文章编号：2095-2163(2020)07-0017-06 中图分类号：TP391 文献标志码：A

一种应用于动态场景的优化语义 SLAM

丁佳惠，刘 翔，奚峥皓
（上海工程技术大学 电子电气工程学院，上海 201610）

摘 要：本文针对动态场景，在 ORB-SLAM 的基础上，提出一种新的语义 SLAM 地图构建方法，提高智能移动机器人对环境感知和场景认知的能力。以 RGB-D 为输入信息，对 RGB 信息和深度信息分别作 ORB 特征匹配和尺度判断，利用 RANSAC 算法进行位姿估计判断关键帧。通过基于金字塔池化改进的 MASK-RCNN 神经网络对关键帧进行语义分割。在分割好的关键帧上，通过查找表法结合语义信息剔除动态目标。处理好的关键帧用于构建语义地图，同时进行局部集束调整，最后再回环检测。原语义分割网络精确率为 81.2%，改进的网络精确率达到 90.5%。

关键词：SLAM；MASK-RCNN；查找表；语义分割；动态场景

An optimized semantic SLAM applied to dynamic scenes

DING Jiahui, LIU Xiang, XI Zhenghao
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201610, China)

【Abstract】 Aiming at dynamic scenes, based on ORB-SLAM, a new semantic SLAM map construction method is proposed to improve the ability of intelligent mobile robots to sense the environment and recognize scenes. With RGB-D as input information, ORB feature matching and scale judgment are performed on RGB information and depth information, respectively, and key frames are determined by pose estimation using RANSAC algorithm. Then MASK-RCNN neural network based on improved pyramid pooling is used to perform semantic segmentation on key frames. On the segmented key frames, the dynamic target is eliminated by the lookup table method combined with the semantic information. The processed key frames are used to construct a semantic map and perform local clustering adjustments at the same time, and then perform loop detection. The accuracy of the original semantic segmentation network is 81.2%, and the accuracy of the improved network reaches 90.5%.

【Key words】 SLAM; Deep Learning; Lookup Table; Semantic Segmentation; Dynamic Scene

0 引 言

近几年来，随着新一代技术（大数据、物联网和云计算等）的迅速发展，人类生产和生活的智能化程度也不断提高，尤其在智能机器人的实现过程中，智能导航系统是智能化改造率最高的环节之一，更是连接实物流与信息流之间的关卡。智能移动机器人是指能够依靠特定装置，在规定范围内行驶，并实现特定功能（如运输，巡检等）的自动机器人。要实现自动引导首先机器人需要具备 SLAM 技术[1]。

SLAM（Simultaneous Localization and Mapping）是指在环境未知的情况下，通过传感器不断地获得当前环境的信息，并且结合自身的位姿信息建立环境地图。而视觉 SLAM 就是以图像为环境感知信息源的 SLAM 系统，是当前最为前沿的技术之一[2]。要想让机器人实现真正的智能自主导航，那么为机器人的 SLAM 地图构建系统添加语义信息就必不可少。语义 SLAM 在对智能机器人进行定位以及感知环境信息的基础上，通过使用语义分割网络获取环境语义信息，构建语义地图[3]。

Civera 等提出了单目语义 SLAM 方法和 SLAM + +方法，但这两种方法建立的语义地图只是简单包含了数据库中的物体；Kundu 等将卷积神经网络应用到 SLAM 中，融合构建了具有实时性的室内语义地图。但是这些语义地图在动态场景中的应用不尽人意，由于动态物体的移动，导致构建的地图出现残影、重影、全局一致性差等问题。本文在语义 SLAM 研究的基础上，提出能够删除动态物且语义信息更加完善的 SLAM。

1 基于 RGB-D 的 ORB-SLAM 结合改进 MASK-RCNN 的语义 SLAM

1.1 语义 SLAM 系统

在 ORB-SLAM 的基础上提出的语义 SLAM 系统，如图 1 所示。输入为 RGB-D 深度图像，同时提取 RGB 图像中的 ORB 特征和深度图像的尺度信

基金项目：上海市科委地方能力建设项目（15590501300）。

作者简介：丁佳惠（1995-），女，硕士研究生，主要研究方向：计算机视觉；刘 翔（1972-），男，博士，副教授，主要研究方向：人工智能。

通讯作者：刘 翔 Email：xiangliu@outlook.com

收稿日期：2020-04-02

息,未初始化时,先利用最初两帧图像进行 ORB 特征点对匹配和尺度判断,得到初始点对[4]。之后的每一帧图像与上一帧进行特征点的匹配,若存在足够的匹配点对,则利用 RANSAC 算法对当前帧与上帧的匹配点对进行结合深度信息的位姿估计。若是匹配点对过少,则先通过 DBoW2 库建立基于 ORB 特征的关键帧词袋向量[5],在已建立的词袋数据库中搜索匹配,以此得到最佳位姿估计。再通过 covisibility graph 获得局部地图对当前帧的位姿进行优化,得到当前帧的位姿信息。达到特定条件,则将当前帧定为关键帧。

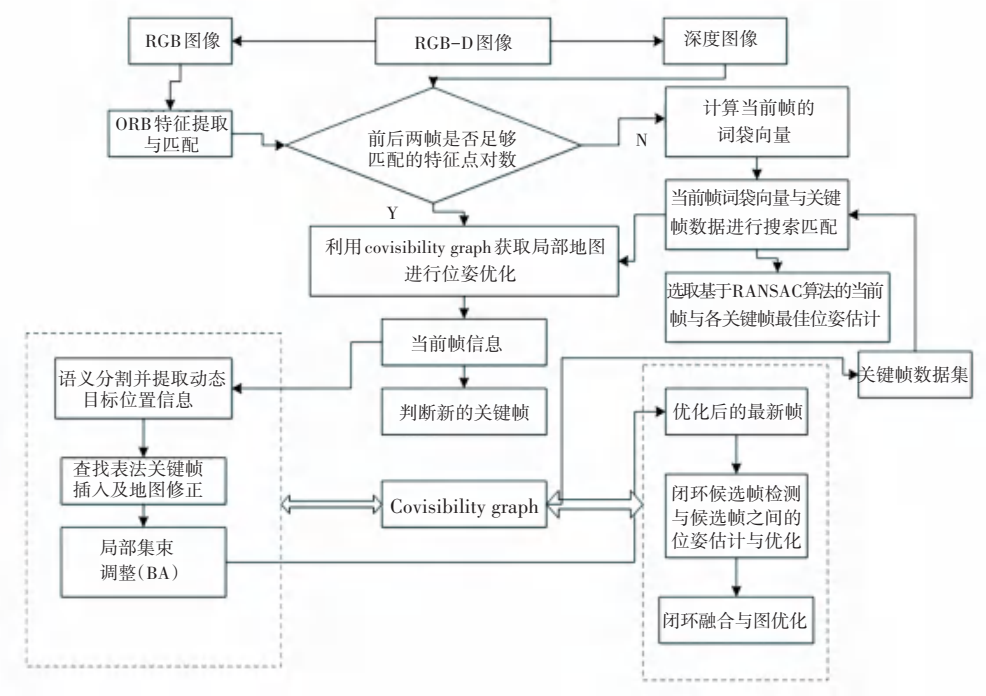


图 1 语义 SLAM 流程

Fig. 1 Semantic SLAM process

每出现一张新的关键帧时(设第一张为关键帧),将此关键帧作为改进神经网络的输入进行语义分割,得到含语义信息的图像和动态目标的位置信息。基于新添加的语义关键帧,利用查找表法,剔除动态目标,维护和拓展新的局部地图点,进行局部集束调整之后进行回环检测[6]。

1.2 回环检测

在全局词袋库中存入当前关键帧的词袋(Words Bag)向量,以此提高后续帧的匹配速度,并检测回环是否存在。若存在回环,则通过位姿图(Pose Graph)优化对全体的关键帧位姿进行优化,以此来减少累计漂移的误差。完成位姿图优化后,利用全局光束平差法得到最优地图和运动(关键帧位姿)[7]。

1.3 基于金字塔池化模型的改进 Mask-RCNN

语义是指机器对周围环境的理解与识别。例如,识别场景中的行人,物体等。随着计算机技术的提升以及深度学习在计算机视觉领域的快速发展,深度学习已经能够结合 SLAM 对场景进行语义分割构建语义地图[8]。

图 2 为 Mask-RCNN 的流程图,在已预训练好的 ResNet 中输入图片得到对应特征图,并对特征图中的每一点设预定 ROI,得到多个候选 ROI。将 ROI 输入 RPN 网络进行 BB 回归和二值分类,并将某些候选 ROI 过滤掉。对剩下的 ROI 进行 ROIAlign 操作(即先对原图和特征图的像素进行对应,再将特征图和固定的特征进行对应)。最后,对这些 ROI 进行分类、BB 回归和 MASK 生成(在每一个 ROI 里面进行 FCN 操作)[9]。

为了提高网络的精度,提出一种改进的 Mask-RCNN 网络,减少误分割从而提高准确率。在 Mask-RCNN 网络的基础特征提取器中,插入金字塔池化模型,即分别使用 1×1,2×2,3×3 以及 6×6 的卷积核处理 ResNet 输出特征图,提取不同尺度下的特征信息,再将经过上采样处理的特征图变换到同一个尺寸下,最终得到的特征图由于包含了局部和全局的上下文信息,就能减少误分割的情况,提高准确率。改进的 Mask-RCNN 网络流程图如图 3 所示。

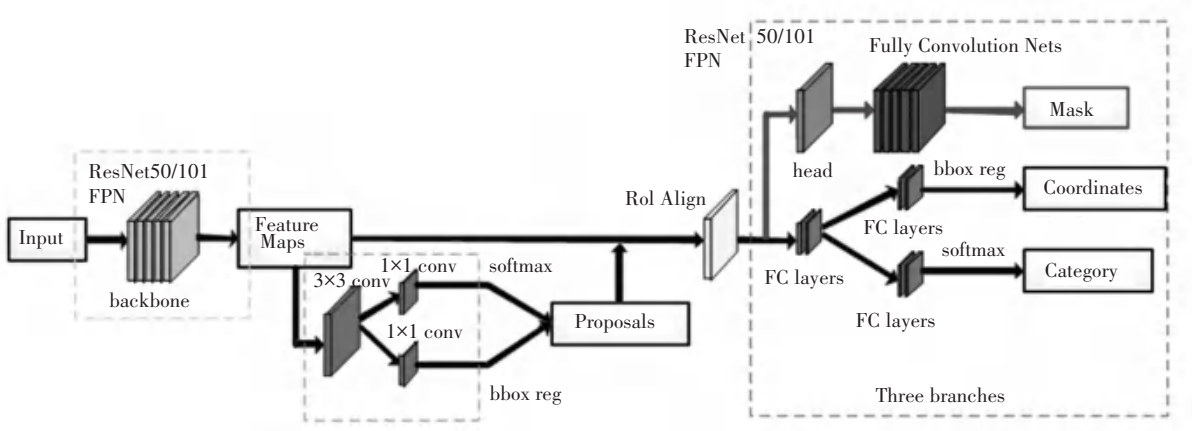


图 2　Mask-RCNN 网络流程图

Fig. 2　Mask-RCNN network flowchart

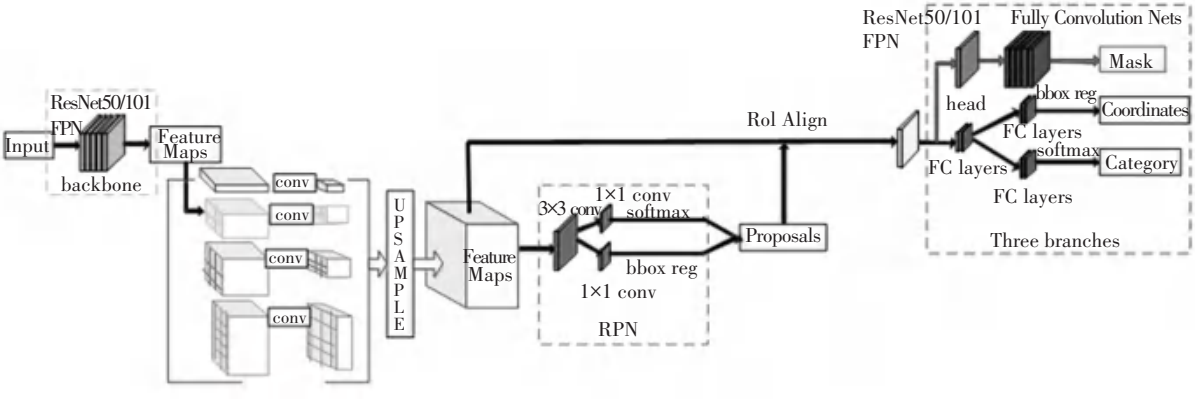


图 3　改进的 Mask-RCNN 网络流程图

Fig. 3　Flow chart of the improved Mask-RCNN network

1.4　结合查找表的 SLAM 建图

由于日常生活工作环境难免有人员的走动，一般情况下 SLAM 会自动将人或者移动的物体加入到地图的构建中，使得视觉 SLAM 适应性差且地图中信息有冗余，出现残影，降低可信度，故采用基于查找表的方法剔除动态目标。

首先，将每帧分成 16 份，如图 4 所示。再进行 ORB 特征点匹配，得到八领域的图像移动方向；其次，将最多统计数量的方向作为该帧的运动方向；最后，构建出查找表。对图像进行语义分割得到动态目标的坐标后，通过表 1 除去对应坐标区域的行人，最终建立出剔除动态目标的语义 SLAM 地图。

图 4　图像分割与 8 领域方向示意图

Fig. 4　Image segmentation and 8-field direction

表 1　查找表示意图

Tab. 1　Schematic diagram of lookup table

方向编号	构图区域编号	图像移动方向
0	0,1,2,3,4,8,12	左上
1	0,1,3,7,11,15	上
2	2,7,11	右上
3	3,7,8,11,14	左上
4	0,4,7,9,12	左
5	2,5,7,11	右
6	1,3,8,19,12	左下
7	1,5,6,8,12,14,15	下

根据表 1 进行地图构建的示意图如图 5 所示。如果将每幅关键帧分割为 n 份，f(cell) 表示每份区域的信息，则该图像可表示为公式(1)：

$$f(image)=\sum_{cell=0}^{n-1} f(cell), \tag{1}$$

地图构建时，每一方向需更新的信息为公式(2)：

$$f(direction)=\sum_{cell\in D} f(cell), \tag{2}$$

其中，D 表示建图时该方向所需更新的区域集合。最后构建的地图为公式(3)：

$$f(map) = \sum f(direction). \quad (3)$$

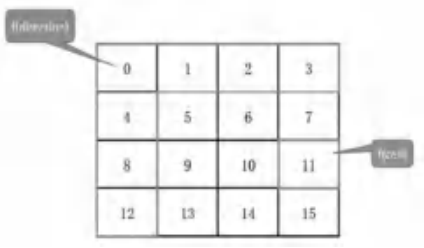


图 5 根据查找表构建地图示意图

Fig. 5 Schematic diagram of building a map based on a lookup table

2 实验结果分析

2.1 关键帧语义分割

对于深度网络的训练采用的是 NYU DepthDatasetsV1 数据集，为增加对动态目标的识别，在原数据集的基础上增加了一些人、车等移动物体数据。原网络精确率为 81.2%，最终改进的网络精确率为 90.5%，比原来的 Mask-RCNN 网络提高了 9.3%，如图 6 所示。

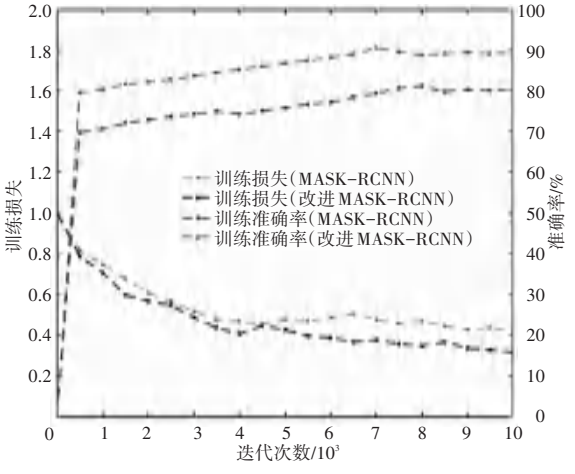


图 6 训练准确率

Fig. 6 Training accuracy

分别选取 NYU DepthDatasets V2 数据集和 ValidationSet 数据集为实验对象，该数据集场景复杂，与生活环境相似，故以此构建语义 SLAM 地图。这两种数据集分别由 780 张、370 张图像构成，动态目标分别为人和水壶，且多处具有回环检测，对于生活场景具有一定的挑战性和代表性。对图 7 中的关键帧进行 Mask-RCNN 语义分割得到结果如图 8 所示，发现存在误分和漏分的情况，改进 Mask-RCNN 语义分割如图 9 所示，误分和漏分现象大大减少。(图 7、8、9 包括两个图，应该为(a)(b))

(a) NYU DepthDatasetsV2 数据集
(a) NYU V2 depth dataset

(b) ValidationSet 数据集
(b) ValidationSet dataset

图 7 关键帧

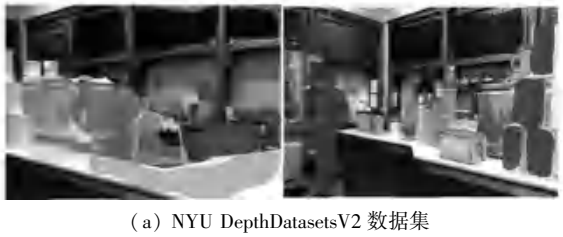
Fig. 7 key frame

(a) NYU DepthDatasetsV2 数据集
(a) NYU V2 depth dataset

(b) ValidationSet 数据集
(b) ValidationSet dataset

图 8 关键帧语义分割

Fig. 8 key frame semantic segmentation



(a) NYU DepthDatasetsV2 数据集
(a) NYU V2 depth dataset

(b) ValidationSet 数据集
(b) ValidationSet dataset

图 9 改进网络关键帧语义分割

Fig. 9 Improved network key frame

利用上述深度网络进行语义分割后，得到两个数据集关键帧中的动态目标的位置坐标分别见表 2、表 3，得到这些坐标后结合查找表法建立剔除动态目标的语义地图。

表 2　NYU 行人位置在图像中的坐标

Tab. 2　Coordinates of NYU pedestrian position in the image

图片	左上角 X 坐标	左上角 Y 坐标	右下角 X 坐标	右下角 Y 坐标
433.png	89.318 62	175.467 74	176.265 89	481.886 94
434.png	82.332 09	172.576 57	167.816 35	476.965 47
435.png	76.544 30	163.326 59	154.421 98	462.632 10
436.png	63.295 63	168.716 23	139.776 35	459.876 45
437.png	59.624 53	159.954 04	133.862 97	451.315 67

表 3　ValidationSet 移动水壶位置在图像中的坐标

Tab. 3　ValidationSet coordinates of the moving kettle position in the image

图片	左上角 X 坐标	左上角 Y 坐标	右下角 X 坐标	右下角 Y 坐标
12.png	81.752 62	162.776 42	152.172 52	458.289 73
13.png	85.475 68	161.563 82	155.784 31	459.776 45
14.png	86.957 12	160.145 77	156.143 25	463.321 74
15.png	89.376 63	161.336 72	160.675 38	471.233 26

2.2　结合查找表的语义地图构建

未进行语义分割以及动态目标的 ORB-SLAM 地图如图 10 所示，其中出现的动态物体会对出现重影，缺乏全局一致性的影响。本文提出的基于改进 MASK-RCNN 的语义 SLAMM 能够构建具有语义信息的地图，增强机器人对周边环境的理解，但是由于移动物体的存在，使得该方法语义分割效率低且出现移动物体的重影，如图 11 所示。本文提出的应用于动态环境的语义 SLAM 方法不仅能够在室内动态环境下语义 SLAM 地图，增加智能移动机器人对动态场景的理解，还能够消除动态目标对建图的影响，除去地图中动态目标带来的重影且保证地图的全局一致性，如图 12 所示。

（a）NYU DepthDatasetsV2 数据集

（a）NYU V2 depth dataset

（b）ValidationSet 数据集

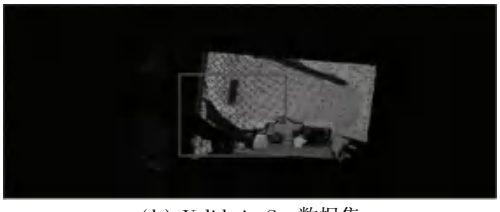
（b）ValidationSet dataset

图 10　传统 RGBD-SLAM 地图

Fig. 10　Traditional RGBD-SLAM map

（a）NYU DepthDatasetsV2 数据集

（a）NYU V2 depth dataset



（b）ValidationSet 数据集

（b）ValidationSet dataset

图 11　基于改进 RGBD-SLAM 的语义地图

Fig. 11　Semantic map based on improved RGBD-SLAM

（a）NYU DepthDatasetsV2 数据集

（a）NYU V2 depth dataset

（b）ValidationSet 数据集

（b）ValidationSet dataset

图 12　融合查找表的改进 RGBD-SLAM 语义地图

Fig. 12　Improved RGBD-SLAM semantic map fused with lookup table

3　结束语

应用于智能生活的语义 SLAM，在深度学习的基础上，提出了一种语义分割的 ORB-SLAM 方法，用于智能移动产品能更好的自主识别未知环境、自主定位以及减小移动物品对产品的影响。该方法以 RGB-D 信息为输入，通过改进 MASK-RCNN 网络对动态场景进行语义分割，将分割后的语义信息通过查找表法剔除动态目标，最后融入 ORB-SLAM 系统中，构建语义 SLAN 场景地图。通过对采集的数据集进行实验，结果表明，所提出方法能够在动态场景中构建剔除动态目标具有全局一致性的语义地图，能够解决智能移动产品自主环境认知定位以及

移动物体产生影响的问题。

语义信息和 SLAM 是互帮互助的,在 SLAM 地图建立中,利用语义信息能够加强机器人对环境的理解,帮助机器人提供更好的服务,使得机器人更加智能化;而 SLAM 建图过程获取的图片位姿信息也能提高语义分割的精度。在未来工作中,将对语义 SLAM 进行更深层次的探讨,利用语义 SLAM 实现智能移动机器人自主导航。

参考文献

[1] 李俊. 基于 SLAM 导航的多目视觉 AGV 系统设计[J]. 包装工程,2018,19(39):181-189.

[2] 揭云飞,王峰. 视觉 SLAM 系统分析[J]. 电脑知识与技术,2018,14(19):221-223.

[3] 王召东,郭晨. 一种动态场景下语义分割优化的 ORB_SALM2[J]. 大连海事大学学报 2018,4(44):121-126.

[4] CIVERA J, GÁLVEZ-LÓPEZ D, RIAZUELO L, et al. Towards semantic SLAM using a monocular camera [C]// IEEE International Conference on Intelligent Robots and Systems. New York: IEEE,2011:1277-1284.

[5] SALASMORENO R F, NEWCOMBE R A, STRASDAT H, et al. SLAM + + :Simultaneous localisation and mapping at the level of objects[C]//IEEE Conference on Computer Vision and Pattern Recognition.New York: IEEE ComputerSociety.2013:1352-1359.

[6] KUNDU A, LI Y, DELLAERT F, et al. Joint semantic segmentation and 3D reconstruction from monocular video [C]// European Conference on Computer Vision. Berlin: Springer International Publishing,2014: 703-718.

[7] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2:An open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics,2017,33(5):1255-1262.

[8] KLEIN G, MURRAY D. Paralleltracking and mapping on a camera phone[C]//Proceedings of IEEE International Symposium on Mixed and Augmented Reality,2009.

[9] GALVEZ-LÓPEZ D, TARDOS J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics,2012,28(5): 1188-1197.

(上接第 16 页)

表 5 算法平均执行时间统计

Tab. 5 Average execution time consumption ms

组号	亮度条件		
	正常亮度	较亮	较暗
1	28.56	26.54	27.72
2	27.32	28.74	27.54
3	27.59	27.69	27.45
4	27.44	28.45	27.12
5	26.77	27.32	26.84
6	27.34	27.48	26.89
7	27.64	28.03	28.37
8	26.89	27.83	26.41
9	27.86	27.91	26.89
10	26.77	26.75	28.43

3 结束语

为克服射击训练系统中光照对于靶面识别的影响,本文提出一种基于 HSV 色彩空间以及 OTSU 算法的靶面识别算法,通过对图像 HSV 通道的分离和结合 OTSU 自适应阈值分割方法识别出射击靶面。通过实验验证,本算法可以有效的在不同亮度条件下识别出射击靶面,识别率高,在三种光照条件下的执行时间稳定在 27 ms 左右,满足射击训练实时性的需求。

参考文献

[1] 罗杰,张之明. 基于图像处理技术自动报靶系统综述[J]. 激光杂志,2016,37(7):1-6.

[2] 苑玮琦,李梦祺. 基于视觉检测的胸环靶自动报靶系统研究[J]. 计算机技术与发展,2019,29(2):147-151.

[3] 刘焱,李敏勇. 靶面目标图像识别算法[J]. 微计算机信息,2006(36):313-314,236.

[4] 李致衡,陈亮,张博程,等. 基于最大熵阈值分割的 SAR 图像溢油检测[J]. 信号处理,2019,35(6):1111-1117.

[5] 郜峰利,陶敏,李雪妍,等. 基于深度学习的 CT 影像脑卒中精准分割[J]. 吉林大学学报(工学版),2020,50(2):678-684.

[6] Xiaoqiang Ji, Yang Li, Jiezhang Cheng, et al. Cell Image Segmentation Based on an Improved Watershed Algorithm[C]// 2015 8th International Congress on Image and Signal Processing. IEEE, 2015:433-437.

[7] Rafika HARRABI, Ezzedine BEN BRAIEK. Color Image Segmentation Using a Modified Fuzzy C-Means Technique and different color spaces: Application in the Breast Cancer Cells Images [C]//1st International Conference on Advanced Technologies for Signal and Image Processing. IEEE, 2014:236-352.

[8] Refik Sametl, Sahin Emrah Amrahoyl and Ali Hikmet Ziroglul, Fuzzy Rule-Based Image Segmentation Technique for Rock Thin Section Images [C]// Image Processing Theory, Tools and Applications. IEEE, 2012.

[9] Chaza Chahine, Racha El Berbari, Corinne Lagorre, et al. Evidence theory for image segmentation using information from stochastic Watershed and Hessian filtering [C]// international Conference on Systems. IEEE, 2018:141-144.

[10] Ye C, Mi H. The technology of image processing used in automatic target - scoring system [C]//International Joint Conference on Computational Sciences & Optimization. IEEE, 2011: 349-352.

[11] Hao Feng, Zhiguo Jiang. Image Segmentation With Hierarchical Topic Assignment [C]// International Conference on Image Processing. IEEE, 2011:2125-2128.