejiang College of Zhejiang University of Technology, Shaoxing Zhejiang 312030, China;

2. School of Computer Science, Zhejiang University of Technology, Hangzhou Zhejiang 310023, China;

3. School of Information Engineering, Zhejiang University of Technology, Hangzhou Zhejiang 310023, China)

**Abstract:** In the field of visual-based object recognition, monocular visual recognition can only obtain partial information of objects, and the computational complexity of binocular visual recognition is high. With the popularization of Internet of Things (IoT) technologies, passive ultra-high frequency (UHF) radio frequency identification (RFID) technology has been widely used in object identification, and has the advantages of fast reading speed and long reading distance. In order to optimize the speed and accuracy of object recognition, this paper proposes a general method to assist visual recognition through prior information. By identifying RFID tags, accurate feature information is read from the database to assist image recognition of objects. The camera collects the image and video information of the object and then transmits it to the control module. The control module obtains the RFID prior information from the database and then corrects the image with the camera calibration algorithm to locate the target object, so as to reproduce the 3-dimensional (3D) image of the object. Traditional edge detection and target detection technologies require two or more cameras to recognize an object in 3D. The proposed method only needs one camera to obtain its 3D position. The improved edge detection algorithm combines the accurate object geometry information and pixel information in RFID tags to determine the weight of the filtering window and to calibrate the real edge and potential edge. The improved target detection algorithm introduces the feature pyramid on the region proposal network (RPN) architecture of the original Faster R-CNN, which makes feature extraction semantics stronger. Finally, the experimental results of two different visual recognition methods demonstrate the effectiveness of the proposed method. The proposed method has higher accuracy of recognition and positioning, lower algorithm complexity and faster identification rate, and can more accurately and reliably detect the characteristics of objects and estimate their shapes and directions.

**Key words:** RFID; visual recognition; prior information; 3D recognition; machine learning

EEACC: 7230

doi: 10.3969/j.issn.1004-1699.2021.09.019

## 射频技术辅助的三维视觉识别系统

张智<sup>1\*</sup>, 刘子瑜<sup>2</sup>, 邱灵龙<sup>2</sup>, 王榭焯<sup>3</sup>, 董旭<sup>1</sup>

(1. 浙江工业大学之江学院信息工程学院, 浙江 绍兴 312030; 2. 浙江工业大学计算机学院, 浙江 杭州 310023;

3. 浙江工业大学信息工程学院, 浙江 杭州 310023)

**摘要:** 在基于视觉的物体识别领域, 单目视觉识别技术往往只能获得物体片面信息, 而多目视觉识别技术的运算复杂度较高。随着物联网技术的普及, 无源超高频射频识别技术已经大规模应用于物体的标识上, 具有的读取速度快、读取距离远的优势。本文提出一种通过先验信息辅助视觉识别的通用方法, 提高物体识别的速度和准确性。通过识别射频识别标签, 从数据库读取准确的特征信息辅助图像识别物体。通过摄像头采集物体的图像、视频信息后传输给控制模块, 控制模块从数据库获取射频识别先验信息后以相机标定算法对图像进行矫正处理, 对目标物体进行定位, 从而复现物体的三维图像。传统的边缘检测和目标检测技术需要两个及以上的摄像头才能对物体进行三维识别, 所提方法只需使用一个摄像头即可获取物体三维位置。针对边缘检测的改进中通过结合射频识别标签中准确的物体几何信息和像素信息来确定滤波窗口的权重, 进行标

定真实边缘和潜在边缘;针对目标检测的改进中在原始的 Faster R-CNN 的 RPN 架构上引入了特征金字塔,使得特征提取时语义更强。最后两种不同视觉识别方式的实验结果证明了所提方法的有效性,所提方法具有更高的识别定位精准度、更低的算法复杂度和更快的识别速率,可以更加准确可靠地对物体特性进行检测及形状、方位的判断。

**关键词:** 射频识别;视觉识别;先验信息;三维识别;机器学习

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1004-1699(2021)09-1270-09

近年来,视觉识别被广泛应用于视觉定位<sup>[1]</sup>、物品检测<sup>[2]</sup>、生物鉴别<sup>[3]</sup>等。随着智慧城市化、快递货运行业等方向的高速发展,视觉识别应用于机器人代替人工从事重复繁杂的工作非常普遍,且逐渐产生分拣机器人、自动包装机器人和视障辅助识别等多元化应用。

随着社会需求的增加,视觉识别的功能与性能也在逐步上升。利用视觉识别技术结合网络化和信息化的模式实现自动分拣成为安全监控、产品检查和自动包装等行业的主要研究和方向之一<sup>[4-6]</sup>。在对物体进行精准操作的应用中,关注物体整体三维形状是十分必要的。通过单目摄像头只能做到二维目标识别,部分研究中通过单目摄像头完成的三维物体识别需要结合图像处理算法获取其他信息,实质上识别的图像还是二维的。视觉识别研究中通常需要多视角或多目摄像头实现物体在三维空间上的目标识别。如 Wiles 等<sup>[7]</sup>提出的从单个和多个视图预测雕塑的 3D 表面学习方法中,使用 2D 剪影图像仍然受图像尺寸的限制,而多视图预测需要合成多个视图来估计三维。三维目标识别一般具有两个以上的摄像机<sup>[8-9]</sup>,随着大量摄像头的安装,监管者需要持续关注视频屏幕并从中提取信息,这对于人员来说在耐力和工作强度上都存在挑战。此外,随着监控时间的增加,视频录像数据的不断增长对存储设备的性能提出了更高的要求。

射频识别(Radio Frequency Identification, RFID)<sup>[10]</sup>通过射频信号获得有关数据,属于无接触式自动识别技术,可有效缩短识别目标对象的时长。通过射频技术与视觉技术融合可以降低数据冗余<sup>[11]</sup>。但是当前关于 RFID 与视频结合的研究更倾向于利用 RFID 获取位置信息或者利用相位值这种无线信号的物理层属性来辅助视觉算法完成目标的定位或追踪任务<sup>[12-13]</sup>,没有充分利用 RFID 标签所关联的物体信息。

在对现有视觉识别技术进行深入研究的基础上,为了提高识别物体的速度和准确率,本文提出了一种可以在不同场景下通用的射频技术辅助的视觉识别系统,借助 RFID 自动识别和动态收集功能,将物体特征数据绑定于 RFID 标签作为先验信息存储于数据库,通过读取物体特征信息可以辅助摄像头

识别物体,以三维立体坐标形式复现物体所在平面的位姿。这样可以对物体进行综合判断,提高原始视频识别系统的智能性,可以广泛的应用在未来的自动分拣、视障辅助、机器辅助操纵等场景。本文主要的创新点包括:①提出了一种通用的多传感器融合的方法,能够通过采用射频标签所关联的先验信息辅助不同视觉识别算法。②以 Canny<sup>[14]</sup>和 Faster R-CNN<sup>[15]</sup>两种算法为例,详述了如何运用本文方法使用标签关联的物体信息作为先验信息辅助现有视觉算法完成单目三维识别,证明了本文方法的通用性。③对不同物体进行大量实验,验证了本文方法的有效性,与使用三维摄像头、多目摄像头实现三维识别的视觉算法相比,通过 RFID 辅助单目摄像头进行物体识别可以降低复杂度,并提升处理速度,多传感器的融合提高了识别的准确度。

## 1 设计原理

本文系统通过 RFID 标签标记物品特性,读取物品的重量、尺寸和材质等先验信息辅助摄像头对物体的识别、距离的判断等操作。如图 1 所示为本文系统框图,系统包含的模块有:传感器模块、控制模块、传输模块、数据库。其中传感器模块包含 RFID 单元(射频模块)和摄像头(视频模块),RFID 单元由电子标签、天线和读写器组成;以电脑作为控制模块;传输模块采用 RFID 读写器和摄像头上的有线网口通信。标签中的物品信息存储于数据库作为先验信息,方便识别物品时读写器读取数据,减少计算机分析视频识别物体的时间。

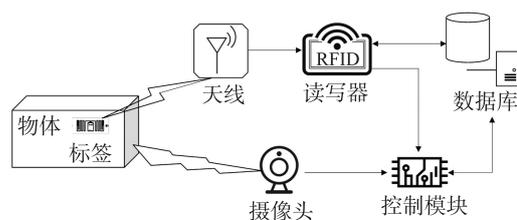


图 1 系统框图

### 1.1 射频识别原理

无接触式自动识别技术 RFID 是由 RFID 电子标签、读写器、天线三部分组成。无源超高频 RFID 系统采用的是反向散射调制,如图 2 所示,系统运行

时,读写器发射查询信号,无源电子标签获得查询信号后,将其中部分信号整流为直流电,用于为电子标签中的电路提供能量,另一部分信号由电子标签反射并调制后,将数据反馈给读写器。读写器的作用是读写电子标签中的内容,天线的作用是在电子标签和读写器间传送射频信号(能量和数据)<sup>[11]</sup>。

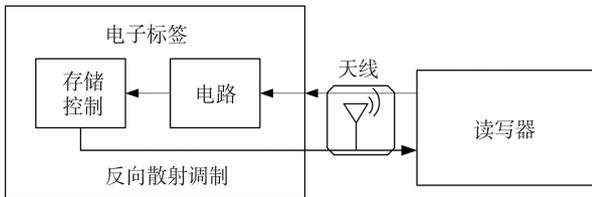


图 2 反向散射调制

## 1.2 视觉识别算法介绍

视觉识别在物体检测中的应用是机器视觉的经典研究方向之一,主要目的是得到图像中物体的目标框,以及给出物体的类别。不管是机器视觉中解决物体检测的基础方法边缘检测,还是更加成熟的深度学习目标检测算法,都无法做到单目摄像头的三维目标识别。

边缘检测在视觉识别中是十分重要的研究方向,经典的边缘检测包括微分法、最优法和拟合法。最常见的是微分法,一般被划分为一阶边缘算法和二阶边缘算法,如常用的 Sobel<sup>[16]</sup>算子属于一阶边缘算法,Sobel 具有较强的抗噪声干扰能力,在灰度渐变或者噪声强的场景中图像处理效果不错,但是定位准确度较低。Laplacian<sup>[17]</sup>和 Canny<sup>[14]</sup>算子属于比较有名的二阶边缘算法,Laplacian 算子在噪声干扰下容易丢失边缘梯度信息,因此更为常用的是 Canny 算子。

在使用 Canny 算法进行边缘检测任务之前,通常使用高斯函数对图像进行平滑处理,以降低高斯噪声,并设定分辨率,使得图像的强度变化能够快速被检测到,这个过程非常有利于检测到灰度变化最大的边缘。通过选择图像梯度可以定位这些最大边缘的区域,返回水平梯度值  $G_x$  和垂直梯度值  $G_y$ 。然后计算欧几里得距离,作为梯度大小  $G$ ,如式(1)所示。图像的边缘往往存在被延伸的情况,只确定梯度值不能很好地确定边缘,因此有必要确定边缘方向  $\theta$ ,计算方法如式(2)所示。

$$G = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (2)$$

但是,Canny 算法也存在一些缺点。例如无法识别物体周围的弱边缘,或者,由于噪声的存在,有可能识别出虚假的边缘。此外,它无法识别出分支

的边缘和一些重要的细节。Canny 的漏检率低于 Sobel,但误检率更高。

随着深度学习的发展,物体检测算法中会加入卷积神经网络(Convolutional Neural Network, CNN)框架对图像进行训练提升识别性能<sup>[18]</sup>。基于深度学习的视频识别过程中,首先要使用摄像头传感器采集视频信息,通过嵌入深度学习算法识别、检测视频图像。

普通的识别方法是对单帧图像进行深度学习表达,单个图像粒度通过 CNN 网络将得到一个输出,CNN 由输入层、卷积层、归一化层、池化层和全连接层组成。在静态的识别过程中,需要提取图像特征进行学习。而在快速运动的视频中,为了减轻背景对视频识别算法性能的影响,需要加强时间域上的学习。因此,为了提高识别精度,在运动量较明显的视频识别过程中,CNN 网络可以通过在更深层进行时间卷积,将同一层的参数传递给不同时间的同一层网络。

R-CNN(Regions with CNN features)<sup>[19]</sup>系列的目标检测算法一般包含候选区域提取、卷积特征提取、训练分类器和边界回归。前人所提的 R-CNN、SPP-Net<sup>[20]</sup>、Fast R-CNN<sup>[21]</sup>算法都对 CNN 在物体检测上的应用做出了重要贡献,但是仍然存在特征重复提取、耗时过大的问题。

Faster R-CNN 使用了区域生成网络(Region Proposal Network, RPN)<sup>[15]</sup>,和之前的 R-CNN 系列算法有所区别的是在特征图上进行候选区域提取,而不是在原图。Faster R-CNN 使用预训练模型训练后可以得到卷积网络层 VGG-16<sup>[22]</sup>,输入图像至卷积层后能够在第五个卷积 block 的第三个卷积层输出最终的特征图,然后使用 RPN 选择候选区域。结合特征图和候选区域的输出结果,并执行感兴趣区域池化(Region of interest pooling, RoI pooling)之后在全连接层可以实现目标的识别和定位。相比 R-CNN 和 Fast R-CNN, Faster R-CNN 在检测精度和耗时方面都有所改进,但对于每个目标候选框的计算量还是比较大。

## 2 射频辅助的三维视觉识别

本文提出的射频辅助视觉识别方法是一种通用的三维识别辅助方法,适用于改进任意物体视觉识别检测算法。本文以 Canny 边缘检测算法和 Faster R-CNN 目标检测算法来证明新提出的射频辅助视觉识别方法的通用性和有效性。

### 2.1 算法流程

射频辅助视觉识别程序流程图如图 3 所示。

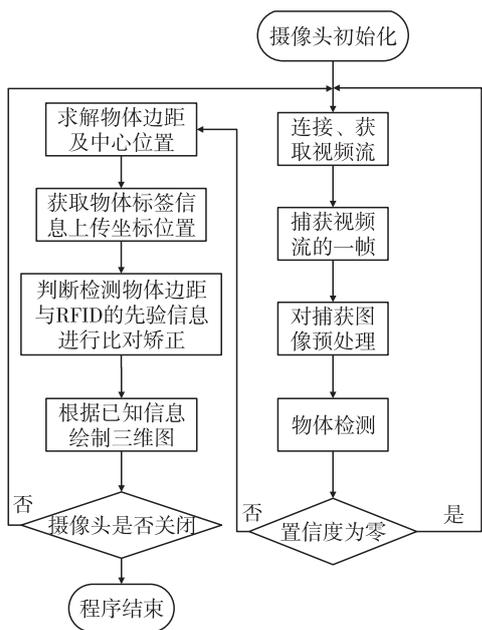


图 3 视频识别原理流程图

首先,启动摄像头后,开启视频流,然后捕获视频的一帧,根据选择的视觉识别算法对捕获的图像进行数据预处理,随后使用视觉识别算法能够较好的对物体进行标定。若置信度过低则重新检测。

在视觉识别过程中,由于相机拍摄画面的畸变容易造成图片桶形失真,因此本文采用相机标定算法<sup>[23]</sup>对图像进行矫正处理。首先将世界坐标转化为相机坐标,如式(3)所示,其中 $(X_w, Y_w, Z_w)$ 表示某点的世界坐标, $(X_c, Y_c, Z_c)$ 表示该点的相机坐标, $\mathbf{R}$ 为世界坐标系分别绕 $X, Y, Z$ 轴旋转的旋转矩阵之积。 $\mathbf{T}$ 表示偏移向量。

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{3 \times 3} & \mathbf{T}_{3 \times 1} \\ \mathbf{O} & 1 \end{bmatrix} \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3)$$

然后结合小孔成像原理和三角形相似性原理计算二维图像坐标,关系表达式如式(4)所示:

$$Z_c \begin{bmatrix} x_g \\ y_g \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (4)$$

式中, $f$ 为图像坐标系原点与相机坐标系原点之间的距离, $(x_g, y_g)$ 表示图像坐标。

最后,通过图像坐标计算像素坐标,虽然图像坐标系和像素坐标系都在成像的平面中,但是两者一般情况下是不垂直的(如图4所示),转换关系如式(5)所示, $(u, v)$ 表示像素坐标, $(u_0, v_0)$ 为像素坐标

系原点在图像坐标系中的坐标值, $\theta$ 为像素坐标系的夹角。

$$\begin{cases} u = u_0 + \frac{x_g}{dx} - \frac{y_g \cot \theta}{dx} \\ v = v_0 + \frac{y_g}{dy \sin \theta} \end{cases} \quad (5)$$

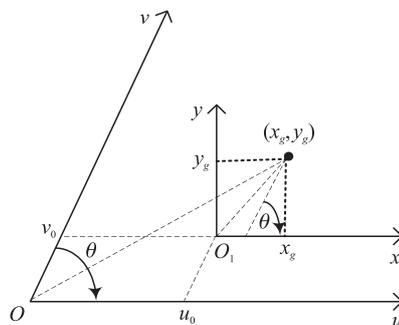


图 4 图像坐标转像素坐标示意图

多目识别系统复杂度较高,而无论用何种单目识别方法得到的结果都是二维的,难以分析物体整体情况,因此本文引入 RFID 标签信息作为先验信息辅助识别物体。视觉识别获得目标框的同时,也将返回物体的长、宽和高等信息,将其两两相乘得到不同面的面积,再将目标框面积与先验信息计算得出的面积进行比较,设定误差在某个区间内为同一个面,此时先验信息中剩余的长(或者宽、高)则为实际的高,得知物体的高后可从目标框的中心点坐标映射出重心坐标,然后以目标框为顶面可以复现出三维物体。以图5为例,加粗黑线框为检测到的目标框,返回目标框的顶点像素坐标后,通过左上角顶点像素坐标 $(x, y, z)$ 计算其真实坐标 $(X, Y, Z)$ ,如式(6)所示,其中 $d_x, d_y, d_z$ 分别是指真实距离 $D_x, D_y, D_z$ 对应的像素距离。通过顶点真实坐标可计算出目标框中心点的真实坐标。由俯视角角度得到了目标框的中心点坐标,则以长方体为例,可映射出的重心坐标为 $(X, Y, \frac{1}{2}Z)$ 。

$$(X, Y, Z) = \left( x \frac{d_x}{D_x}, y \frac{d_y}{D_y}, z \frac{d_z}{D_z} \right) \quad (6)$$

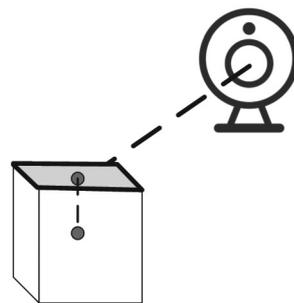


图 5 识别目标框示意图

本文以 Canny 边缘检测算法和 Faster R-CNN 目标检测算法为例来详述如何使用本文方法对现有算法进行改进,对 Canny 和 Faster R-CNN 的改进分别见 2.2 节和 2.3 节。

## 2.2 Canny 的改进算法

当物体检测过程中采用边缘检测算法时,为了尽可能减少噪声对边缘检测结果的影响,采用高斯滤波对图像进行平滑及过滤噪声处理。然后计算图像中不同像素点的梯度强度和方向,图像边缘一般会指向各个方向,于是 Canny 算法使用四种算法来识别图像中的水平、垂直和对角边缘。在传统的 Canny 边缘检测中,伪边的存在会导致物体尺寸的确定出现误差<sup>[14]</sup>,因此基于原始的算法,本文做了如下的改进。

首先对图像进行平滑时,通过每个像素的水平梯度和垂直梯度的高斯核计算梯度的绝对值,并得到边缘方向,分别如式(7)和式(8)所示。

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (7)$$

$$\theta = \arctan \frac{|G_x|}{|G_y|} \quad (8)$$

本文改进的 Canny 边缘检测方法中,为了消除边缘检测引起的虚假响应,采用非最大值抑制,然后采用双阈值检测方法进行标定真实边缘和潜在边缘,实质是结合了 RFID 标签中准确的物体几何信息和像素信息来确定滤波窗口的权重,像素值越接近几何中心的像素,则权重越大,数学表达如式(9)所示。

$$w(u_i, v_j, u_k, v_l) = \exp \left[ \frac{(u_i - u_k)^2 + (v_j - v_l)^2}{2\delta_d^2} - \frac{\|f(u_i, v_j) - f(u_k, v_l)\|^2}{2\delta_g^2} \right] \quad (9)$$

式中,  $(u_i, v_j)$  为当前点的像素坐标,  $(u_k, v_l)$  为几何中心点的真实坐标,获取物体几何信息后由具体的几何中心计算公式得到。  $f(u_i, v_j)$  和  $f(u_k, v_l)$  分别表示当前点和几何中心点的像素值。  $\delta_d$  和  $\delta_g$  分别表示空间距离标准差和灰度距离标准差。由上式可知,若像素值和附近目标点的像素很接近或者差异很大时,难以达到去噪目的,因此需要缩小边缘的搜索范围,本文采取减小搜索范围阈值进行自适应双阈值,表达式如下:

$$\sigma_B^2 = P_1(m_1 - m_c)^2 + P_2(m_1 - m_c)^2 \quad (10)$$

式中,  $\sigma_B^2$  是由大律算法<sup>[24]</sup>计算得到的最大类间方差,  $m_c$  为图像平均像素灰度值,  $m_1$  为边缘像素平均灰度值,  $P_1(m_1 - m_c)^2$  表示检测到像素点是物体边缘的概率,  $P_2(m_1 - m_c)^2$  表示检测到像素点为背景的概率。

最后,通过抑制孤立的弱边缘来完成边缘检测。分级后的强边与弱边相关联,通过保留局部极大值和抑制所有其他的非极大值点可以实现将模糊的边缘转换成锐利的边缘。

## 2.3 Faster R-CNN 的改进算法

为了更全面对视觉识别方法进行改进,来展示此方法的通用性,本算法也提供了对目标检测算法的辅助识别,输入的数据集标定了 RFID 标签中准确的物体几何信息。由于对于分类较多的目标检测任务, ResNet<sup>[25]</sup>模型在精确度和运行速度上都优于 VGG 模型,因此本文采用的基础网络为 ResNet-50,预训练模型选择 ImageNet,且在卷积层中运用 ImageNet 中的批归一化 (Batch Normalization, BN) 方法。训练迭代轮数为 24,初始学习率设置为 0.002 5,批大小为 6,观察到训练集损失发散时降低学习率。

本文为了增强识别精度,改进的方法在原始的 Faster R-CNN 的 RPN 架构上引入了特征金字塔 (Feature Pyramid Network, FPN)<sup>[26]</sup>,在每次的卷积下采样时都生成对应的特征图,然后通过自顶向下的上采样方式将特征图放大,同时语义也越强,这使得网络更具鲁棒性,提高了 Faster R-CNN 算法对多尺度的小目标的检测精度。将前后两种特征图进行横向特征融合,可得图 6 中右边的 FPN 特征图组。

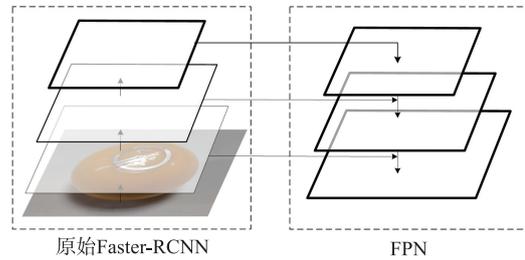


图 6 FPN 结构图

RPN 对特征图输出提议框之后,由 Softmax 模型确定生成的提议框中的候选目标部分,RPN 的损失函数如式(11)所示:

$$L(\{P_i\}, \{b_i\}) = \frac{1}{N_c} \sum_i L_c(P_i, P'_i) + \lambda \frac{1}{N_p} \sum_i P'_i L_p(b_i, b'_i) \quad (11)$$

式中,  $N_c$  表示类别数目,  $N_p$  表示候选区域中目标的数量,  $i$  为候选目标的编号,  $P_i$  表示候选目标是真实目标的概率,  $P'_i$  表示预测正确性标签,获取先验信息计算候选目标的预测框与真实目标框重合面积,当此面积大于等于 80% 时,  $P'_i = 1$ ,当重合面积小于 30% 时,  $P'_i = 0$ 。  $b_i$  表示预测框的边界参数坐标,  $b'_i$  代表  $P'_i = 1$  时的目标标注框边界坐标向量。  $\lambda = 10$ ,是

损失函数在训练总损失函数中占比的平衡参数。

由于 RoI Pooling 会导致区域错误匹配的问题, 因此为了提高多层次目标的检测精度, 本文在选择候选区域之后增加 RoI 对齐<sup>[19]</sup>层, 替换 Faster R-CNN 的 RoI Pooling 层。

### 3 实验与结果分析

#### 3.1 实验环境介绍

本文所提系统基于 Python3.6 和 PaddlePaddle 实现, 算法采用边缘检测 Canny 算子和 Faster R-CNN 目标检测算法。PaddlePaddle 用于训练 Faster R-CNN 模型并进行数据增强以提高模型准确度, 模型使用时需要导入依赖库 paddle。硬件配置为 Intel i7-8700K 3.70GHz CPU 和 16GB 内存, Nvidia GeForce RTX 2080 显卡, Nvidia 依赖为 CUDA10.0、cuDNN7.5.1。射频模块读写器由 INDY R2000 芯片和双 CPU 架构设计组成。视频模块硬件包括海康威视摄像头 DS-2DC2204IW-DE3/W, 该摄像头可变焦, 焦距范围 2.8 mm~12 mm, 其使用的图像传感器能够对图像进行精准的显示, 且具有降噪功能, 在对相邻帧图像进行比较滤波后, 能够对找出的噪声点位置进行增益控制, 以弱化弱信号图像的噪声干扰; 支持水平 350° 无死角监控, 通过网口传输视频流, 结合物体检测可实现对物体坐标的精准定位。

本文基于两种不同的算法进行的对比实验过程分别包含双目识别和单目 RFID 辅助识别两种不同的识别方式。双目识别采取如图 7 所示的通过两个摄像头获取物体三维信息, 其中摄像头 1 获取单个截面目标框, 摄像头 2 可获取物体的高度; 单目 RFID 辅助识别方式的测试演示图如图 8 所示, 包括 RFID 读写器、RFID 天线、带有标签的物体和一个摄像头。图 7 中的摄像头 1 与图 8 中的摄像头 1 位置一致, 即摄像头 1 在双目、单目的情况下位置一致。

为体现本研究的可靠性及效果, 本文以视野范围左上角顶点处为原点建立三维坐标系, 对不同规格的物体进行了不同方位、不同距离的测试, 结果表现出物体越小识别误差越大, 因此最终选择了较小

规格的物体识别结果数据进行分析, 实物如图 9 所示, 包括长方体、立方体、直棱柱、平行六面体、球体和圆锥形状的物体。

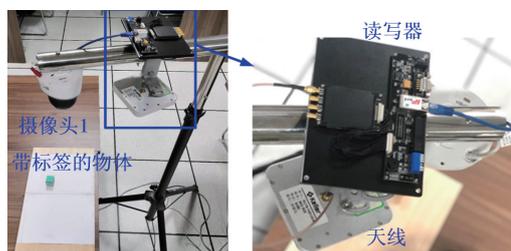


图 8 RFID 辅助单目摄像头识别演示图

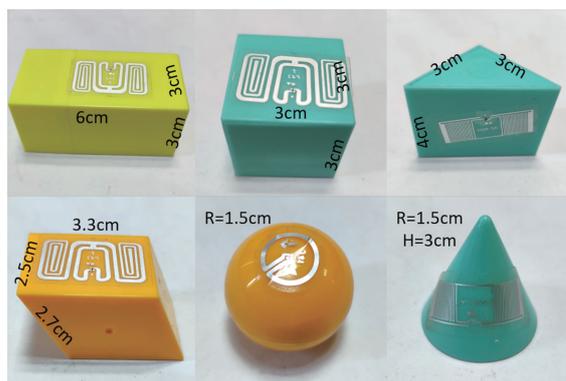


图 9 部分测试物体实物图

#### 3.2 性能评估

双目识别需要靠视觉识别技术获取物体的有效信息。而基于 RFID 技术的单目视频识别辅助系统中, RFID 标签存有物体实际的先验信息, 通过真实的信息辅助摄像头识别物体加快了识别速度、减少了估计误差。

在定位准确度方面, 如图 10 所示为边缘检测实验的中心距离偏差对比图, 图 11 为 Faster R-CNN 目标检测实验的中心距离偏差对比图。相同尺寸物体基于同样的环境因素进行测试, 由于边缘检测比目标检测算法更易受环境因素的影响, 因此定位误差起伏较大。对比不同形状的物体的检测结果, 本文通过 RFID 标签信息辅助单目边缘检测进行三维视觉识别的方法误差均比双目边缘检测的误差要小, 总体平均降低定位误差为 25.60%。通过 RFID 标签



图 7 双目识别演示图

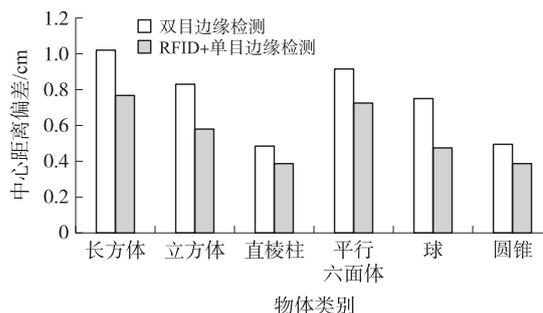


图 10 边缘检测误差对比图

信息辅助单目 Faster R-CNN 进行三维视觉识别的方法误差均比双目 Faster R-CNN 三维识别方式的误差更小,总体平均降低定位误差为 25.70%。

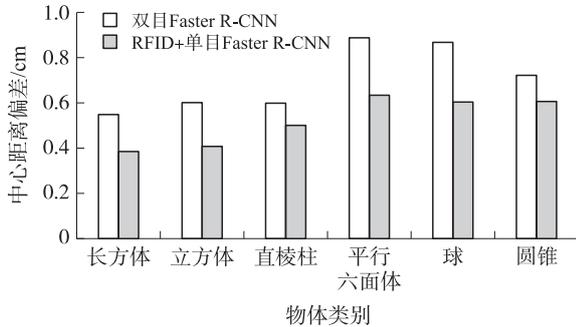


图 11 目标检测误差对比图

在识别准确度方面,边缘检测能标识物体的边缘,但得到的结果中目标框会大于物体实际边缘。目标检测能准确识别出物体标记好的标签,但不论是边缘检测或是目标检测算法都会由于环境因素的影响而使得目标框不够标准,导致返回物体的长、宽和高有一定的误差,如图 12 中实际俯视应该是黄色的框,但是检测返回的结果却是红色的框,这时通过 RFID 读写器读取物体标签能得到准确的物体尺寸信息,使得检测的计算结果更加准确。本实验中,对于物体三维尺寸信息,边缘检测的平均识别误差为 9.02%,目标检测的平均识别误差为 9.25%,通过从数据库中获取读写器上传的真实的物体信息,能去除三维尺寸的识别误差,有助于未来应用于各种机器,提高机器对物体的准确操纵程度。

在算法速度方面,运行多目摄像头进行物体识别,产生的数据量更大,而且需要对多个摄像头画面进行关联分析,算法计算更耗时。单目 RFID 可以减少视角,利用单目识别出的二维图形加上 RFID 标签关联的物体尺寸等信息,可以识别出物体的三维位置,并且射频读取的速度很快,因此从原理上是能够提高物体识别的速度的,而且实验也证实了单目 RFID 确实提升了识别速度,该方法的代价是在物体上需要贴上电子标签,将物体的信息与标签相

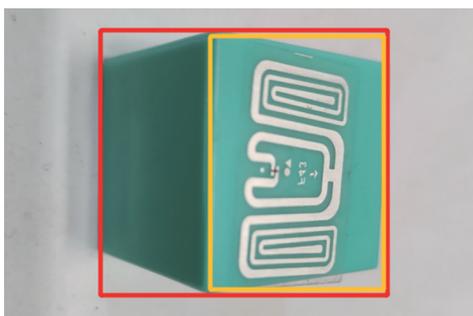


图 12 物体识别目标框示意图

关联并存储在数据库中,这部分的工作需人工操作。如图 13 所示的边缘检测识别速度对比图中可看出在所有的测试中,双目边缘检测所需的时间均高于基于 RFID 改进的单目边缘检测算法,本文在速度上平均提升约 32.82%。

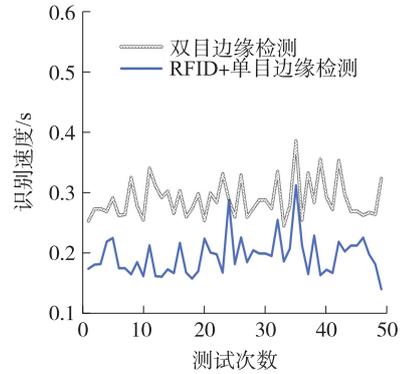


图 13 边缘检测识别速度对比图

如图 14 所示为目标检测识别速度对比图,由于 Faster R-CNN 计算量更大,加以调用了第三方 paddle 依赖库,在准确度提升的同时会牺牲速率,因此基于 RFID 的单目 Faster R-CNN 在速度的提升上更加明显,速度上提升约 48.38%。本文以一个摄像头代替两个摄像头实现三维识别,降低了算法复杂度,提高了识别速度。

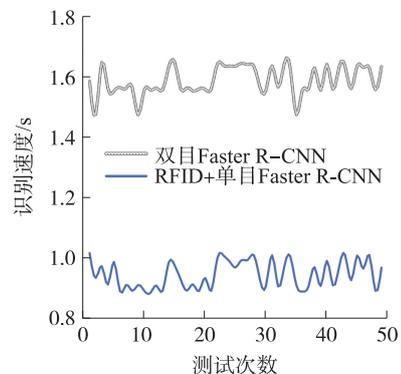


图 14 目标检测识别速度对比图

### 4 总结

本文提出了一种通用的射频技术辅助的视觉识别系统,通过读取 RFID 标签信息辅助单目视觉识别算法识别定位物体,并获取数据库中存储的标签数据实现三维坐标可视化复现,对今后自动分拣、视障辅助、机器辅助操纵等不同场景的智能化发展具有极大的研究价值。实验数据证明,针对边缘检测算法的辅助识别,本文方法可提高 25.60%的定位准确度,降低 9.02%的三维尺寸识别误差,速度提升了 32.82%;针对 Faster R-CNN 目标检测算法的辅助识别,本文方法可提高 25.70%的定位准确度、降低

9.25%的三维尺寸识别误差以及提升48.38%的识别速度。在本文系统的辅助之下,视觉识别效果能得到整体的提升,后续工作将继续优化识别的准确度,使得本系统能够复现不规则物体的三维图像。

### 参考文献:

- [1] 曹蜜,胡凌燕,熊彭文,等. 基于PROSAC算法与ORB-SLAM2的RGB-D相机室内视觉定位研究[J]. 传感技术学报,2019,32(11):1706-1712.
- [2] Meshram V V, Patil K, Meshram V A, et al. An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People[J]. IEEE Transactions on Human-Machine Systems, 2019, 49(5):449-460.
- [3] Hughes B, Burghardt T. Automated Visual Fin Identification of Individual Great White Sharks[J]. International Journal of Computer Vision, 2017, 122(3):542-557.
- [4] Lin Y, Zhou H, Chen M, et al. Automatic Sorting System for Industrial Robot with 3D Visual Perception and Natural Language Interaction[J]. Measurement and Control, 2019, 52(1-2):100-115.
- [5] Liu L, Ouyang W, Wang X, et al. Deep Learning for Generic Object Detection: A Survey[J]. International Journal of Computer Vision, 2020, 128(2):261-318.
- [6] Wang Z, An Q, Cao C, et al. Design and Implementation of Automatic Sorting Control System for Melon and Fruit Products [C]//International Conference on Application of Intelligent Systems in Multi-Modal Information Analytics. Springer, Cham, 2019:1455-1459.
- [7] Wiles O, Zisserman A. Learning to Predict 3d Surfaces of Sculptures from Single and Multiple Views[J]. International Journal of Computer Vision, 2019, 127(11):1780-1800.
- [8] Liu X, Jonschkowski R, Angelova A, et al. Keypose: Multi-View 3d Labeling and Keypoint Estimation for Transparent Objects [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:11602-11610.
- [9] Yuan J, Abdul-Rashid H, Li B, et al. A Comparison of Methods for 3D Scene Shape Retrieval[J]. Computer Vision and Image Understanding, 2020, 201:103070.
- [10] 李建宇,张金艺,蔡春艳,等. 大空间双模多频段射频识别复合定位[J]. 传感技术学报,2015,28(11):1664-1669.
- [11] Xu H, Yuan C, Li P, et al. Design and Implementation of Action Recognition System Based on RFID Sensor[C]//2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE, 2017:3021-3025.
- [12] Martinez-Martin E, Ferrer E, Vasilev I, et al. The UJI Aerial Librarian Robot: A Quadcopter for Visual Library Inventory and Book Localisation[J]. Sensors, 2021, 21(4):1079.
- [13] Shi X, Cai H, Wang M, et al. TagAttention: Mobile Object Tracing With Zero Appearance Knowledge by Vision-RFID Fusion [J]. IEEE/ACM Transactions on Networking, 2021, 29(2):890-903.
- [14] Canny J. A Computational Approach to Edge Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986(6):679-698.
- [15] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards Real-Time Object Detection with Region Proposal Networks [C]//Advances in neural information processing systems. 2015:91-99.
- [16] Kittler J. On the Accuracy of the Sobel Edge Detector[J]. Image and Vision Computing, 1983, 1(1):37-42.
- [17] LeCun Y, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognition [J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [18] Zhao K, He T, Wu S, et al. Application Research of Image Recognition Technology Based on CNN in Image Location of Environmental Monitoring UAV [J]. EURASIP Journal on Image and Video Processing, 2018, 2018(1):1-11.
- [19] He K, Gkioxari G, Dollár P, et al. Mask r-cnn [C]//Proceedings of the IEEE International Conference On Computer Vision. 2017:2961-2969.
- [20] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1904-1916.
- [21] Girshick R. Fast r-cnn [C]//Proceedings of the IEEE International Conference on Computer Vision. 2015:1440-1448.
- [22] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [C]//International Conference on Learning Representations. 2015:1-14.
- [23] Gao Z, Zhu M, Yu J. A Novel Camera Calibration Pattern Robust to Incomplete Pattern Projection[J]. IEEE Sensors Journal, 2021, 21(8):10051-10060.
- [24] Otsu N. A Threshold Selection Method from Gray-Level Histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1):62-66.
- [25] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [26] Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:2117-2125.



张智(1983—),男,通讯作者,博士,浙江工业大学之江学院信息工程学院物联网实验室主任,杭州立宸科技有限公司总经理,主要研究领域为物联网、大数据和人工智能技术, zhangzhi@zjc.zjut.edu.cn;



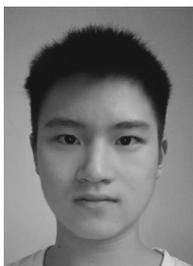
刘子瑜(1996—),女,浙江工业大学计算机学院硕士研究生,研究方向为传感数据分析、目标识别和目标追踪, lzyu@zjut.edu.cn;



邱灵龙(2000—),男,浙江工业大学计算机学院在读本科生,研究方向为大数据分析和目标检测,qll7521@gmail.com;



王植焯(2000—),男,浙江工业大学信息工程学院在读本科生,研究方向为物联网、传感信号处理、目标识别,wxuanye@yeah.net;



董旭(2000—),男,浙江工业大学之江学院信息工程学院在读本科生,研究方向为自动化,嵌入式开发和机器人学,1728411347@qq.com。