

复杂场景下的运动目标识别算法^①

宫法明, 李隽然, 马玉辉

(中国石油大学(华东)计算机与通信工程学院, 青岛 266580)

通讯作者: 李隽然, E-mail: lixiaoran93@163.com

摘要: 目标识别是计算机视觉的基本目的, 同时也是人工智能领域的重要组成部分之一. 随着信息化时代的来临, 视频采集工具的普及, 海量的视频数据给人工识别带来了巨大挑战. 现阶段, 在智能交通领域、生产质检领域等简单场景中, 视频识别技术已经得到广泛的应用. 如何从复杂场景中实现目标的识别和检测则成为了更加重要和困难的问题. 针对该问题, 本文提出了一种复杂场景下的运动目标识别算法. 首先, 提出一种改进的光流算法, 通过时间序列以及空间像素变化对运动目标区域进行快速标记; 其次, 对目标区域进行滑动窗口检测, 匹配人体各部位模型, 并将反馈信息利用树形结构进行人体建模, 实现在复杂场景下识别运动目标. 通过实验进行评估, 该方法能够在保证较高准确率的情况下, 相比基于深度学习的检测算法检测速度更快, 可以满足实时监测的要求.

关键词: 计算机视觉; 光流算法; 滑动检测; 模型匹配; 目标识别

引用格式: 宫法明, 李隽然, 马玉辉. 复杂场景下的运动目标识别算法. 计算机系统应用, 2018, 27(8): 193-197. <http://www.c-s-a.org.cn/1003-3254/6465.html>

Recognition Algorithm of Moving Target in Complicated Scenes

GONG Fa-Ming, LI Xiao-Ran, MA Yu-Hui

(College of Computer & Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: Target recognition is the basic purpose of computer vision, it is also one of the key components in the field of artificial intelligence. With the advent of the information age, the popularity of video capture tools, massive video data to human identification has brought great challenges. At this stage, video recognition technology has been widely used in simple scenes such as intelligent transportation field and production quality inspection field. How to realize the target recognition and detection from complex scenes has become a more important and difficult issue. In response to this problem, this paper presents a moving target recognition algorithm in complex scenes. First, an improved optical flow algorithm is proposed to mark the moving target region quickly by time series and spatial pixel changes; Secondly, the sliding window of the target area is detected to match the model of each part of the human body, and the feedback information is modeled by a tree structure. Through experiments, this method can detect faster than detection algorithm based on depth learning while ensuring high accuracy, and can meet the requirements of real-time monitoring.

Key words: computer vision; optical flow algorithm; sliding detection; model matching; target recognition

1 引言

近年来, 运动目标识别^[1]已经成为计算机视觉领域的研究热点, 引起了众多学者的关注. 随着大数据的到

来, 引用大数据的治理体系包括对大数据的采集、存储、管理、应用的全生命周期的管理, 人们希望计算机能够自主的识别并且理解人类的日常动作, 使计算

① 基金项目: 科技部创新方法工作专项 (2015IM01030)

Foundation item: Special Project for Innovative Work Methods of Ministry of Science and Technology (2015IM01030)

收稿时间: 2017-12-05; 修改时间: 2017-12-27; 采用时间: 2018-01-08; csa 在线出版时间: 2018-07-28

机能够和人们更加自然的交流。目前运动目标识别主要应用于人机交互、医疗保健、智能安防、家居智能等领域。

针对这一课题,国内外的研究人员提出了多种解决方法。根据种类可以分为基于背景建模的目标检测和基于前景建模的目标检测。其中常见的基于背景建模的方法有光流法^[2]、背景减法、帧间差分法^[3]等。该方法的方法精确度低且受环境因素的影响非常大。基于前景建模的目标检测方法通过对大量的目标数据进行训练学习,得到相应的检测器对目标进行分类,通常分为基于全局模板的目标检测器^[4]、基于部件模型的目标检测器^[5]以及基于深度学习的目标检测器^[6],该方法的方法可以更准确的检测目标,但是对计算机计算硬件要求过高,检测速度慢,很难实现实时检测。

综上所述,运动目标检测技术在我们的生活中得到了应用,但大多适用于场景简单、目标单一的情况。对于场景复杂的环境,往往得不到满意的结果。针对这一问题,本文提出了一种复杂场景下的运动目标识别算法。

我们的实验数据来源于海上平台的监控视频。由于在海上平台上工作存在一定的危险性,以及与石油生产相关的化工原料和设备至关重要,所以对海上平台状况的监控就显得尤为重要。所有的海上平台都配备了视频监控,以监测工作人员的安全和非法人员入侵等情况,这就要求我们能够在大量的视频中识别出运动目标。在传统的行人检测算法中,高分辨率的图像提高了滑动窗口检测^[7]程序的操作复杂性。此外,人体还可能被海上平台的复杂管道所遮挡,海水的流动会影响移动区域划分的准确性,使得目标识别更加困难。

针对海上工作平台监控视频中管道遮挡复杂、色调单一、人体角度变形等问题。我们提出适用于该复杂场景下的运动目标识别算法进行解决。如图1所示,为目标识别算法流程图。首先,我们改进了光流算法,以图像序列中光流向量为基础,利用数据和空间约束检测出图像序列之间的光流向量变化,利用鲁棒性^[8]对参数进行优化,进行迭代和矫正获得光流信息,对运动目标区域进行快速标记,从而过滤掉视频中的静态目标区域,大大降低计算量,提高了目标检测的速度;其次,将运动目标区域进行矩形标准化处理,提取的运动区域进行滑动窗口检测,比对图像与身体部分外形模板的相似度,利用树形结构对人体进行3D骨架建模^[9],将图像对人体各个部分的响应以信息传递的方式综合

计算,实现人体在图像中的定位;通过回传进行人体部分的定位^[10],使用滑动窗口的方法获取运动目标的动作时序,通过连接各个局部动作片段的特征描述整个动作序列;通过构建的人体模型进行复杂场景下的运动目标检测。

本文的以下部分组织如下:在第2节中,介绍了在复杂场景下实现运动目标检测的算法原理。在第3节中,在海上石油平台的监控视频数据集中对本文提出的算法进行实验验证,并分析了其优劣。在第4节中对本文内容进行了总结,并提出了未来的研究方向。

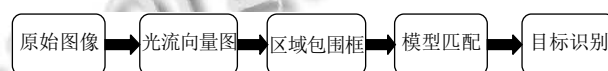


图1 目标识别算法流程图

2 复杂场景下的运动目标检测算法

2.1 预处理和光流检测

传统的光流算法可以看作一种简单实用的图像运动表达式。通常情况下,该算法会根据物体的运动信息来判断物体是否发生移动。从图像像素的角度观察物体,图像中亮度模式的变化是物体表面上的点在空间中移动速度的表达方式。光流的研究是通过时域变化来确定像素位置的移动以及图像序列中像素强度数据的相关性,即研究图像灰度在时间上的变化和物体结构与运动之间的关系。基于匹配、灰度、频域和神经动力学四个方面,我们得到了光流的计算方法。此外,因为缺乏鲁棒性的最小二乘估计,所以光流问题的许多常见解决方案由此而制定。光流估计中的当前工作忽略了违背亮度恒定性假设的问题,反而侧重解决违背运动边界的空间平滑度的问题,使得计算量较大,检测速度较慢,检测结果为物体轮廓。针对传统光流算法存在的不足,我们提出了改进后的光流法。如图2所示为改进的光流算法流程图。

首先,我们需要解决的问题是初始化参数。我们假设 $I(x, y, t)$ 是时间为 t 时的图像强度,所以维持约束的目标函数是:

$$I(x, y, t) = I(s + u\delta t, y + v\delta t, t + \delta t) \quad (1)$$

其中, (u, v) 表示点 (x, y) 在时间 t 时的状态, (u, v) 是一个点的水平和垂直方向的速度,当 δt 很小时,数据约束的目标函数为:

$$E_D(u, v) = \sum_{(x,y) \in \mathfrak{R}} (I_x u + I_y v + I_t)^2 \quad (2)$$

其中, I_x 表示在 \mathfrak{R} 目标区域上 I 对 x 的部分推导. 当 \mathfrak{R} 非常小时, 需要通过增加空间一致性来进一步限制 $u = (u, v)$ 的解, 空间一致性的目标函数是:

$$E_S(u, v) = u_x^2 + u_y^2 + v_x^2 + v_y^2 \quad (3)$$

其中, u_x 表示 u 对 x 的部分推导, 通过不断改变目标函数和优化鲁棒性估计, 结合数据约束和空间一致性约束, 总目标函数为:

$$E(u) = E_D(u) + \lambda E_S(u) \quad (4)$$

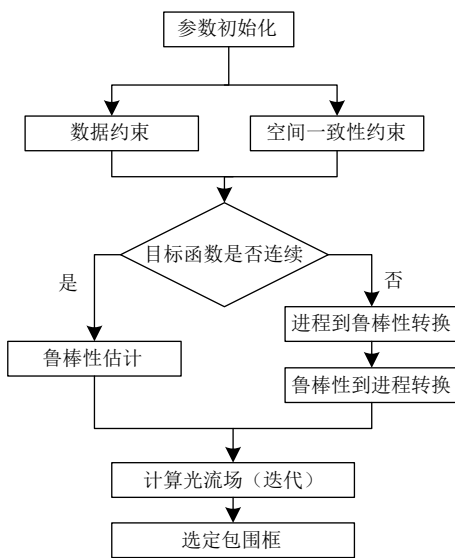


图2 改进的光流算法流程图

鲁棒性估计解决了在一组测试数据中找到最适合模型的参数值问题. 我们的目标是找到使剩余错误达到最小的值, 在获得估计值之后, 我们需要完善光流算法的目标函数. 经过这些步骤之后, 根据目标函数和优化约束得到的集合, 再经过几次迭代, 计算出与最优解一致的优化约束和目标函数. 此外, 通过使用同时过度松弛算法 (SOR)^[11] 找到局部最小值.

在我们的方法中使用的图像具有各种分辨率, 并且采用了多层的金字塔表示法. 第一次迭代中的图像输入就具有 4 层金字塔的高分辨率, 下抽样时逐层减少, 迭代次数为 3 时, 每次迭代的金字塔有 2 层. 此时, 第二层是第一层的 0.8 倍.

通过上述过程, 我们通过计算获得了描述图像运动属性的光流场, 并且根据运动背景的具体特点, 设定

光流的阈值. 然后我们提取每个有效运动区域以及每个区域和边界框的像素列表.

因为两个连续图像之间的时间间隔很短, 所以我们可以得出结论: 连续图像中的对象的位置是相似的. 此外, 两个连续图像中的一个对象表明它们的位置的边界框可以重叠, 如果边界框的重叠比例较大, 那么两个边界框代表相同目标的可能性将会更大. 由于运动的连续性, 我们计算出两个连续帧之间的运动物体的速度, 以便估计该物体在第三帧中的位置, 并把该结果与通过找到两个运动矢量的最小点乘积而获得的计算结果进行比较. 我们还提出一个公式来计算可能性, D 表示两个边界框之间的距离, 比率是这两个边界框的重叠系数. 当比值为 1 时, 即边界框完全重叠, 此时计算结果即为边界框之间的距离.

2.2 滑动窗口模型匹配法

我们采用了部位混合 (MoP) 模型, 其中人体的左侧部分和右侧部分由图像中的位置定义, 躯体部分用多个模板的混合来表示. 通过对训练人员身体部位轮廓的相对位置模型进行图像计算, 计算出每个滑动窗口位置对身体部位轮廓的响应值^[12].

我们认为如头、颈、肩、肘、手、躯干、胯部、膝盖和脚等主要关节作为树结构的节点, 人的肢体作为树结构中的节点与节点之间的连接. 头部是根节点, 双手和双脚都是叶节点, 这种树型结构可以表示人体的铰链模型.

利用训练好的人体各部位相对位置关系模型, 对图像进行卷积操作, 计算图像中每个滑动窗口位置对身体部分外形的响应 $S(p_i)$:

$$S(p_i) = \sum_{i \in V} \omega_i^{t_i} \cdot \phi(I, p_i) \quad (5)$$

利用身体部分混合模型, 将每个身体部分的响应通过消息传递的方式传递到人体树形模型的根节点, 得到图像各个位置对整个身体模型的响应值 $S(t)$:

$$S(t) = S(t) + \sum_{i \in V} \omega_i^{t_i} \cdot \phi(I, p_i) + \sum_{i,j \in E} \omega_{ij}^{t_i, t_j} \cdot \psi(p_i - p_j) \quad (6)$$

将利用滑动窗口得到的每个动作片段用特征直方图表达; 使用一个长向量把所有片段的特征直方图表达串联起来, 使用局部训练法, 通过单独对每个动作样本集训练特征词汇从而得到覆盖整个动作训练集的特征词汇表; 特征码本得到后, 对每一个目标特征描述子^[13]

进行量化编码;令 P 为从动作序列中获得的目标特征描述子,每个特征描述子的维度为 D ,总共有 N 个特征:

$$p = [p_1, p_2, \dots, p_N] \in R^{D \times N} \quad (7)$$

令训练得到的码本有 M 个码字,通过改进局部坐标编码方式进行编码:

$$c_i = \arg \min_{c \in R^M} \|p_i - F_c\| + \lambda \|d_i \Delta c\|^2, \text{ s.t. } 1^T c = 1 \quad (8)$$

其中, Δ 表示元素间相乘, d_i 是局部适配因子,它是根据输入目标 p_i 与码本中各码字间的相似性来赋值,看作特征基的权重; d_i 的选取首先要满足相似距离越近权重越大;选取 d_i 为非线性的表达式,同时引入了最近相似距离 DM 作为比较对象:

$$d_i = \exp\left(-\frac{((D(p_i, F) - D_M)/D_M)^2}{2\sigma^2}\right) \quad (9)$$

其中, $D(p_i, F)$ 表示 p_i 和 F_i 的欧氏距离.

$$D_m = \min\{D(p_i, f_j), D(p_i, f_2), \dots, D(p_i, f_M)\} \quad (10)$$

利用身体部位的混合模型,通过消息传递的方式将人体各部位的响应值传递给人体树结构模型的根节点(相应的头节点),然后,将每个部分的响应值都转换为人体模型.根据对人体定位的结果,在树结构中,MoP 从树结构的叶节点开始,并从每个子节点向其父节点传递最大匹配的分数值.通过消息传递的方式,所有的身体部位都会向根节点传递分数值.然后,根节点位置的分数值是评定在该位置的人体姿态的最终得分.最后,通过从根节点追溯到叶节点,MoP 将固定身体各部位的位置.

2.3 目标识别

利用树形结构对人体建模,将图像对人体各个部分的响应以信息传递的方式综合计算,实现人体在图像中的定位^[14].通过回传进行人体部分的定位,使用滑动窗口的方法获取运动目标的动作时序,通过连接各个局部动作片段的特征描述整个动作序列.最后,根据

构建的人体模型来完成特征匹配识别工作,计算出运动区域的对应关系,实现复杂场景下运动目标识别.

3 实验结果与分析

在计算机视觉领域,基于深度学习的卷积神经网络算法是当前热门的一种视频识别算法.本实验将使用同样配置的计算机,分别使用基于深度学习的算法和本文提出的基于复杂场景的识别算法对相同视频进行检测,并得出两组实验的速度参数与准确率参数.

本实验采用普通的 PC 机,将 CPU 为 2.4 GHz,内存为 8 G 的 PC 机作为硬件平台,搭载英伟达 GTX 1060 型号显卡.采用 Matlab R2014a 和 Caffe 作为 Windows 10 系统中的程序环境.

实验所用的视频来自海上平台,由监控设备提供.在实验中,监控设备保持不动,视频序列以海洋工作平台作为背景.

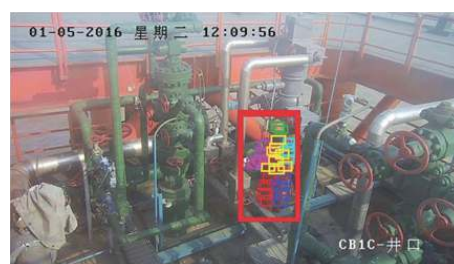


图3 复杂场景下的目标检测法所得结果图



图4 卷积神经网络法所得结果图

表1 实验检测结果对比表

检测算法	检测视频时长 (s)	包含目标帧	不包含目标帧	误测率 (%)	漏测率 (%)	检测时长 (s)
复杂目标检测法	73	865	887	5.62	6.15	296
卷积神经网络法	73	865	887	3.76	4.31	389

如图3、图4所示,分别是使用本文提出的复杂场景下的目标检测算法和使用卷积神经网络算法所得出的结果图.其结果表明两种算法均能在复杂的场景下

实现目标物体的准确识别.本文提出的目标检测算法可以通过估算边界框的大小来判断待检测人体的大小,然后检测人体的所有部位,并用不同的颜色标记它们,

以便于实现人体的姿态识别。

如表1所示,为两组检测算法在同一数据集的基础上进行实验的结果图。其中,包含目标帧为视频帧中包含运动目标的个数,不包含目标帧为视频帧中不包含运动目标的个数,误测率为检测错误的视频帧占不包含目标帧的比率,漏测率为检测遗漏的视频帧占包含目标帧的比率,检测时间为检测该组数据集所消耗的时间。通过实现可以得出:

(1) 在相同的数据检测中,本文提出的方法在检测效率上要优于基于深度学习的检测算法。

(2) 在相同的数据检测中,本文提出的方法在检测质量上略低于基于深度学习的检测算法,但在资源的消耗程度上要优于深度学习检测算法。

实验表明,我们所提出的基于复杂场景下的目标识别算法,在复杂场景下进行目标检测比基于深度学习的方法更具适应性。

4 结论和未来工作

本文所提出的复杂场景下的运动目标识别算法,可以在保证在较高准确率的情况下,更快地实现运动目标的识别与跟踪,该方法可以对需要安全监控的特定区域(如复杂的工作环境)内的异常事件进行有效自动的报警。由于海上平台远离陆地,且工作环境相对复杂,因此,该项技术的应用有效地保证了工作人员的安全以及平台工作的顺利开展。下一步工作,将计划利用检测到的运动目标进行特征提取^[15],实现运动目标的多姿态识别。

参考文献

- Mishne G, Talmon R, Cohen I. Graph-based supervised automatic target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(5): 2738–2754. [doi: 10.1109/TGRS.2014.2364333]
- 肖军, 朱世鹏, 黄杭, 等. 基于光流法的运动目标检测与跟踪算法. *东北大学学报(自然科学版)*, 2016, 37(6): 770–774. [doi: 10.3969/j.issn.1005-3026.2016.06.003]
- 屈晶晶, 辛云宏. 连续帧间差分与背景差分相融合的运动目标检测方法. *光子学报*, 2014, 43(7): 213–220.
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA. 2005. 886–893.
- Yan JJ, Lei Z, Wen LY, *et al.* The fastest deformable part model for object detection. *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 2497–2504.
- Ramakrishnan SK, Pal A, Sharma G, *et al.* An empirical evaluation of visual question answering for novel objects. *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 7312–7321.
- 邵奇可, 李路, 周宇, 等. 一种基于滑动窗口优化算法的行人检测算法. *浙江工业大学学报*, 2015, 43(2): 212–216. [doi: 10.3969/j.issn.1006-4303.2015.02.019]
- 杨涛, 李静, 潘泉, 等. 基于场景模型与统计学习的鲁棒行人检测算法. *自动化学报*, 2010, 36(4): 499–508.
- Belagiannis V, Amin S, Andriluka M, *et al.* 3D pictorial structures for multiple human pose estimation. *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 1669–1676.
- Tang SY, Andriluka M, Schiele B. Detection and tracking of occluded people. *International Journal of Computer Vision*, 2014, 110(1): 58–69. [doi: 10.1007/s11263-013-0664-6]
- Wen ZW, Yin WT, Zhang Y. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 2012, 4(4): 333–361. [doi: 10.1007/s12532-012-0044-1]
- Wei SE, Ramakrishna V, Kanade T, *et al.* Convolutional pose machines. *arXiv: 1602.00134*, 2016: 4724–4732.
- 方智文, 曹治国, 朱磊. 基于梯度角度的直方图局部特征描述子的图像匹配算法. *计算机应用*, 2015, 35(4): 1079–1083. [doi: 10.11772/j.issn.1001-9081.2015.04.1079]
- Rother C, Minka T, Blake A, *et al.* Cosegmentation of image pairs by histogram matching - Incorporating a global constraint into MRFs. *Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, NY, USA. 2006. 993–1000.
- 刘丽, 匡纲要. 图像纹理特征提取方法综述. *中国图象图形学报*, 2009, 14(4): 622–635. [doi: 10.11834/jig.20090409]