DOI:10.19344/j.cnki.issn1671-5276.2019.04.039

基于长短期记忆网络的挖掘机器人视觉跟踪系统设计

丁盼,庞晓平,陈进

(重庆大学 机械传动国家重点实验室,重庆 400044)

摘　要:为通过视觉系统向挖掘机器人输入挖掘目标以及解决因运动容易导致挖掘目标丢失的问题,提出一种基于长短期记忆网络的视觉跟踪方法,无需限定特定目标类别,直接从摄像头图像中框选目标,经过 Dlib 提取特征后,通过训练优化长短期记忆网络来跟踪挖掘目标的位置。针对亮度变化、障碍物遮挡和背景干扰三类常见问题进行长短期记忆网络建模和训练,经图片集模拟测试和实际跟踪测试均显示该网络模型能够有效纠正干扰并稳定输出目标位置。网络模型将神经网络和传统机电控制方法结合,提高视觉跟踪精度,为将神经网络应用于挖掘机智能化进行了初步探索。

关键词:长短期记忆网络;视觉跟踪;时间序列;挖掘机器人

中图分类号:TP242; TP311.5　　**文献标志码**:A　　**文章编号**:1671-5276(2019)04-0145-04

Visual Tracking System Design of Excavation Robot Based on Long Short-Term Memory Network

DING Pan, PANG Xiaoping, CHEN Jin

(State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China)

Abstract: To input the excavation target to the excavating robot through the visual system and solve the easy loss of the excavation target in movement, a visual tracking method based on the long and short term memory network is proposed. The target tracking system is directly selected from the camera image without limiting the specific target category, After the feature is extracted by Dlib, the position of the mining target is tracked by training to optimize the short-term memory network. Long-term and short-term memory network modeling and training are conducted for three common problems of brightness change, obstacle occlusion and background interference. The simulation training of the picture sets and actual tracking experiment shows that the network model can be used to effectively correct the interference and output the target position stably. In the network model, the neural network is combined with the traditional electromechanical control method to improve the accuracy of visual tracking. This paper makes a preliminary exploration that the neural network is applied to the intellectualization for the excavating robot.

Keywords: long and short-term memory network; visual tracking system; time sequence; excavating robot

0　引言

挖掘机是在工程施工中普遍使用的工程机械,如何在现有挖掘机基础上进行智能化改造,实现挖掘机机器人化是行业未来发展的必然趋势。王福斌[1]等人采用视觉光流场估计法实现挖掘机器人在有道路导引线情况下进行障碍物识别和自主移动。Hyun-Seok Yoo[2]等人使用二维激光扫描仪进行环境建模,实现了挖掘机机器人的挖掘、卸料和回转的工作任务规划。于华琛[3]等人利用三帧差法以及高斯混合建模法实现铲斗目标检测,实现挖掘机器人对自身挖掘姿态的识别。张建忠[4]等人探索出通过视觉引导对挖掘机械臂进行寻迹控制,实现对人类通过视觉控制机械臂的控制策略模拟。挖掘机器人要进行环境感知[5]和工作任务规划,视觉能够提供最为丰富的信息[6]。

挖掘机器人在向挖掘区域移动时,出于避障需要会出现绕行等情况,需要视觉跟踪系统保持对挖掘区域跟踪,同时伺服电机转角可为挖掘机器人提供控制参数。近年来,随着深度学习的迅速发展,很多学者将深度神经网络特别是卷积神经网络(convolutional neural networks, CNN)运用于视觉跟踪,因此诞生了很多性能优异的视觉跟踪算法。Ross Girshick[7]等人提出一种基于 CNN 提取目标特征的递归卷积神经网络(recurrent convolutional neural networks, RCNN)视觉跟踪算法。该算法依赖于 ImageNet ILSVC 2012 图像识别库,其包含 1 000 类物体的 1 千万张图片。该算法使用 CNN 提取候选区域图像特征,再输入每类物体的 SVM 分类器进行判别,最后使用回归器修正目标候选框的位置。Joseph Redmon 和 Ali Farhadi[8]等人于 2015 年提出 YOLO 算法(you only look once, YOLO),该算法对 RCNN 进行了大幅改进。该算法和 RCNN 算法

基金项目:国家自然科学基金项目(51475056)

作者简介:丁盼(1991—),男,四川大竹人,硕士研究生,研究方向为挖掘机器人智能化控制。

一样,都需要使用分类图片库训练分类器,跟踪目标种类必须属于训练分类器的图片库。基于 CNN 的视觉跟踪算法只能识别分类图片库中的物体种类,缺乏对跟踪目标的普遍适应。

人们在观察运动物体时,会不断将当前眼睛接收到的图片和记忆中的图片进行比对,提取出目标的运动轨迹,这需要从环境中提取具有同一目标特征区域,不断与记忆进行比对。长短期记忆网络(long short term memory networks,LSTM)具有控制遗忘的结构设计,非常适合处理时序任务[9]。视觉跟踪任务从时间角度来看,视频流是由一个有序排列、单向流动的图片序列组成的。根据这一原理,采用机器学习特征提取工具 Dlib[10] 对图片进行特征提取并返回预测位置,LSTM 神经网络根据输入和自身记忆[11] 进行修正,输出最终目标位置。本文对挖掘机器人自动工作视觉跟踪技术进行初步探索,介绍视觉跟踪系统技术验证的硬件平台、软件功能;分析 LSTM 神经网络结构并对网络结构和参数进行调整优化;通过测试集和跟踪系统的实际测试对整个视觉跟踪算法进行评价,结果证明 LSTM 神经网络可以运用于挖掘机器人视觉跟踪任务。

1　开发硬件平台

硬件包括开发 PC 机 1 台,树莓派 3 model B 型嵌入式开发板 1 个,罗技 C270 网络摄像头 1 个,Tower Pro SG90 9 克舵机 2 个,5 V2 A 电源适配器 1 个,2.4 G 无线路由器 1 个,USB 键鼠 1 套(图 1)。

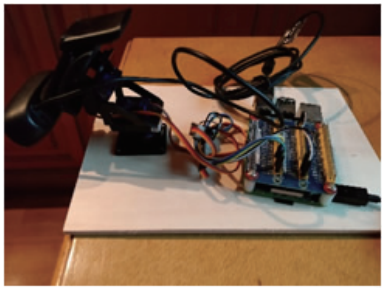

图 1　视觉跟踪平台

视频采集流程:树莓派平台安装 mjpg-streamer 软件库,摄像头采集的视频流被 mjpg-streamer 通过 WIFI 网络传送给 PC 主机,PC 主机可以 IP 地址访问(树莓派 ip 地址+8080 端口)

视觉跟踪流程:Dlib 运行于 PC 主机上,可以通过鼠标对跟踪目标进行框选,Dlib 会在后续视频帧中进行特征匹配,并返回相似度最高的预测目标框坐标值(目标框四边坐标),坐标值传入 LSTM 神经网络,LSTM 神经网络根据历史传入坐标值数据,对当前帧的目标预测位置进行修正,修正值作为结果输出,并用于 Dlib 跟踪目标初始化。

跟踪动作流程:舵机控制端软件工作于 PC 主机上,软件会以 6 帧为单位接受 LSTM 输出的坐标框值,分别计算前后 3 帧预测框中心点坐标均值,判别后 3 帧中心点运动方向。软件通过 WIFI 网络向树莓派平台发送运动指令,使目标保留在摄像头视野范围中心位置附近,完成跟踪任务(图 2)。

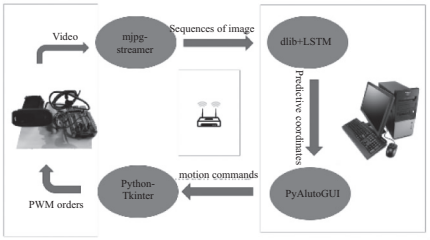

图 2　视觉跟踪系统工作示意图

2　LSTM 神经网络

2.1　LSTM 算法简介

LSTM 神经网络是在循环神经网络(recurrent neural networks, RNN)基础上发展而来的。RNN 神经网络的输入不仅有当前时刻的输入,还包含上一时刻 RNN 细胞隐藏层的状态。从原理上讲,RNN 的输出会受到以前所有输入的影响,但是在实际使用中发现,随着时间序列的变长,在优化 RNN 网络时会出现梯度消失或者梯度爆炸[12],网络不再收敛。为了解决这一问题,减少历史状态对 RNN 细胞状态的影响,在 RNN 细胞的输入层、输出层、隐藏层添加了 3 个门结构,这就成为了 LSTM 细胞。LSTM 细胞按信息传递时间轴连接成一层 LSTM 神经网络如图 3 所示[13]。

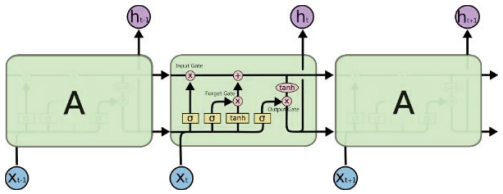

图 3　LSTM 细胞结构

LSTM 细胞上有 3 个门控制器:输入门(Input Gate)、遗忘门(Forget Gate)和输出门(Output Gate)。相对于 RNN 细胞,LSTM 增加了两个状态:隐藏层状态 h 与细胞状态 c。隐藏层状态 h 反映 LSTM 细胞输入输出变化,细胞状态 c 反映 LSTM 细胞记忆变化。这两层状态是 LSTM 长短期记忆遗忘功能的基础。

1) 遗忘门决定 t 时刻从细胞状态中丢失信息的比例。取决于 $t-1$ 时刻隐藏层状态 h_{t-1} 和细胞状态 c_{t-1}。

$$f_t=\sigma(\boldsymbol{W}_f[h_{t-1},c_{t-1},x_t]+\boldsymbol{b}_f) \tag{1}$$

2) 输入门 i_t 影响当前时刻 t 的细胞状态更新,输入门输入数据 x_t 与上一时刻 $t-1$ 隐藏层 h_{t-1}、细胞状态 c_{t-1} 相乘,再输入激活函数 $\acute{O}$,激活结果由一个 tanh 层表示,再加上 $t-1$ 时刻细胞状态乘以遗忘门,作为 LSTM 细胞当前细胞状态 c_t。

$$i_t=\sigma(\boldsymbol{W}_i[h_{t-1},c_{t-1},x_t]+\boldsymbol{b}_i) \tag{2}$$

$$c_t=i_t\cdot\tanh(\boldsymbol{W}_c[h_{t-1},x_t]+\boldsymbol{b}_c)+f_t\cdot c_{t-1} \tag{3}$$

3) 输出门 O_t 影响着 t 时刻 LSTM 细胞的输出信息比例,并影响着隐藏层状态变化。

$$O_t=\sigma(\boldsymbol{W}_o[h_{t-1},c_t,x_t]+\boldsymbol{b}_o) \tag{4}$$

$$h_t=O_t\cdot\tanh(c_t) \tag{5}$$

式(1)-式(5)中:$\boldsymbol{W}_f$、$\boldsymbol{W}_i$、$\boldsymbol{W}_o$分别为遗忘门、输入门、输出门的权值矩阵;$\boldsymbol{b}_f$、$\boldsymbol{b}_i$和$\boldsymbol{b}_o$均为初始偏置矩阵;σ与tanh均为神经网络激活函数。

2.2　用于视觉跟踪的LSTM网络模型建立

本文使用TensorFlow 1.2版本搭建用于挖掘机器人视觉跟踪系统的LSTM模型。

LSTM神经网络的输入张量形状为[batch_size, time_steps, input_size],batch_size为每次训练时输入值的个数,time_steps为每个batch需要多少次输入,input_size为每次输入的维度数即包含多少个值。因为LSTM网络t时刻输出值会在t+1时刻重新输入,所以输入值与输出值张量形状相同。

从LSTM网络输出张量中取出最后一次time_step的输出值,其张量形状为[1,1,*cell_size*],*cell_size*表示每层LSTM网络所包含的细胞数。需要在LSTM网络后添加一个线性输出层,将LSTM网络[1,1,*cell_size*]输出结果平滑输出为[1,1,4]的预测坐标值。根据公式(6)计算预测坐标值B_{pred}与真实坐标值B_{target}之间均方差,TensorFlow通过AdamOptimizer优化器按照已确定的学习率使用随机梯度下降法不断迭代,减少均方差到局部最小值,LSTM神经网络不断学习到目标的运动规律。

$$L_{MSE}=\frac{1}{n}\sum_{i=1}^{n}\parallel B_{\text{pred}}-B_{\text{target}}\parallel_2^2 \tag{6}$$

挖掘机工作在粉尘大、具有较多障碍物的工作场所,这使得视觉跟踪目标会因为光线明暗变化(iiiumination variation,IV)、目标遮挡(occlusion,OCC)、环境物体相似干扰(background clutters, BC)等问题。从学界常用于视觉跟踪器性能评估的TB-100视频序列数据库[14]选择具有代表性数据集(表1)进行LSTM网络训练及参数调优,强化其对跟踪视觉特征变化目标的跟踪能力。

表1　训练数据集描述

数据集名称	训练样本数	测试样本数	数据集特征
Woman	149	448	IV、OCC
*Human*3	445	1 253	OCC、BC
*Car*1	255	765	IV、BC
*Skating*1	100	300	IV、BC
*David*3	63	189	OCC、BC

依次使用上述数据集进行训练,在每次训练后,恢复网络模型,并使用测试集进行验证,统计网络模型在测试集中的损失函数值、预测坐标值等数据。根据预测坐标与真实坐标计算预测目标区域与真实目标区域的重叠率即*IOU*值,学界普遍认为*IOU*>50%为跟踪成功。

2.3　用于视觉跟踪的LSTM网络模型参数调优

对LSTM网络参数进行调整,可以使得跟踪更加精确和稳定。其中参数包括输入序列长度、学习率和LSTM网络记忆率这3个显著影响LSTM网络性能的参数。评价视觉跟踪算法性能主要指标为*IOU*值(intersection-over-union),即交叉重叠率,指预测目标框与真实目标框交集与并集之商。IOU值>0.5视为跟踪成功,反之失败。*IOU*最小值表明网络性能最差值,应使得*IOU*最小值尽可能大,以减少目标丢失的发生。*IOU*均方差表明网络工作稳定程度,值越小表明网络性能越稳定,控制系统舵机摇摆的可能性更低。因此网络调优应对上述3个参数综合考虑,使得跟踪网络平稳工作。

使用控制变量法进行网络参数调优,结果如下。

表2为不同输入序列长度时LSTM网络跟踪性能。由表2可知,随着输入序列变长,*IOU*均值稳定上升,均方差值也随之上升,跟踪对象的不稳定性随之增加。当序列长度为3时,*IOU*最小值>0.5,在整个测试阶段均跟踪成功。

表2　输入序列长度评估

输入序列长度	*IOU*均值	*IOU*均方差	*IOU*最小值
2	0.777 94	0.007 47	0.426 30
3	0.791 36	0.007 60	0.520 66
4	0.800 79	0.007 93	0.459 91

表3为不同学习率时LSTM网络跟踪性能。由表3可知,*IOU*均值最高点出现在学习率为0.03时,*IOU*均方差最小值和*IOU*最小值的最大值出现在学习率为0.06时。

表3　学习率评估

学习率	*IOU*均值	*IOU*均方差	*IOU*最小值
0.01	0.794 96	0.015 11	0.148 34
0.03	0.817 53	0.010 17	0.333 81
0.06	0.787 56	0.008 45	0.459 29
0.09	0.720 78	0.011 39	0.320 65

表4为不同记忆率时LSTM网络跟踪性能。由表4可知,*IOU*均值最大值、*IOU*最小值的最大值和*IOU*均方差最小值均出现在记忆率为0.95时。

表4　记忆率评估

记忆率	*IOU*均值	*IOU*均方差	*IOU*最小值
0.8	0.591 41	0.025 65	0.072 91
0.85	0.643 44	0.019 90	0.047 21
0.9	0.716 79	0.014 12	0.237 28
0.95	0.779 15	0.008 40	0.483 71

经过上述反复实验,确定LSTM网络在输入序列长度为3、学习率为0.06,记忆率为0.95时跟踪效果最佳。

3　视觉跟踪系统测试

3.1　LSTM网络跟踪性能测试

在测试阶段,LSTM网络在t时刻的输出会作为最终结果,用于Dlib在t+1时刻的目标初始化。测试阶段开始后,第1、2两个页面跟踪使用Dlib返回坐标,以减少LSTM网络初始化使用零张量填充的影响。使用图片测试集测试时,需关闭视觉跟踪系统控制输出功能。

图4a)在90~93帧显示,跟踪系统纠正目标周围行人

a）测试集为 TB-100 Human3

b）测试集为TB-100 Singer2

c）测试集合为 TB-100 Suv

图 4　视觉跟踪系统在应对相似物体干扰、光线明暗变化和物体遮挡的性能表现

的背景干扰;图 4b)在 322~325 帧显示,跟踪系统有效纠正瞬间亮度变化引起跟踪框变化;图 4c)在 511~514 帧,跟踪系统能够修正行道树对 SUV 的视野遮挡影响。经过训练集优化训练后的基于 LSTM 网络的视觉跟踪系统能有效应对红色框为标记真实位置,蓝色框为 LSTM 网络输出预测位置(因本刊系黑白印刷,如有疑问请咨询作者)。

3.2　视觉跟踪系统测试

打开系统控制输出功能,测试其实际工作表现。

a）亮度变化

b）背景相似干扰

c）背景遮挡

图 5　视觉跟踪系统实际跟踪性能展示,蓝色框为跟踪位置

在图 5a)中,跟踪系统能够有效跟踪处于逆光环境的跟踪目标,并适应目标亮度变化;在图 5b)中,跟踪系统能够有效区别和目标相似的物体,并纠正其对跟踪的干扰;在图 5c)中,跟踪目标被障碍物遮挡,跟踪系统稳定工作,使得跟踪目标保持在视野较中心位置。

由以上数据集和实地实验可知,基于 LSTM 网络视觉跟踪系统能够有效应对挖掘机器人在工作环境中上述工作问题,提高跟踪精度,为机器人识别环境、理解环境奠定基础。

4　结语

为了解决挖掘机器人视觉跟踪丢失目标的问题,提出了一种基于 LSTM 网络的视觉跟踪算法,并组建了硬件系统进行验证。实验表明:

1) LSTM 网络在解决时序依赖性问题有着突出优势,本文将其应用与视觉跟踪领域,显示出在时间和空间方面出色的回归预测能力,可以有效利用视频序列的记忆信息,从中找出运动规律,做出准确的预测。

2) 本视觉跟踪系统使用通用视觉特征提取库 Dlib,使得算法不再像以往基于 CNN 网络的视觉跟踪算法局限于训练时的物体种类,实现了对环境、目标的广泛适应。网络不需要针对特定目标进行训练,针对机器人的工况进行训练即可。

3) 本视觉跟踪系统可与其他机器人控制系统进行整合,系统可将伺服舵机转角参数用于移动控制。相比于王福斌等的研究,使得挖掘机器人不再依赖于引导线等特殊的视觉标识,降低了挖掘机器人对工作环境的硬件要求。

本文对挖掘机智能化进行了技术尝试,将神经网络与传统工控算法结合,用于解决挖掘机器人通过视觉感知环境的一些突出问题,为神经网络应用于挖掘机智能化奠定了理论基础。

参考文献:

[1] 王福斌,刘杰,焦春旺,等. 基于多传感信息的挖掘机器人导航及避障技术[J]. 辽宁工程技术大学学报(自然科学版),2011,30(3):430-433.

[2] Yoo H S, Kim Y S. Development of a 3D local terrain modeling system of intelligent excavation robot[J]. KSCE Journal of Civil Engineering, 2017, 21(3): 565-578.

[3] 于华琛,袁祖强. 基于机器视觉的铲斗目标检测[J]. 机械制造与自动化,2016,45(4):165-167.

[4] 张建忠,颜景平. 基于视觉的未标定机械臂路径规划[J]. 机械制造与自动化,2006,35(6):25-28.

[5] 杨晓宁,叶锦华. 基于 KINECT 和全局视觉的机器人避碰研究[J]. 机械制造与自动化,2016,45(5):156-160.

[6] 徐德,谭民. 机器人视觉测量与控制 [M]. 北京:国防工业出版社, 2008.

[7] He, Kaiming, Georgia Gkioxari, Piotr Dollár, et al. Girshick. "Mask R-CNN." 2017 IEEE International Conference on Computer Vision (ICCV) Venice[C]. New York: IEEE, 2017: 2980-2988.

[8] Redmon, Joseph, Ali Farhadi. "YOLO9000: Better, Faster, Stronger." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Venice[C]. New York: IEEE, 2017: 6517-6525.

[9] Tax N, Verenich I, La Rosa M, et al. Predictive business process monitoring with LSTM neural networks. In Pohl K, Dubois E, editors, Advanced Information Systems Engineering: 29th International Conference[C]. Berlin Springer: CAiSE Essen Germany, 2017: 477-492.

[10] King D E. Dlib-ml: A machine learning toolkit [J]. Journal of Machine Learning Research, 2009, 10: 1755-1763.

[11] 王国栋, 韩斌, 孙文赟. 基于 LSTM 的舰船运动姿态短期预测 [J]. 舰船科学技术, 2017, 13: 69-72.

[12] Ericson G. Biomedical entity recognition using Team Data Science Process (TDSP) Template [M]. [S.I.]: [s.n.], 2017.

[13] Song E, Soong F K, Kang H-G. Effective Spectral and Excitation Modeling Techniques for LSTM-RNN-Based Speech Synthesis Systems [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(11):52-61.

[14] Lim J. Visual Tracker Benchmark [DB/OL]. (2014-09-16) [2017-12-19]. http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html.

收稿日期:2018-03-15