

移动相机下基于三维背景估计的运动目标检测*

郭会文^{1,2}, 吴新宇^{1,2}, 苏士娟¹, 傅睿卿¹

(1. 中国科学院深圳先进技术研究院 深圳 518055; 2. 中国科学院大学深圳先进技术学院 深圳 518055)

摘要:室内环境中的运动目标检测是计算机视觉领域的研究热点,而移动相机造成的动态背景是运动目标检测的难点。本文提出一种基于同步定位与地图创建(ORB-SLAM)三维背景估计的运动目标检测算法,首先使用移动相机遍历整个室内环境,采用ORB-SLAM技术建立当前全局环境的三维背景特征点云模型;然后基于局部视频建立局部三维特征点云,根据定位信息将当前局部三维特征点云与环境三维背景特征点云进行嵌入,基于环境背景信息,采用三维均值漂移(3DMS)算法,对局部三维特征点云进行前景特征点提取;运用深度卷积神经网络,对提取的前景特征点所在候选区域进行运动目标确认。通过多个室内场景的实际实验进行验证,结果表明本文方法具有较高的运动目标检测准确率和召回率,提出的运动目标检测算法充分利用了三维背景信息,采用深度卷积神经网络进行确认,有效地改善了检测的准确性和鲁棒性。

关键词:室内运动目标检测;三维背景估计;三维均值漂移;深度卷积神经网络

中图分类号: TP391.4 TH79 文献标识码: A 国家标准学科分类代码: 520.20

3D Background estimation for moving object detection using a single moving camera

Guo Huiwen^{1,2}, Wu Xinyu^{1,2}, Su Shijuan¹, Fu Ruiqing¹

(1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China;

2. Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract: Moving object detection in indoor environment is a research hotspot in the field of computer vision. The dynamic background caused by moving the camera is a challenge in moving target detection. In this paper, a moving object detection algorithm based on ORB-SLAM (oriented FAST and rotated BRIEF-Simultaneous Localization and Mapping) is proposed. Firstly, the entire indoor environment is traversed using a moving camera, ORB-SLAM is used to establish the 3D feature cloud model of the global background. Then, based on the environmental information, local 3D feature point cloud is built. By embedding the local 3D feature points into the global 3D background feature cloud, 3D Mean-shift is applied to extracting the foreground points from the local 3D feature points. Finally, deep convolution neural network is utilized to confirm the moving target of the candidate region where the extracted foreground feature points are located. The experimental results on multi indoor scenes show that the proposed method has high detection accuracy and recall rate. The proposed moving object detection algorithm makes full use of the background information, and the depth convolution neural network is used to confirm candidate regions, which effectively improves the detection accuracy and robustness.

Keywords: 3D background estimation; moving object detection; 3D Mean-Shift (3DMS); deep convolutional neural networks

0 引言

主要应用领域包括视频监控、视频编码、目标识别与跟踪、行为分析等^[2-4]。此外,该技术经常被用于其他计算机视觉相关工作的预处理当中,如视频抠图^[5]等。传统的移动目标检测技术假设相机是固定静止不动

运动目标检测是计算机视觉中的一项重要技术^[1],

收稿日期:2017-04 Received Date: 2017-04

* 基金项目:国家自然科学基金面上项目(61473277)、机器人与智能制造国家地方联合工程实验室(2015581)、深圳市智能机器人与智能制造工程实验室(20141722)、深圳市技术攻关项目(JSGG20150930154605341)资助

的,近年来研究人员提出多种背景差分技术处理动态背景、恶劣天气情况以及夜间拍摄视频等问题。然而,随着移动设备(如智能手机、移动机器人、手持摄像机、无人机载体等)的快速发展,移动相机视频越来越普及,针对移动相机下的运动目标检测技术越来越重要。相对于静止相机,移动相机下的运动目标检测更为困难。目前,针对运动相机下的运动目标检测问题主要有3类算法:基于运动信息补偿、基于运动聚类和基于图像的算法^[6-9]。

为了使用运动信息,一类典型的做法是计算移动相机姿态变化,对相机的运动进行补偿,抵消相机运动产生的运动信息,从而提取有效的运动信息。文献[10]利用移动机器人相关的传感器估计机器人的姿态变化,从而间接推断相机的姿态变化。在特征点选择方面,文献[11]采用著名的尺度不变特征变换(scale-invariant feature transform, SIFT)特征点进行匹配;为了提高特征点提取速度,文献[12]采用能够快速计算的局部不变特征(oriented fast and rotated brief, ORB)进行匹配。在匹配方面,为了进一步消除特征点畸变和噪声干扰,文献[13]采用递归的思想,排除噪声匹配点对的影响,通过多次递归计算,得到最终的姿态矩阵。文献[14]将图像分成多个区块,并分别计算每个区块的姿态变化矩阵。文献[15]针对移动小目标的检测,在运动补偿的基础上,使用分类器对候选前景区域进行筛选,获得准确度更高的运动目标检测结果。这类算法虽然在一定程度上缓解了动态背景造成的影响,但是仍然存在姿态矩阵估计不准确,计算复杂度高等问题。

另一类算法使用基于运动聚类的算法来避免复杂的计算。这类算法基于前提假设:移动相机的运动相对于运动目标的运动来说是微弱的。基于该假设,该类算法将具有相似运动属性的点进行聚类,从中选取运动强度大的区域,认为是运动目标前景区域。已有文献[16-18]从光流强度图中进行聚类。其中,文献[16]使用迭代聚类算法获取鲁棒的运动聚类信息,而文献[18]使用动态阈值算法从光流聚类结果中筛选出前景运动目标;文献[19]通过提取相邻图像的边缘,并使用边缘进行匹配,获取稀疏运动信息,并将这些运动信息进行聚类,提取前景运动目标轮廓,从而对运动目标进行检测;文献[20]使用鲁棒角点(Kanade-Lucas Tomasi, KLT)特征匹配点对提供的运动信息进行聚类,快速发现移动的目标;文献[21]提取相似运动特征和色度空间特征对每个像素点进行描述,通过聚类得到运动目标。该类算法可对移动目标进行快速检测,然而,由于前提假设的缺陷,无法对一些运动速度缓慢的移动目标进行高精度检测。

为了避免有效运动信息提取困难的问题,一些学者直接使用外观信息进行运动目标检测,即将问题退化为

基于图像的目标检测问题。该类算法主要关注人体目标,采用滑窗的方式对每一帧图像进行扫描检测是否存在运动人目标^[22-26]。除了传统的采用滑窗方式的深度神经网络框架如残差网络(residual networks, ResNet)^[18]等,为了提高检测速度,一些无需滑窗处理、端到端的深度模型如快速区域卷积神经网络(Faster region convolutional neural network, Faster-RCNN)^[23]、实时目标检测系统(you only look once, YOLO)^[24]、单次多框检测器(single shot multiBox detector, SSD)^[25]也被用于运动目标检测。此外,一些改进也被采用。如文献[26]使用级联卷积神经网络进行行人目标检测;文献[27]提取多种显著性特征,从而加快运动目标的检测;文献[28]从信号处理的角度,提出一种简单但有效的时空滤波算法,也能快速筛选出候选区域。尽管这类算法随着分类器的性能不断提高而越来越有效,但由于真实世界中的环境复杂多变,虚警和误检依旧难以下降。

针对上述问题,本文借鉴运动目标检测中背景估计算法^[29-31]的成功,考虑到当前移动相机环境,提出基于ORB-SLAM(oriented FAST and rotated BRIEF-simultaneous localization and mapping)^[32]三维背景估计的运动目标检测算法。算法的核心思想是尽管相机移动带来视频中背景的变化,但从全局角度来看,背景并未发生改变,通过大量预先获取的视频建立不变的背景模型,并将当前构建的局部环境模型进行背景减除,即可将当前的前景快速提取出来。首先使用移动相机,在当前室内环境中采集大量视频建立全局环境三维背景模型;然后,基于短视频构建局部环境三维模型,并利用ORB-SLAM的定位功能,将局部环境三维模型与全局环境三维背景模型进行匹配嵌入;接着,采用三维均值漂移技术(three dimensional Mean-Shift, 3DMS),将局部环境三维模型中的背景减除,快速的检测到候选运动目标;最后,使用预训练的深度卷积神经网络对候选运动目标进行确认,得到最终的运动目标检测结果。

1 算法描述

本文算法详细的流程如图1所示。本算法采用背景减除的思想,将视频映射到三维特征点云中,利用全局环境背景三维特征点云,将背景剔除,从而提取前景区域,再使用预训练的深度卷积神经网络进行候选区域确认,获取运动目标检测结果。算法首先简要介绍ORB-SLAM技术,描述基于视频帧的全局环境背景三维点云模型建立;在将局部环境三维点云与全局环境三维背景点云匹配的基础上,介绍了基于3DMS预处理的背景减除和前景提取;最后,详细介绍了采用的深度卷积神经网络的结构和训练过程,利用深度卷积神经

网络对提取的候选前景目标进行确认,得到运动目标检测结果。

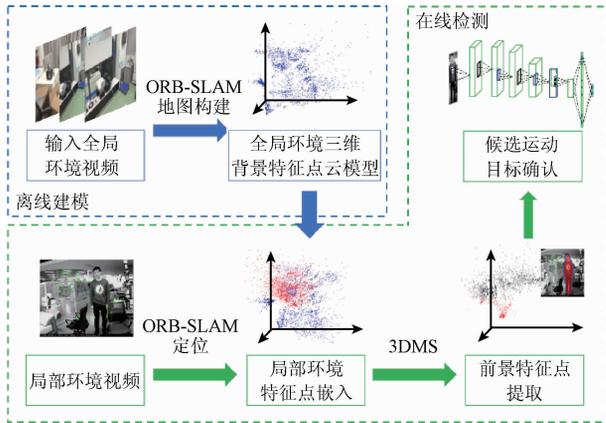


图 1 本文算法流程

Fig. 1 The flowchart of proposed algorithm

1.1 基于 ORB-SLAM 的背景估计

ORB-SLAM 是一种基于 ORB 特征的三维定位与地图构建算法,该算法基于并行跟踪与地图创建 (parallel tracking and mapping, PTAM) 架构,其目标是使用特征点,构建当前环境的三维特征点云模型,并进行全局定位。整个算法分为 3 个并行部分:跟踪、地图构建和闭环检测,算法示意图如图 2 所示。

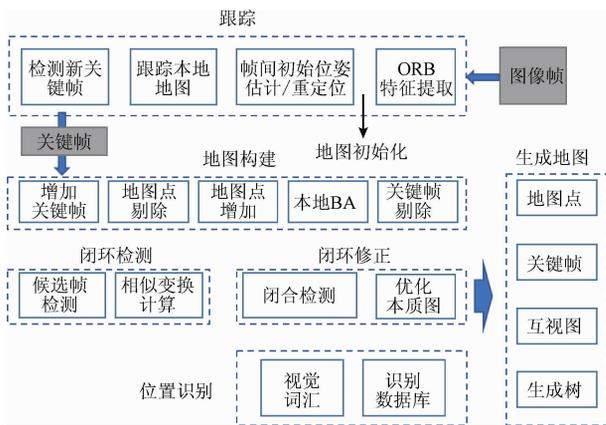


图 2 单目视觉 ORB-SLAM 算法

Fig. 2 Monocular ORB-SLAM algorithm

在跟踪部分开始之前,首先通过标定算法对相机的畸变参数进行标定,并对每一帧图像进行畸变校正。在跟踪阶段,算法采用具有方向信息的多尺度超快特征 (features from accelerated segment test, FAST),用 32 字节 256 位比特数的简短特征 (binary robust independent elementary features, BRIEF) 作为描述符,即一个特征点

用一个 256 位的 0 或 1 表示,提取图像中的 ORB 特征点。对于前后帧图像的特征点,使用词袋技术进行预分类,在此基础上,快速进行特征点匹配。基于匹配特征点对,计算单应性矩阵来估位姿。通过起始多帧完成位姿初始化后,将插入关键帧。

在地图构建环节,针对新插入的关键帧,首先找到对应匹配点,并执行点的剔除保证只保留高质量的特征点,去除关键帧中冗余的特征点。并将这些特征点更新到地图中。为了优化更新到地图中特征点的位置,将在当前帧和局部地图之间寻找尽可能多的特征点对应关系,来优化当前帧的位姿。同时,算法将维护一个运动模型,用来快速估计位姿,在结束更新后,将对运动模型进行更新。

根据上述描述,ORB-SLAM 最终将得到当前环境的三维背景模型和当前相机在环境中的位姿,全局环境三维背景模型采用具有三维位置的特征点云进行描述。由于特征点的数量多少依赖于当前环境纹理丰富程度,因此,很难保证全局环境背景模型中三维特征点云的稠密性。同时,由于采用了计算速度快的 FAST 特征,难以保证不同图像中同一位置均可检测到特征点,此外,对同一位置的特征点描述也可能不同。

1.2 基于 3DMS 的背景减除与前景提取

在全局环境背景三维特征点云模型建立的基础上,利用移动相机获取的局部视频,采用 ORB-SLAM 技术,可建立局部环境三维特征点云模型,并获得该模型相对于全局环境的位置,典型的示例如图 3 所示。

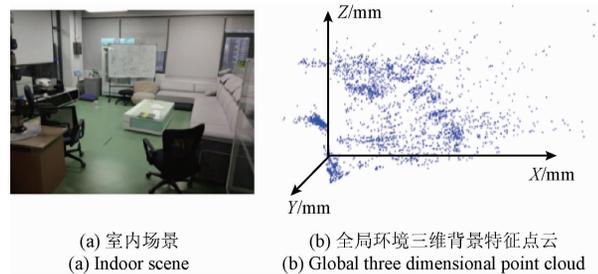


图 3 全局环境三维背景特征点云模型建立

Fig. 3 Establishment of 3D cloud model for global background

利用该位姿,可以将局部环境三维特征点云嵌入到全局环境背景三维特征点云中:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_1 \\ a_{21} & a_{22} & a_{23} & t_2 \\ a_{31} & a_{32} & a_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix} \quad (1)$$

式中: X_L 、 Y_L 、 Z_L 表示特征点在局部环境中的三维位置,

X_G, Y_G, Z_G 表示特征点在全局环境中的三维坐标。 a_{11}, \dots, a_{33} 为旋转矩阵的元素, t_1, t_2, t_3 为平移矩阵的元素。一个典型示例如图4所示。

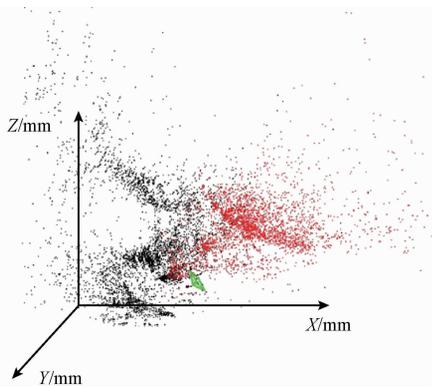


图4 局部环境三维特征点云嵌入全局环境三维特征点云模型(黑色点属于全局环境,红色点属于局部环境,绿色四边形表示当前相机朝向)

Fig. 4 The embedding of local 3D feature point cloud into global environment 3D feature point cloud model (The black points belong to global, the red points belong to local, the green quadrilateral represents the direction of the camera)

由于 ORB-SLAM 采用 ORB 特征,存在两个问题:1) 同一位置的特征描述不一致;2) 不同图像帧在同一位置不一定存在特征点。一方面,无法直接基于特征点描述子进行匹配的方法进行特征点匹配,从而进行背景减除,提取前景;另一方面,无法直接使用特征点三维位置来确认局部环境三维特征点云中的特征点是否属于背景。本文提出基于三维均值漂移算法对特征点位置信息进行聚类,根据类内特征点的分布,来判断局部环境三维特征点云是否属于前景目标。令 $P_G = \{p_G^i\}, i = 1, \dots, N$ 为全局环境背景三维特征点云集合,其中 N 为全局环境背景特征点的个数,且 $p_G^i = [X_G^i, Y_G^i, Z_G^i]$ 。令 $P_L = \{p_L^i\}, i = 1, \dots, M$ 为局部环境三维特征点云集合,其中 M 为局部环境特征点的个数,可知 $M < N$, 且 $p_L^i = [X_L^i, Y_L^i, Z_L^i]$ 。使用式(1)将 P_L 转化为全局坐标 P'_L 。基于 3DMS 的背景减除和前景提取算法的具体步骤如下。

1) 对每个局部环境三维特征点 p_L^i , 计算其邻域的特征点 $N_{p_L^i} = \{p_k : d < T_r\}$, 其中 d 为两点间的欧氏距离, $d = \sqrt{(X_G^i - X_G^k)^2 + (Y_G^i - Y_G^k)^2 + (Z_G^i - Z_G^k)^2}$, T_r 为距离阈值;

2) 更新该点 p_L^i 的坐标为邻域点坐标的均值: $\bar{p}_L^i = \frac{1}{K} \sum_{p_k \in N_{p_L^i}} p_k$, 其中 K 为邻域点个数;

3) 对所有点,重复步骤 1) 和 2), 直至所有点的坐标

变化均小于终止阈值 T_Δ ;

4) 对更新后的每个局部环境三维特征点 \bar{p}_L^i , 其邻域特征点集合为 $N_{\bar{p}_L^i}$, 邻域点个数为 K 。当且仅当邻域点中属于全局环境背景特征点的个数 $K_1 > \frac{1}{3}K$ 时,该特征点视为背景特征点。否则该点为前景特征点。典型的 3DMS 迭代过程如图 5 所示。

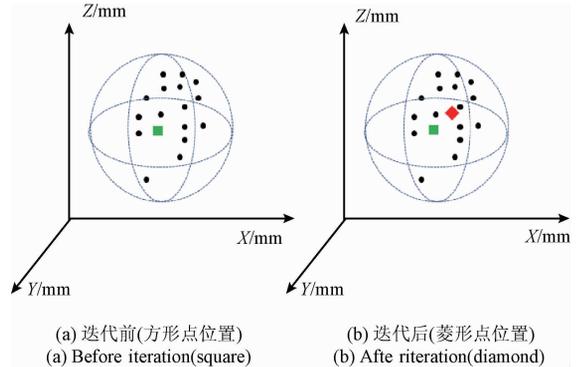


图5 3DMS 算法示意图

Fig. 5 Schematic diagram of 3D Mean-Shift

1.3 基于深度卷积神经网络的运动目标检测

在上述步骤后,基于前景特征点可从图像帧中获取到前景候选区域。由于噪声的影响和感兴趣运动目标选择的需要(运动的人),算法采用深度卷积神经网络进行候选区域的最终确认。由于算法在前述步骤已获取到较为可靠的候选区域,因此,采用较为简单的深度神经网络结构。该网络由 3 个卷积层、3 个最大池化层、1 个全连接层和 1 个 softmax 层组成。网络的输入为归一化为 80×40 大小的图像。网络输出二维结果确定候选区域是否为感兴趣运动目标。具体的网络结构和详细的网络配置如图 6 所示。

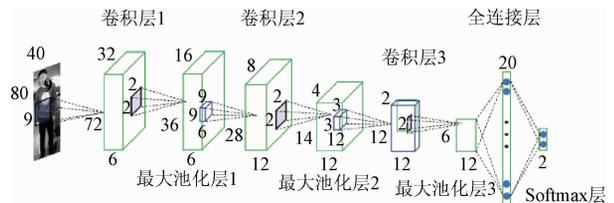


图6 深度卷积神经网络结构

Fig. 6 Structure of deep convolutional neural network

为了使网络保持较高检测准确率,首先使用通用大型数据库数据对网络进行了预训练。由于目前室内环境大多关注于人体运动目标,因此手工采集大量运动行人数据,对网络进行精训练。为了有效应对移动

相机带来的观测视角、视觉尺度等变化,本文从大量真实环境中采集 14 232 张人体运动目标正样本,在使用方向梯度直方图 (histogram of oriented gradient, HOG) 与支持向量机 (support vector machine, SVM) 组合算法的辅助下,手工协同将数据标注。为了平衡训练数据集,从多种背景中提取到 12 000 张负样本,包括柱子、凳子等。

在精调阶段,随着迭代次数的增加逐步降低学习率。在 1 000 次训练后,学习率从 0.05 降到 0.01,在 10 000 次训练后降到 0.001,在 20 000 次训练后降到 0.000 1。丢弃率设置为 0.5 以避免过拟合。批处理数量为 128。

2 实验结果及分析

为了验证本文提出的算法的有效性,我们利用先锋移动机器人平台搭建实验平台,平台如图 7 所示。利用移动平台,采集了 10 组场景数据进行测试。10 组场景来源于实验室环境,包括实验室搭建的典型家居环境(见图 3)和办公环境等。所有测试视频均手工标注,包括包含运动目标的视频帧和运动目标在视频帧中的位置。每组测试视频均有 2 个运动人体目标出现,每次出现的视频帧数平均为 32 帧,且各个场景中的运动目标运动速度、运动方向不同。



图 7 实验设备

Fig. 7 Experimental equipment

在测试本文算法之前,首先对每个场景建立全局环境背景三维特征点云模型,并在此基础上利用 ORB-SLAM 的定位技术进行测试。

为了客观评价算法的有效性,使用帧级检测准确率 FD 和召回率 FR 以及帧级定位准确率 FL 对算法进行评价。其中,帧级检测准确率即被检测到真实包含运动目标的视频帧数量与被检测出包含运动目标视频帧数量的比值;帧级召回率即被检测到真实包含运动目标的视频

帧数量与所有真实包含运动目标的视频帧数量的比值。帧级定位准确率 FL 即检测到目标区域与目标真实区域的重合程度。

本文算法包含距离阈值 T_r 和终止阈值 T_Δ 两项参数。距离阈值的大小决定了算法对噪声的容忍度,距离阈值越大,则对噪声容忍越大,但对于与背景(墙壁等)贴近的目标越难检测到,反之亦然。终止阈值 T_Δ 决定了算法对坐标点更新的敏感度,当终止阈值越小,则需要更多迭代次数,耗时更多,但同时,前背景特征点的区分更为准确。根据 ORB-SLAM 算法本身的距离度量尺度,设置不同 T_r ,检测结果如表 1 所示。通过比较可以看到,检测准确率变化不大。这是由于算法在后端使用深度卷积神经网络进行再确认,保持算法检测的准确率。此外,随着阈值的不断增加,算法的召回率不断提高,且定位越来越准确。这是由于随着阈值的增加,同时考虑周围特征点的个数也随之增加,提高前景点提取准确率。但注意到算法的召回率并未有极大的提高,这是由于测试数据中存在一组非常贴近背景(墙壁)的运动目标。在算法耗时上,随着阈值的增加,迭代次数不断增长,耗时越来越多。因此,根据对比结果,结合实际应用,设置距离阈值 $T_r = 0.1$,并经验地设置 $T_\Delta = 0.01$ 。

表 1 算法不同设置下的实验结果对比

Table 1 Comparison of experimental results under different settings

设置	$FD/\%$	$FR/\%$	$FL/\%$	前景提取耗时/(s/帧)
$T_r = 0.01$	98.1	81.2	72.5	0.05
$T_r = 0.05$	97.6	87.9	75.3	0.12
$T_r = 0.10$	98.2	91.4	81.9	0.21
$T_r = 0.20$	97.9	92.5	82.1	0.45

整个算法中,3DMS 操作的算法复杂度为 $O(Tn^2)$,其中, T 为迭代次数, n 为特征点个数。其操作计算时间开销很大,因此,算法在实现过程中,不使用球形邻域,而采用正方体邻域。将特征点距离的计算由开方与乘法变为加减,在保证性能的基础上,减少运算开销。

最后,将本文提出的算法与先前算法在十个场景视频数据上进行比较,测试结果如表 2 所示。文献[15]通过运动补偿进行前景提取,同时,为了确保准备率,其在其后端加入一个深度卷积神经网络进行再确认。因此,其检测准确率较高,但由于其使用运动补偿算法,无法获得较高的召回率和定位准确度。通过比较结果,可以看到,本文算法较文献[15]和文献[29]的算法整体略优。其中,文献[29]基于分块相对光流的背景特征点筛选算法

来估算相机运动,并检查光流与相机运动的一致性,得到背景减除的最终结果。而本文的算法利用三维背景信息,通过鲁棒三维背景估计与减除技术,在各个测试场景中都得到了较好的效果。本文算法在测试场景中的前景特征点检测结果示例如图8所示。

表2 不同算法的实验结果对比

Table 2 Comparison of experimental results with different algorithms (%)

算法	FD	FR	FL
文献[11]	95.2	80.3	75.8
文献[29]	87.9	85.9	80.4
本文算法	98.2	91.4	81.9

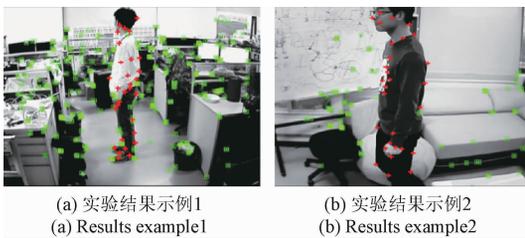


图8 前景特征点提取结果示例(十字点为前景特征点,方形点为背景特征点)

Fig. 8 Examples of foreground feature points (cross points are foreground and square points are background)

然而,本文提出的算法针对于运动目标紧贴背景的情况检测结果略低,这是由于特征点描述子的不稳定,本文算法仅使用特征点的全局位置信息,而未利用特征描述子信息,导致前景特征点难以与背景特征点进行区分。

3 结论

本文提出一种基于三维背景估计的运动目标检测算法。算法避免运动相机造成的动态背景问题,从不变的三维背景出发,提出利用 ORB-SLAM 建立全局环境三维背景特征点云模型,并基于 3DMS 算法进行局部环境三维特征点云的前景特征点提取,最后使用训练的深度卷积神经网络进行候选运动目标区域进行最终确认。该方法充分利用背景信息,在 10 组室内场景中进行测试,实验结果验证了本文方法的准确率和鲁棒性。

参考文献

[1] 孙斌,黄神治. 移动背景下运动目标检测与跟踪技术研究[J]. 电子测量与仪器学报, 2011, 25(3):

206-210.

SUN B, HUANG SH ZH. Target detection and tracking under moving background [J]. Journal of Electronic Measurement and Instrument, 2011, 25(3): 206-210.

[2] 吴克伟,段伟伟,杨学志,等. 雨夜条件下的红外可见光视频融合目标跟踪[J]. 仪器仪表学报, 2016, 17(5): 1131-1139.

WU K W, DUAN W W, YANG X ZH, et al. Infra-visible video fusion object tracking under rainy night condition [J]. Chinese Journal of Scientific Instrument, 2016, 17(5): 1131-1139.

[3] 戴鹏,王雪,谈宇奇,等. 面向行人检测的异构视觉传感网络自适应标定[J]. 仪器仪表学报, 2016, 37(3): 683-689.

DAI P, WANG X, TAN Y Q, et al. Self-adaptive calibration of hybrid visual sensor networks for pedestrian detection [J]. Chinese Journal of Scientific Instrument, 2016, 37(3): 683-689.

[4] 谢英红,庞彦伟,韩晓微,等. 基于 Grassmann 流形和投影群的目标跟踪[J]. 仪器仪表学报, 2016, 17(5): 1140-1147.

XIE Y H, PANG Y W, HAN X Y, et al. Object tracking based on grassmann manifold and projection group [J]. Chinese Journal of Scientific Instrument, 2016, 17(5): 1140-1147.

[5] ZIVKOVIC Z. Improved adaptive gaussian mixture model for background subtraction [C]. IEEE International Conference on Pattern Recognition (ICPR), 2004: 1-4.

[6] 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 M, YANG F B, SHE Q SH. Human motion detection based on the depth image of Kinect [J]. Chinese Journal of Scientific Instrument, 2015, 36(2): 386-393.

[7] 李庆武,仇春春,俞楷,等. 基于多尺度聚合通道特征的实时行人检测[J]. 电子测量与仪器学报, 2015, 29(11): 1691-1697.

LI Q W, QIU CH CH, YU K. Real time pedestrian detection based on multi-scale aggregation channel features [J]. Journal of Electronic Measurement and Instrumentation, 2015, 29(11): 1691-1697.

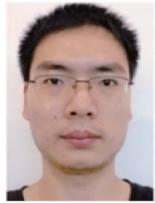
[8] 陆兴华,陈锐俊,池坤丹,等. 基于时频分析的人体红外热信号检测算法[J]. 国外电子测量技术, 2016, 35(2): 55-58.

LU X H, CHEN R J, CHI K D. Infrared thermal signal detection algorithm based on time frequency analysis [J]. Foreign Electronic Measurement Technology, 2016,

- 35(2): 55-58.
- [9] 王永龙,孙东辉,冯登超,等. 移动目标重复轨迹的可视化技术[J]. 电子测量技术, 2014,37(9):53-55.
WANG Y L, SUN D H, FENG D CH, et al. Moving target repeat trajectory visualization technology [J]. Electronic Measurement Technology, 2014, 37(9): 53-55.
- [10] TALEGHANI S, ASLANI S, SAEED S. Robust moving object detection from a moving video camera using neural network and kalman filter[C]. Robot Soccer World Cup XII, 2008:638-648.
- [11] YU X Q, CHEN X N, XU H Q, et al. Moving object detection from moving camera sequences [J]. Proceedings of SPIE, 2011, 7544(12):115-120.
- [12] 薛丽霞, 罗文浩, 王佐成. 动态背景下基于ORB特征匹配的运动目标检测算法[J]. 计算机应用与软件, 2015(10):294-297.
XUE L X, LUO W H, WANG Z CH. Moving object detection algorithm based on orb feature matching under dynamic scene[J]. Computer Applications and Software, 2015(10):294-297.
- [13] SADEGHITEHRAN P, ANGELOV P. ARTOD: Autonomous real time objects detection by a moving camera using recursive density estimation [J]. Novel Applications of Intelligent Systems. 2016, 586(1): 123-138.
- [14] 于红绯, 刘威, 袁淮, 等. 基于子块运动补偿的运动目标检测[J]. 电子学报, 2017, 45(1):173-180.
YU H F, LIU W, YUAN H, et al. Moving object detection based on sub-block motion compensation [J]. Acta Electronic Sinica, 2017, 45(1):173-180.
- [15] ROZANTSEV A, LEPETIT V, FUA P. Detecting flying objects using a single moving camera [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(5):879-892.
- [16] ZHANG H, YUAN H, LI J. Moving object detection in complex background for a moving camera [C]. International Conference on Machine Vision (ICMV), 2013:1-8.
- [17] KIM J, YE G, KIM D. Moving object detection under free-moving camera [C]. International Conference on Image Processing (ICIP), 2010:4669-4672.
- [18] KIM D S, KWON J. Moving object detection on a vehicle mounted back-up camera[J]. Sensors, 2015, 16(1): 23.
- [19] GHOSH A, SUBUDHI BN, GHOSH S. Object detection from videos captured by moving camera by fuzzy edge incorporated markov random field and local histogram matching[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(8):1127-1135.
- [20] DEGOL J, NAM M. A clustering approach for detecting moving objects captured by a moving aerial camera [C]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014:6538-6542.
- [21] 徐克虎, 王天召, 陈金玉, 等. 动态背景下的运动目标检测定位算法研究[J]. 计算机测量与控制, 2013, 21(12):3319-3321.
XU K H, WANG T ZH, CHEN J Y, et al. Research on algorithm of detection and localization of moving object in dynamic scenes [J]. Computer Measurement and Control, 2013, 21(12):3319-3321.
- [22] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]. Computer Vision and Pattern Recognition, 2016:1-13.
- [23] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016: 1-14.
- [24] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015:779-788.
- [25] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision (ECCV), 2016:21-37.
- [26] ANGELOVA A, KRIZHEVSKY A, VANHOUCKE V, et al. Real-time pedestrian detection with deep network cascades [C]. British Machine Vision Conference (BMVC), 2015:1-12.
- [27] 高智勇, 唐文峰, 贺良杰, 等. 基于运动显著性的移动镜头下的运动目标检测[J]. 计算机应用, 2016, 36(6):1692-1698.
GAO ZH Y, TANG W F, HE L J, et al. Moving object detection with moving camera based on motion saliency [J]. Journal of Computer Applications, 2016, 36(6):1692-1698.
- [28] CAI S, GUO H, BAO G, et al. Real-time running event detection via a community patrol robot [J]. International Journal of Advanced Robotic Systems 2016, 13(6): 1-14.
- [29] 孙丰, 秦开怀, 孙伟, 等. 一种针对移动相机的实时视频背景减除算法[J]. 计算机辅助设计与图形学学报, 2016, 28(4):573-579.
SUN F, QING K H, SUN W, et al. A real-time background subtraction algorithm for freely moving

- cameras [J]. *Journal of Computer-Aided Design and Computer Graphics*, 2016, 28(4):573-579.
- [30] YUN K, LIM J, JIM Y C. Scene conditional background update for moving object detection in a moving camera[J]. *Pattern Recognition Letters*, 2017, 88(1): 57-63.
- [31] WU Y, HE X, NGUYEN T Q. Moving Objects detection with freely moving camera via background motion subtraction [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(2):236-248.
- [32] MUR A R, MONTIEL J M M, TARDOS J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. *IEEE Transactions on Robotics*, 2015, 31(5):1147-1163.

作者简介



郭会文,2010年于吉林化工学院获得学士学位,2013年于湖南大学获得硕士学位,2017年于中国科学院大学获得博士学位,现为中国科学院深圳先进技术研究院智能仿生中心研究助理,主要研究方向为计算机视觉、机器人、视频行为分析等。

E-mail:hw.guo@siat.ac.cn

Guo Huiwen received his B. Sc. degree from Jilin Institute of

Chemical Technology in 2010, received his M. Sc. degree from Hunan University in 2013 and received his Ph. D. degree from Chinese Academy of Science in 2017. Now he is a research assistant in Intelligent Biomimetic Center of Shenzhen Institute of Advanced Technology of the Chinese Academy of Sciences. His main research interests include computer vision, robots and intelligent video behavior analysis.



吴新宇,分别在2001年和2004年于中国科学技术大学获得学士学位和硕士学位,2008年于香港中文大学获得博士学位,现为中国科学院深圳先进技术研究院集成所智能仿生中心研究员,主要研究方向为外骨骼机器人、计算机视觉等。

E-mail:xy.wu@siat.ac.cn

Wu Xinyu received his B. Sc. and M. Sc. degrees both from the Department of Automation, University of Science and Technology of China in 2001 and 2004, respectively, and received his Ph. D. degree from Chinese University of Hong Kong in 2008. Now he is a professor in the Center for Intelligent Biomimetics of Shenzhen Institute of Advanced Technology of the Chinese Academy of Sciences. His main research interests include exoskeleton robot and computer vision.