

基于巡逻机器人的实时跑动检测系统

吴庆甜^{1,2} 郭会文^{1,2} 吴新宇^{1,2} 贺 涛³

¹(中国科学院大学深圳先进技术学院 深圳 518055)

²(中国科学院深圳先进技术研究院 深圳 518055)

³(五邑大学信息工程学院 江门 529020)

摘 要 文章提出了一种基于巡逻机器人系统的快速运动人体目标检测方法,采用卷积神经网络作为运动人体目标检测器,在不同摄像头视角和背景条件下,采集了不同姿态的跑动目标正负样本图像,完成了卷积神经网络的训练。为区分前景目标的运动和机器人造成的背景运动,采用了光流特征来描述目标的运动情况并提取出感兴趣区域;为提高跑动目标的检测准确率,将跑动人物的表面特征和运动特征结合起来形成双流数据通道,并输入到卷积神经网络中进行识别。实验结果表明,该系统在室外环境下能够实现 85% 的跑动人体目标检测准确率,并达到 20 帧/秒检测速度。

关键词 跑动检测; 实时系统; 巡逻机器人; 光流

中图分类号 TP 391.3 **文献标志码** A

Real-Time Running Detection from a Patrol Robot

WU Qingtian^{1,2} GUO Huiwen^{1,2} WU Xinyu^{1,2} HE Tao³

¹(Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

²(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

³(College of Information Science, Wuyi University, Jiangmen 529020, China)

Abstract In this paper, a real-time running targets detection method was investigated based on a patrol robot system. The convolutional neural network method was used as the classifier. Running targets with various poses under different camera viewpoints and backgrounds were collected for the training of the neural network. To discriminate the foreground target and the changing background caused by the robot motion, an optical flow-based method was applied. Optical flow of two successive frames taken by on-board camera was used to extract region of interest. To boost the detection efficiency and accuracy, both appearance and motion information of the target are used as input of the convolutional neural network. Experimental results show that under real outdoor scenarios, the detection accuracy can reach 85% with a running efficiency of 20 frames per second.

Keywords running detection; real-time detection; patrol robot; optical flow

收稿日期: 2016-08-31 修回日期: 2017-02-07

基金项目: 国家自然科学基金(61473277)

作者简介: 吴庆甜, 硕士研究生, 研究方向为机器视觉和机器学习; 郭会文, 博士研究生, 研究方向为计算机视觉和机器学习; 吴新宇(通讯作者), 研究员, 研究方向为智能机器人、机器视觉、智能控制和机器学习, E-mail: xy.wu@siat.ac.cn; 贺涛, 硕士研究生, 研究方向为计算机视觉。

1 引言

智能视频分析与监控技术已经成为计算机视觉与机器学习领域的一个研究热点, 在无人驾驶系统^[1]、人群异常事件检测^[2,3]、人机交互或者军事目标检测等领域有着广泛的应用前景。目前, 视频监控技术^[4,5]已经在社区、交通、广场等公众区域, 银行、办公楼、停车场等室内场合得到了广泛应用。

在过去二十年里, 国内外的研究者们已研发了多种智能视频监控系统, 如 1997 年美国国防高级研究项目署设立了以卡内基梅隆大学为首, 麻省理工学院等高校参与的视觉监控重大项目 (Visualsurveillance and Monitoring)^[5], 主要研究用于战场及普通民用场景监控的自动视频理解技术。由马兰里大学开发的实时视觉监控系统^[6]不仅能定位人和人体的各个部位, 通过建立外观模型来实现对多人的跟踪, 还能检测人群中一些简单的异常事件。目前, 国内不少高校和研究单位都在开展相关的研究工作并在智能视频监控技术的应用方面取得了不错的成果, 如中国科学院自动化研究所谭铁牛研究员团队研发的实时智能视频监控预警系统^[7], 已应用于北京地铁 13 号线, 并使该区域的犯罪率降低至原来的 1/8。针对视频序列中的局部异常事件检测, 周灵娟^[8]提出了一种基于稀疏编码模型异常事件检测方法, 但该方法需要提前建立人群事件的类中心字典。为了提高密集人群中异常事件的检测效率, 潘磊^[9]通过分析不同状态下的图像熵来判断异常事件的发生, 该算法在视角较窄的 320×240 像素视频中取得较高的检测效率和实时性。以上大多数监控系统借助与采用大量的固定摄像头来监控场景, 这对于异常情况出现的地点具有随机性, 大量群体异常事件发生在固定摄像头无法监控的区域, 因而无法对整个事件进行及时有效的监控和警报。很多实际应用需要摄像头的运动来实现

目标的连续跟踪, 增大空间范围。因此, 本系统新颖地利用社区巡逻机器人搭载摄像头在室外复杂的环境下进行巡逻, 并用于检测具有随机性的人群异常事件检测, 这对社区环境下的安全防护具有重要意义。

人群异常事件的发生通常聚集在事件的焦点上, 这个焦点就是事件发生的对象与主体, 因此检测感兴趣的焦点目标对于人群异常事件性质的判断具有重要作用。基于视觉的目标检测与追踪, 特别是行人检测, 成为了计算机视觉与机器学习领域一个非常热门的研究方向。Viola 和 Jones^[10]提出了一种结合积分图来快速计算特征和级联的 AdaBoost 算法实现 15 fps 的目标检测速度; Papageorgiou 等^[11]提出了一个通用的可训练框架用来检测静态背景下的目标。由于大多数视频监控系统的感兴趣目标是作为行为主体的人, 因此基于人物目标的智能视频监控系统自然成为人们研究的主要方向。2005 年, 法国研究人员 Dalal 在计算机视觉与模式识别会议 (Computer Vision and Pattern Recognition) 上发表结合梯度直方图 (Histogram of Gradient, HOG) 和支持向量机 (Support Vector Machine, SVM) 实现快速、高精度的行人检测算法^[12], 该算法作为一种经典的目标检测算法也集成到开源计算机视觉库 (OpenCV) 里面。另外, 随着硬件设备, 例如, 图形处理器 (Graphics Processing Unit, GPU) 不断发展和高性能技术的突破, 深度学习成为了机器学习领域一个爆发式增长的研究热点。深度学习的概念由 Hinton 等^[13]于 2015 年提出, 是机器学习研究中的一个新领域。它目的在于建立、模拟人脑进行分析学习的神经网络, 其原理是模拟大脑复杂的神经网络结构来构建一个深度复杂的数学模型, 实现监督或者非监督的学习训练, 特别是卷积神经网络 (Convolutional Neural Network, CNN)。深度学习在图片分类方面有着天然的优势, 并在人脸识别、手写数字识别中取得超过

人眼的识别率。Lecun 和 Bengio^[14]提出的卷积神经网络是第一个真正多层结构学习算法,它利用空间相对关系减少参数数目以提高训练性能。此外,深度置信网络(Deep Belief Networks)^[15]作为一种非监督式的算法,通过贪心逐层训练,为解决深层结构相关的优化难题带来希望。

对于人群异常事件的检测,行人通常不构成异常事件的焦点,而在大多数异常事件中(打劫、追逐、抢劫等),通常都伴随着人物跑动的动作。因此,基于巡逻机器人的实时跑动物体检测系统的目的是作为智能视频监控系统中的人群异常事件检测的基础工作。通常,我们将人群异常事件定义成两类:个体异常事件和群体异常事件^[2]。其中,个体异常事件是指单独个体或目标的行为不同于大部分其他个体的行为,如人行道上的快速跑动、入口的逆行等;群体异常事件是指监控区域中多个个体或者目标的行为迥异于平时群体的行为,如群体斗殴、马路上的追逐,甚至是偷窃、抢劫等恶劣行为。目前,人类的行为动作识别与理解已经在世界范围内展开了大量的研究工作,其中一部分工作是建立一个标准的数据库用于统一的数据输入和算法性能的比较。常用的人群异常事件数据库包含 UMN 数据库(群体恐慌异常)^[16]、UCSD^[17]数据库(人行道上的机动车、骑行等异常)、subway surveillance(地铁监控)数据库^[18]等。这些数据库包括的人群异常事件有:室外环境下的大密度人群恐慌,步行街上的骑行、轮滑和机动车,检票口的逆行、逃票等。

通过以上数据库容易发现人物目标的跑动物体伴随着人群异常事件的发生。因此,我们以人物跑动物体的检测作为出发点,构成检测人群异常事件检测的基础工作。此外,本文通过构建一个实时的目标检测框架,可以用于除跑动物体之外的其他目标检测,如跌倒、攀爬等异常。但基准数据集的限制就是获取的途径相对单一,其场景大部分是在固定摄像头下拍摄的,运动目标检测

与追踪的背景比较单一,与在社区复杂、动态的室外环境下利用巡逻机器人实时地检测异常事件有很大的不同。首先,社区室外环境下,存在建筑物、树木、道路、草坪等复杂背景,对运动目标检测造成较大的干扰;其次,由于摄像头的运动导致场景中背景发生相对运动,造成运动的前景目标与背景难以区分,增加了运动目标检测和追踪的难度,尤其检测非刚体的跑动物体目标,更加具有挑战性。

2 系统设计

本系统是针对在室外环境下基于移动摄像头的跑动物体的实时检测算法进行研究。

系统主要从两个方面考虑,即实时性和鲁棒性。在实时性方面,为了达到 20 fps 甚至更快的实时检测速度,我们使用了一系列处理方法来提取感兴趣区域。由于跑动物体的光流幅值超过了因机器人运动带来的背景光流均值,可以通过阈值分割找到跑动物体所在的位置。由于人体肤色或衣服颜色的相似性,通常通过计算光流无法提取到一个完整的人体外轮廓。因此,在进行光流运算前考虑使用均值滤波器对图像进行钝化,过滤掉图像的高频成分;光流运算完毕后,使用形态学上的闭操作来填充这个不完整的人体外轮廓;最后通过提取外轮廓找到一个完整的跑动物体,将作为候选集输入到卷积神经网络中判别是否为跑动的人体。在鲁棒性方面,由于在室外环境下,很多因素(光照、背景变化、非刚体形变等)影响着跑动物体的检测,给高准确率检测的目标检测提出了挑战。卷积神经网络有着强大的非线性分类能力和自学习提取特征的特点,对于大型图像处理有着出色的表现。因此,我们采用了一种人体表征信息和运动信息相结合的双流卷积神经网络作为分类器。为了训练网络,我们进行大量的室外实验,收集人物跑动过程中多组连

续的跑动动作图片作为 CNN 网络的训练数据。该系统通过卷积神经网络的迭代与训练, 自学习 CNN 特征, 实现了复杂环境下的跑动目标检测。

系统设计流程图如图 1 所示, 主要包括图像预处理、感兴趣区域的提取和双流卷积神经网络的检测。下面将从三个方面详细地阐述系统设计的工作。

2.1 双流卷积神经网络模型

卷积神经网络通过卷积运算使得权值共享, 降低连接的数量级, 同时兼顾二维特征。从算法层面上, CNN 的核心还是与误差反向传播网络相似, 即权值正向传播误差反向传播, 并利用误差来更新每一层的权值。这一结构使得卷积神经网络能够利用输入数据的二维结构, 响应一部分

覆盖范围内的周围单元, 对于大型图像处理有着出色的表现。

但是, 卷积神经网络面临着训练耗时, 其内部参数, 如卷积核、输出层图、网络层数的调优没有严格的数理推导的问题。因此, 我们借鉴一个传统的 5 层卷积神经网络作为检测器。该网络包含: 2 个卷积层、2 个下采样层和 1 个全链接层。通过不断迭代训练网络, 在合理的误差范围内结束训练, 每一层 Kernel 代表不同尺度的局部特征, 最后将全链接层, 也就是 CNN 特征输入到 SVM 中进行分类。跑动动作的局部特征, 如手、脚弯曲的幅值等构成了跑动与非跑动物体的分类特征。由于感兴趣区域的提取依赖于光流, 并且光流图在一定程度上也能反映跑动人物的特征, 结合带有原始信息的灰度图和运动信息的光

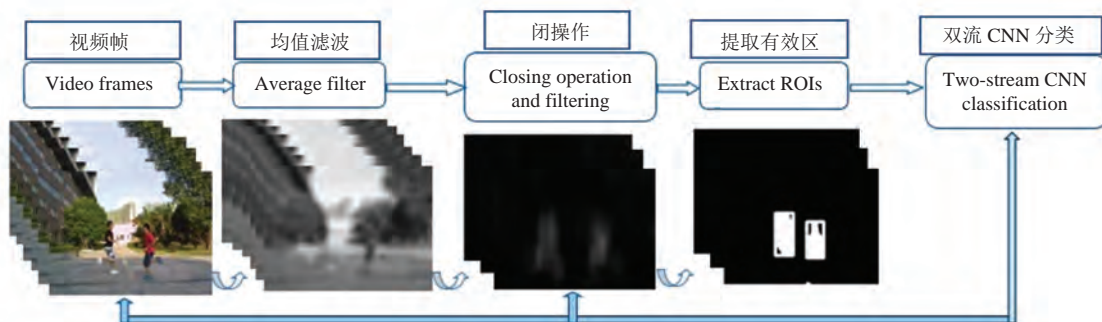


图 1 系统设计的流程图

Fig. 1 The architecture of the real-time running detection system

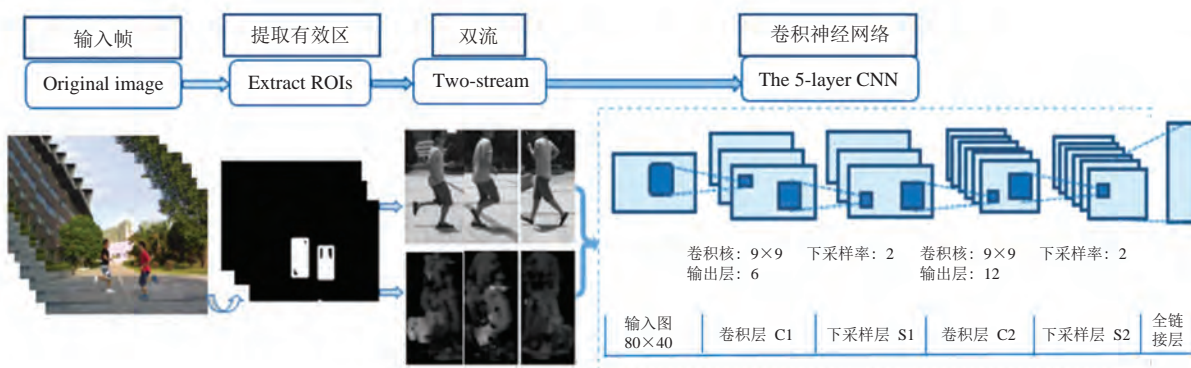


图 2 双流卷积神经网络网络模型

Fig. 2 The structure of the two-stream convolutional neural network

流图构成双流数据并输入到卷积神经网络。该网络的输入图片大小为 80×40 ，输出结果是对感兴趣区域识别的置信度。置信度越接近 1，表明该感兴趣区域被识别为跑动人物的可能性越大。该网络的可视化见图 2，并在图中展示了系统取得最好分类效果时采取的一组参数。

2.2 感兴趣区域的提取

提取感兴趣区域是满足系统实时性的关键操作。对于一幅大小 640×480 像素的原帧图片，如果采用步长为 5 的滑窗在原图上进行滑窗搜索，将会产生 9 600 个候选集要送到卷积神经网络进行识别，这个过程是缓慢的、带有大量的冗余性操作，无法满足实时性要求。由于跑动人物的运动信息很丰富，采用光流来快速提取感兴趣区域。光流是一种图像亮度模式的表观运动。光流表达了图像的变化，由于它包含了目标运动的信息，因此可被观察者用来确定目标的运动情况。

在计算机视觉中，光流扮演着重要的角色，在目标对象分割、识别、跟踪等都有着非常重要的应用。由光流的定义可以引申出光流场，它是指图像中所有像素点构成的一种二维瞬时速度场，其中的二维速度矢量是景物中可见点的三维速度矢量在成像表面的投影。因此，光流不仅包含了被观察物体的运动信息，而且还包含有关景物三维结构的丰富信息。

光流是一种由摄像头与场景发生相对运动造成的运动模式。其基本的方程表达式为：

$$-f_t = f_x \mathbf{u} + f_y \mathbf{v} = \nabla \mathbf{f} \times \mathbf{c}^T \quad (1)$$

其中， $\mathbf{c} = [u, v]$ ； $\nabla \mathbf{f} = [f_x, f_y] \times \mathbf{c}$ ； $\nabla \mathbf{f}$ 和 f_t 分别表示在像素 (x, y) 处的光流向量和在 t 时刻的空间梯度。 f_x 、 f_y 和 f_t 可以表达成如下所示：

$$f_x = \frac{1}{4} (f_{i+1,j,k} + f_{i+1,j+1,k} + f_{i+1,j,k+1} + f_{i+1,j+1,k+1}) - \frac{1}{4} (f_{i,j,k} + f_{i,j+1,k} + f_{i,j,k+1} + f_{i,j+1,k+1}) \quad (2)$$

$$f_y = \frac{1}{4} (f_{i,j+1,k} + f_{i+1,j+1,k} + f_{i,j+1,k+1} + f_{i+1,j+1,k+1}) - \frac{1}{4} (f_{i,j,k} + f_{i+1,j,k} + f_{i,j,k+1} + f_{i+1,j,k+1}) \quad (3)$$

$$f_t = \frac{1}{4} (f_{i,j,k+1} + f_{i+1,j,k+1} + f_{i,j+1,k+1} + f_{i+1,j+1,k+1}) - \frac{1}{4} (f_{i,j,k} + f_{i+1,j,k} + f_{i,j+1,k} + f_{i+1,j+1,k}) \quad (4)$$

其中， i 、 j 、 k 表示水平坐标、垂直坐标和帧数。假设 \mathbf{u}_x 、 \mathbf{u}_y 、 \mathbf{v}_x 和 \mathbf{v}_y 表示 x 、 y 方向上的两个光流向量。通过拉格朗日乘法计算光流，表示如下：

$$F = \mathbf{u}_x^2 + \mathbf{u}_y^2 + \mathbf{v}_x^2 + \mathbf{v}_y^2 + \lambda (f_x \mathbf{u} + f_y \mathbf{v} + f_t)^2 \quad (5)$$

公式(6)给出了运用上述方法计算出的光流值。

$$\begin{cases} \mathbf{u}_{ij} = \mathbf{u}_{ij} - \frac{f_x \overline{u_{ij}} + f_y \overline{u_{ij}} + f_t}{1 + \lambda (f_x^2 + f_y^2)} f_x \\ \mathbf{v}_{ij} = \mathbf{v}_{ij} - \frac{f_x \overline{v_{ij}} + f_y \overline{v_{ij}} + f_t}{1 + \lambda (f_x^2 + f_y^2)} f_y \end{cases} \quad (6)$$

其中， \mathbf{u}_{ij} 、 \mathbf{v}_{ij} 表示运动向量 $\mathbf{u}(i, j)$ 和 $\mathbf{v}(i, j)$ ； $\overline{u_{ij}}$ 、 $\overline{v_{ij}}$ 表示 $\mathbf{u}(i, j)$ 和 $\mathbf{v}(i, j)$ 的近邻均值。

通过计算连续两帧图片的光流，快速找到感兴趣区域是至关重要的。为了进一步提高实时性的检测性能，系统采用了 Lucas-Kanade (LK) 光流算法来提取感兴趣区域。这是因为 LK 光流作为一种稀疏的局部光流法，使用局部区域的像素运动的一致性替代全局速度平滑的约束条件，使得求解速度更快。

图 3 展示了系统每一步操作的效果图。其中，图 3(a) 是一幅彩色原帧图片，大小为 640×480 ；图 3(b) 是采用 15×15 的均值滤波器模糊化的结果图；图 3(c) 是对连续两帧图片进行光流运算后的灰度图；图 3(d) 是进行闭操作的效果；图 3(e) 是感兴趣区域的提取，白色框表示人体外轮廓；图 3(f) 是卷积神经网络将感兴趣区域识别为跑动人物的结果。

2.3 感兴趣区域的识别

光流信息在该系统中起着重要的作用，系统



图 3 效果处理流程图

Fig. 3 Example of the cascade post-processing results

不仅采用光流来提取感兴趣区域, 还将光流图作为一个数据通道输入到双流卷积神经网络中。为尽可能地减少误检和漏检及提高正确检测率, 系统采用了卷积神经网络作为分类器, 并结合了跑动人体的表面特征信息和跑动人体的运动信息构成双流数据, 并输入到卷积神经网络得到对应的识别置信度 P_1 、 P_2 。我们这样给出最终结果的置信度 P 定义为:

$$P = \alpha \times P_1 + \beta \times P_2 \quad (7)$$

该公式表明最终识别结果取决于双流的置信度, 但给予两个置信度不同的权重, 分别为 α 、 β 。这是因为, 虽然光流图在一定程度上刻画了人物跑动过程中的手脚细节, 但原始的灰度图始终包含了最全面的人物跑动信息。因此, 给予两个置信度不同的权值, 得到候选集的最终置信度, 如公式(7)所示。

3 实验结果与分析

在室外环境下, 我们进行了多次实验来评

估该系统在实时性和检测准确率上的性能。实验平台是基于一台 16 G 内存、3.5 GHz CPU 的 surface 平板电脑, 机器人移动平台是 Pioneer 2, 摄像头默认的分辨率是 640×480 。

3.1 实时性

实时检测对于多数应用系统来说是一个不可或缺的要求。该系统将内置于一款搭载摄像头的社区巡逻机器人上。如果没有使用提取感兴趣区域的手段, 采用步长为 5 的滑动窗口在一幅原图上进行滑窗检测跑动的人物, 将会产生 9 600 个滑窗并输入到卷积神经网络中识别, 整个过程将会耗时 2.5 s, 意味着滑窗法的检测速度是 0.4 fps, 这个检测速度远远不能达到实时性检测的要求。通过提取感兴趣区域, 一般针对一幅原图产生的感兴趣区域的数量级为 10^1 , 远远小于滑窗法产生 10^4 (9 600) 数量级的候选集, 这在很大程度上避免了冗余的候选区域的重复识别, 大大降低了运算量和时间复杂度。通过提取感兴趣区域, 整个流程下来检测一帧图像的时间是 50 ms, 达到了 20 fps 的检测速度。表 1 详细地展示了系统

每一步处理所消耗的时间。由于在动态摄像头检测跑动的人物目标相对新颖，我们比较了几种实现快速目标检测的算法，结果如表 2 中所示。通过比较发现，我们的算法用于动态背景下的跑动人物目标检测，不仅可以在复杂的动态背景下有效，还能够达到最高的 20 fps 的实时检测速度。值得一提的是，其他几种方法均为对行人目标检测，而本文方法是对跑动目标检测。

表 1 流程操作的时间消耗情况

Table 1 Distribution of computing time

模块化操作	实时耗时(ms)
计算光流	10
闭操作处理	6
提取候选集	25
CNN 分类	10
合计	51

表 2 静态背景下的行人检测速度比对

Table 2 Speed comparison of several fast picture-based method for object detection

方法	在 640×480 视频上的检测速度 (fps)
HOG-LBP+SVM ^[19]	4
GPU HOG+SVM ^[20]	10
Integral Image+AdaBoost ^[21]	15
C4 ^[22]	20
本文方法	20

3.2 检测准确率

与在标准数据库上进行识别准确率的比对不一样，跑动人物检测的数据库很少，特别是在室外场景下。对于应用在社区巡逻机器人上的智能视频分析系统来说，最终目的是对异常情况的检测和警报。因此，对于一个完整的人物跑动动作来说，漏检其中的一两帧不影响跑动动作的检测与警报，这在一定程度上反映了该系统能够容忍少量的漏检。由于在室外环境下，系统对跑动人物检测的鲁棒性要求高，在标准数据库下容易陷入过拟合。因此我们进行了很多次的室外实验，

并手工采集了 30 000 个跑动人物样本来建立跑动数据库。在离线的数据库上，我们分别比较了使用 SVM 和双流 CNN 作为分类器来识别跑动的人体目标，实验结果如图 4 所示。结果显示，SVM 作为一种高维小样本分类器，达到了 93.5% 的识别准确率，而双流 CNN 达到了 99.85% 的识别准确率，这也充分说明了卷积深度神经网络在图片识别中的优势。

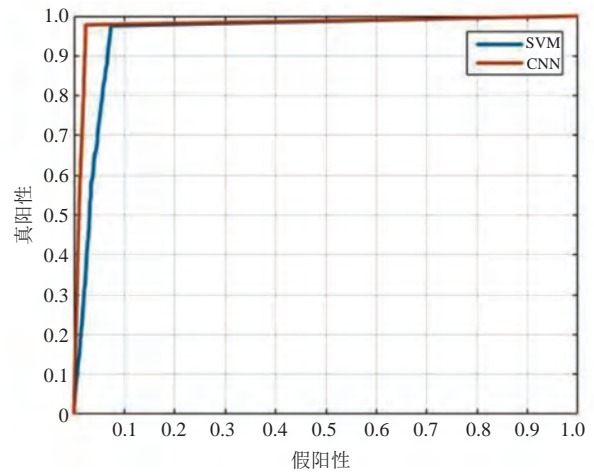


图 4 跑动检测准确率的 ROC 曲线

Fig. 4 ROC curves for running detection accuracy

此外，在评价该系统在室外环境下的正确检测率时，我们选取了一个包含 400 帧的视频流，该视频是在室外复杂的条件下拍摄的。通过人工标记跑动人物的数量、系统检测出来的跑动人体数量和系统误检的数量进行评估与分析。通过数据统计，基于双流 CNN 的跑动目标检测系统在室外动态的背景环境中达到 85% 的正确检测率，但伴随着少量的漏检和误检。其中，漏检的原因是跑动人物距离摄像头太远或者跑动人速度过慢导致对应区域的光流场太弱而被当成背景区域，无法检测出跑动的人物；误检是因为样本空间不够大，导致在室外环境下遇到非刚体的行人发生误检。另外，图 5 分别展示了在不同的室外场景下检测到的跑动人体目标。虽然该系统在复杂的室外环境中取得实时性的检测效果，但不

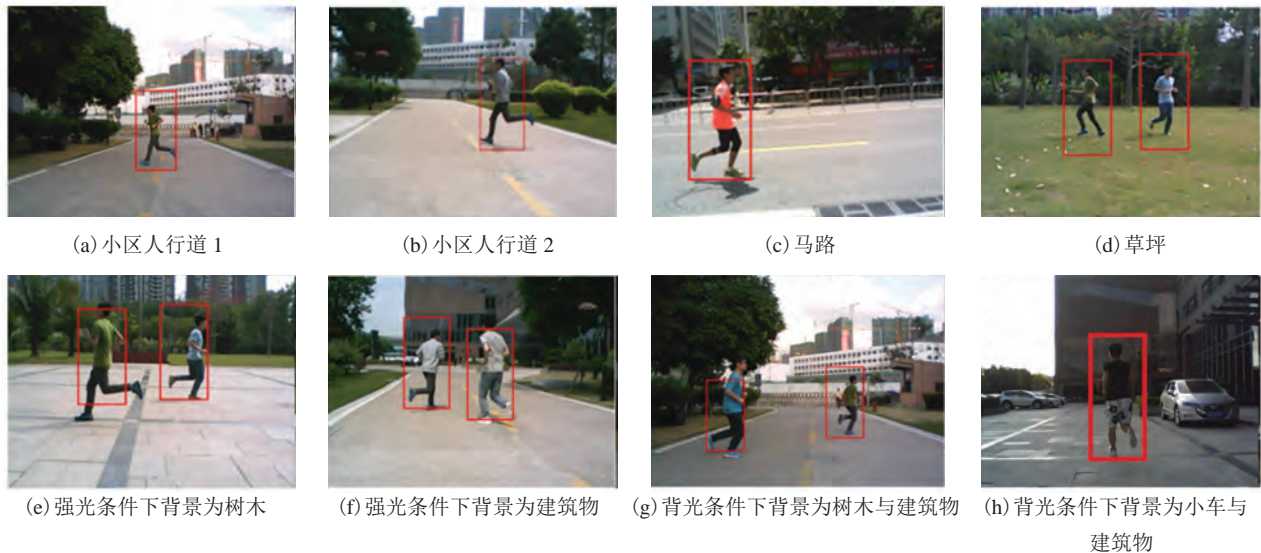


图 5 不同场景下的跑动目标检测结果

Fig. 5 Example of running people detection

能很好地检测 12 m 以外的跑动目标。这是由于摄像头的透视投影效应, 远处的人体目标成像较小, 计算对应区域的光流值时也偏小, 在预处理阶段会通过闭操作而滤除到光流值偏小的区域, 因此远处的跑动目标会发生漏检情况。

4 总结与展望

本文针对摄像头相对检测场景运动情况下的人体跑动提出了一个实时性、较高检测精度的系统。对于实时性, 该系统利用光流快速提取感兴趣区域, 并通过均值滤波、形态学上的闭操作进行了外轮廓提取。对于复杂的室外环境下的跑动目标检测, 该系统利用大量的样本来训练了一个 5 层的卷积神经网络, 并充分利用光流信息与表面特征结合构成了双流通道以提高检测精度。通过大量的室外实验, 证明该系统在复杂的社区环境下能够达到 20 fps 的跑动检测速度和较高的跑动人物检测准确率。考虑到异常事件都是一个连续的动作, 后续工作会结合卡尔曼滤波或模板匹配来进行跑动目标的追踪。

参 考 文 献

- [1] 乔维高, 徐学进. 无人驾驶汽车的发展现状及方向 [J]. 上海汽车, 2007(7): 40-43.
- [2] 吴新宇, 郭会文, 李楠楠, 等. 基于视频的人群异常事件检测综述 [J]. 电子测量与仪器学报, 2014, 28(6): 575-584.
- [3] Ren HM, Liu WF, Olsen SI, et al. Unsupervised behavior-specific dictionary learning for abnormal event detection [C] // The 26th British Machine Vision Conference, 2015: 28.1-28.13.
- [4] Ravanbakhsh M, Nabi M, Mousavi H, et al. Plug-and-play CNN for crowd motion analysis: an application in abnormal event detection [J]. arXiv:1610.00307v1, 2016.
- [5] Collins RT, Lipton AJ, Kanade T, et al. A system for video surveillance and monitoring [R]. CMU-RI-TR-00-12, Pittsburgh: Carnegie Mellon University, 2000.
- [6] Haritaoglu I, Harwood D, Davis LS. W4: Who? When? Where? What? A real time system for detecting and tracking people [C] // Proceedings of the 3rd International Conference on Face & Gesture Recognition, 1998: 222.
- [7] Huang KQ, Tan TN. Vs-star: a visual interpretation

- system for visual surveillance [J]. *Pattern Recognition Letters*, 2010, 31(14): 2265-2285.
- [8] 周灵娟. 人群中局部异常事件检测 [J]. *电脑知识与技术*, 2015, 11(33): 125-128.
- [9] 潘磊. 基于图像熵的密集人群异常事件实时检测方法 [J]. *计算机科学与探索*, 2016, 10(7): 1044-1050.
- [10] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C] // *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001: 511-518.
- [11] Papageorgiou CP, Oren M, Poggio T. A general framework for object detection [C] // *The 6th International Conference on Computer Vision*, 1998: 555-562.
- [12] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005: 886-893.
- [13] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [14] Lecun Y, Bengio Y. Convolutional networks for images, speech, and time-series [C] // *The Handbook of Brain Theory and Neural Networks*, 1995: 255-257.
- [15] Hinton GE. Deep belief networks [J]. *Scholarpedia*, 2009, 4(6): 5947.
- [16] University of Minnesota. Detection of events: detection of unusual crowd activity [EB/OL]. 2009-04-12[2016-08-31]. http://www.mha.cs.umn.edu/proj_events.shtml.
- [17] University of California. UCSD anomaly detection dataset [EB/OL]. 2010-10-10[2016-08-31]. <http://www.svcl.ucsd.edu/projects/anomaly/dataset.html>.
- [18] Adam A, Rivlin E, Shimshoni I, et al. Robust real-time unusual event detection using multiple fixed-location monitors [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(3): 555-560.
- [19] Wang X, Han TX, Yan S. An HOG-LBP human detector with partial occlusion handling [C] // *IEEE International Conference on Computer Vision*, 2009: 32-39.
- [20] Wojek C, Dorkó G, Schulz A, et al. Sliding-windows for rapid object class localization: a parallel technique [C] // *Proceedings of the 30th DAGM Symposium on Pattern Recognition*, 2008: 71-81.
- [21] Zhu Q, Yeh MC, Cheng KT, et al. Fast human detection using a cascade of histograms of oriented gradients [C] // *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2006: 1491-1498.
- [22] Wu JX, Geyer C, Rehg JM. Real-time human detection using contour cues [C] // *IEEE International Conference on Robotics and Automation*, 2011: 860-867.