

基于地面估算与目标位置估计的多目标跟踪算法^{*}

何 嘉¹, 奚峥皓^{1,2}, 阚 秀¹

(1. 上海工程技术大学 电子电气工程学院, 上海 201620;

2. 清华大学 计算机系 智能技术与系统国家重点实验室, 北京 100091)

摘 要: 提出了一种基于地面移动平台的多目标跟踪系统。系统将利用移动平台上摄像头的部分外参, 构建与视频流对应的地面栅格图。系统将用 HOG 算法检测目标, 并在构建的地面栅格图中计算地面占用, 来跟踪目标。为了保证移动平台下多目标跟踪的鲁棒性, 算法可以根据视频序列中消失点变化情况补偿目标运动模型, 并与地面栅格图中的目标三维信息相结合。通过与其他算法在相同视频序列中的结果比较, 发现所提出的算法是有实际效果的。

关键词: 多目标跟踪; 移动平台; 概率占用图; 消失点; 目标估计

中图分类号: TN919.82; TP23

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.190737

中文引用格式: 何嘉, 奚峥皓, 阚秀. 基于地面估算与目标位置估计的多目标跟踪算法[J]. 电子技术应用, 2020, 46(8): 112-116.

英文引用格式: He Jia, Xi Zhenghao, Kan Xiu. Multi-target tracking algorithm based on ground projection and target position estimation[J]. Application of Electronic Technique, 2020, 46(8): 112-116.

Multi-target tracking algorithm based on ground projection and target position estimation

He Jia¹, Xi Zhenghao^{1,2}, Kan Xiu¹

(1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science, Tsinghua University, Beijing 100091, China)

Abstract: This paper presents a multi-target tracking system based on terrestrial mobile platform. The system will use some external parameters of the camera on the mobile platform to construct the ground raster map corresponding to the video stream. The system will use HOG algorithm to detect the target and calculate the occupancy of the ground in the raster map to track the target. In order to guarantee the robustness of multi-target tracking under mobile platform, we also propose a target estimation algorithm based on the change of vanishing point in video sequence. By comparing the results of other algorithms in the same video sequence, it is found that the proposed algorithm is effective.

Key words: multi-target tracking; ground moving platform; probabilistic occupancy map; vanishing point; target estimation

0 引言

多目标跟踪按照摄像头数量, 可分为两大类: 单目跟踪方法与多目跟踪方法。

(1) 单目跟踪方法

传统方法^[1-2]多是在目标区域建立简单目标特征, 但复杂背景大大提高了目标特征的提取难度, 且遮挡鲁棒性低。对此, Chu Chunte^[3]提出多核的目标跟踪算法, 解决了局部遮挡下的目标跟踪; 段沛沛^[4]等人则是通过稀疏算法, 压缩样本数据提高运算速度。

神经网络算法在目标跟踪中同样有较为广泛的应用^[5-7], 但样本与真实目标间的差异往往导致跟踪效果

不佳, 因此 HYEONSEOB N^[8]开始在线调整训练样本。对于目标的运动估计, 通常使用粒子滤波算法, 但重采样过程运算时间较长。对此, Li Peihua^[9]就图像颜色信息重新构建了粒子滤波跟踪算法, POJALA C^[10]将自适应加权算法引入重采样过程。

(2) 多目跟踪方法

在多摄像机情况下, 融合不同摄像机下的光学遮挡结构, 可以有效解决遮挡问题^[11-12], 但无法有效剔除局部的虚假目标。对此, OTSUKA K^[11]引入了递归贝叶斯估计, 而 MITTAL A^[12]通过卡尔曼滤波器来约束目标。UTASI A^[13]与 BENEDEK C^[14]则侧重于多视角下的目标特征提取。前者旨在提取目标像素级特征, 而后者旨在

^{*} 基金项目: 国家自然科学基金资助项目(61701295, 61803254, 61703270)

计算机技术与应用 Computer Technology and Its Applications

构建三维对象模型。FLEURET F^[15]等人首次提出一种概率占用图(POM)模型,计算目标在场景平面网格图上的占用概率,但算法运行时间较长。LIN Y S 等人提出一种根据消失点的快速人员定位方法^[16]。JEROME B 等人利用前后帧目标关系,通过 k-shortest 路径算法快速优化轨迹^[17]。

多目跟踪方法尽管有着测量精度高、目标信息丰富等优点,但高纬度的信息往往会占据较大的运行时间,不利于算法实时性。本文将提出一种基于单目摄像机的背景建模方法,并根据消失点变化情况估补偿目标运动模型。所提出的算法将具有以下特点:(1)准确表现场景三维信息;(2)检测场景变化情况并反馈给目标模型。

1 模型建立

1.1 背景建模

依据透视原理,将实际拍摄过程简化为图 1 所示。集合 $P=(P_1, \dots, P_N)$ 为实际场景中按间隔 μ 排列的点,集合 $l=(l_1, \dots, l_N)$ 为集合 P 在图像中对应点的间距,为画出场景空间模型,需要求解集合 l 中任意相邻两段之间的关系式。其中, C 为摄像机, h 为摄像机对地高度, D 为摄像机拍摄到的最近距离, d 为摄像机焦距。

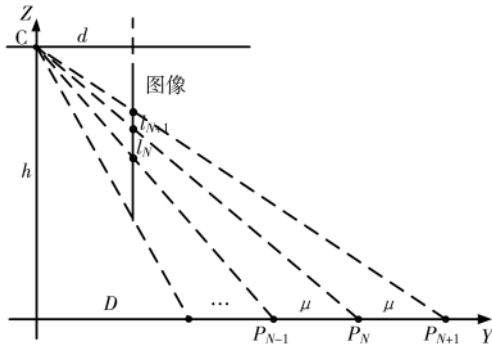


图 1 摄像头拍摄示意图

根据三角形相似原理,可以得到:

$$l_N = dh \left(\frac{1}{D + (N-1) \cdot \mu} - \frac{1}{D + N \cdot \mu} \right) \quad (1)$$

可推得递推公式:

$$\frac{l_{N+1}}{l_N} = \frac{D + (N-1) \cdot \mu}{D + (N+1) \cdot \mu} \quad (2)$$

纵向网格将依靠标定的方法来确定。设转换公式 $f(x, y)$, 可实现栅格图与世界坐标系的转换。

1.2 最优轨迹

用集合 $L^n = (L_s^n, L_{s+1}^n, \dots, L_H^n)$ 表示目标 n 自进入检测区到离开时的所有轨迹,其中 s 表示目标首次出现时的帧数, H 表示目标离开时的帧数,且其中任意时刻的目标轨迹均可以用图像中确定的坐标位置来表示,即 $L_t^n = (x, y)$ 。用 $I = (I_1, \dots, I_N)$ 表示从第 1 帧开始到第 tN 帧的视频序列,则多目标轨迹的计算过程可以表示为 $P(L^1, \dots, L^{NT} | I_1, \dots, I_N)$ 的最优求解过程,其中, NT 表示最大跟踪

目标数量。

1.3 目标建模

FLEURET F^[15]等人提出了一种概率占用计算模型,模型将多摄像头拍摄范围划分为一定数量的格子,并用集合 M 表示所有位置,则 $L_t^N = k, k \in M$ 可以表示为第 t 帧时目标 N 处于位置 k 。算法依据检测结果与跟踪目标之间可靠性函数计算结果来依次计算目标轨迹,在目标所处位置不重叠的约束条件下,第 t 帧时目标 N 处于 k 处的轨迹计算可以表示为:

$$P(L_t^N = k | I, \forall i < N, L_t^i \neq L_t^N) \quad (3)$$

式中, L_t^i 表示当前帧中已经确定的目标轨迹位置。而目标 N 处于 k 处的可靠性函数计算满足:

$$\Psi_i(k) = P(I_t | L_t^N = k) \cdot \max_{\tau} P(L_t^N = k | L_{t-1}^N = \tau) \Psi_{t-1}(\tau) \quad (4)$$

式中, $P(I_t | L_t^N = k)$ 和 $P(L_t^N = k | L_{t-1}^N = \tau)$ 分别表示目标外观模型和目标运动模型。

2 轨迹计算

2.1 目标模型匹配

用颜色直方图特征来表示目标外观模型,则第 t 帧中目标 N 的颜色直方图特征可以表示为 η_t^N ,且处于 k 处的检测结果 $T_{t,k}$ 的颜色直方图特征可以表示为 $\eta_{t,k}$ 。使用欧式距离 $ED()$ 计算直方图之间的匹配相似度,则外观模型的匹配计算可以表示为:

$$P(I_t | L_t^N = k) = ED(\eta_{t-1}^N, \eta_{t,k}) \quad (5)$$

为减少背景影响,算法将如图 2 所示,缩小目标区域,提高匹配准确率。对于目标运动模型的匹配,将依赖于目标在相邻帧中的位置和速度 v_{t-1}^N 来计算可靠性,满足:

$$P(L_t^N = k | L_{t-1}^N = \tau) = \begin{cases} 1/Z \cdot e^{-\rho \|L_t^N - (L_{t-1}^N + v_{t-1}^N)\|} & \|L_t^N - (L_{t-1}^N + v_{t-1}^N)\| \leq c \\ 0 & \text{其他} \end{cases} \quad (6)$$

式中,常数 Z 和 ρ 用以调节结果影响程度,参数 c 则用来与欧几里得范数 $\|\cdot\|$ 结果相比较,划定目标的置信区间,当与上一帧中的目标相距较远时,匹配结果为 0。针对目标丢失以及遮挡情况,算法将基于两方面处理:

(1) 通过运动模型估计目标轨迹:

$$L_t^n \approx L_{t-1}^n + v_{t-1}^n \quad (7)$$

(2) 根据前后帧目标位置,拟合目标轨迹。

2.2 转向变化

设第 t 帧中消失点位置为 p_t ,则消失点位置变化的判断式为:

$$\text{Dis}(p_t, p_{t+2}) \geq \varepsilon \quad (8)$$

式中, ε 为阈值参数,通过 $\text{Dis}(\cdot)$ 计算两点横坐标变化量,当结果大于 ε 时,认为摄像头发生转向变化。

摄像头的转向过程可以视作固定坐标系下目标的



(a) 视频序列的原图



(b) 通过目标检测算法检测到的目标



(c) 改变选区大小后用于提取直颜色方图的区域

图2 目标特征提取示意图

旋转过程。如图3所示,第 t 帧时目标处于 A 处记作 L_t^A ,第 $t+1$ 帧时目标处于 B 处,期间旋转角为 θ ,此时 A 的运动状态满足:

$$L_{t+1}^A \approx (L_t^A + v_t^A) + T(\theta) = \hat{L}_t^A + \hat{v}_t^A \quad (9)$$

图3 旋转示意图

式中, $T(\theta)$ 表示摄像头转向带来的变化,将转向影响补偿到目标后即可得到目标的补偿位置 \hat{L}_t^A 与补偿速度 \hat{v}_t^A 。

2.2.1 目标位置变化

按照图4进行旋转角 θ 的求解, E 和 E' 分别为消失点移动前和移动后的位置,在图像中分别对应 C 和 D 。以相机位置 O 为坐标原点,纵轴为摄像头方向。设消失点 E 和 E' 与原点 O 连线后与纵轴方向的夹角分别为 θ 和 θ' ,以纵轴为 0° 方向,顺时针方向为正方向。则任意位置处的横坐标 X 都可以与夹角 θ 和焦距 d 组成三角关系,满足:

$$\theta = \arctan(kX/d) \quad (10)$$

式中,常数 k 可以实现像素距离到实际距离的转换,可利用图像边缘处的三角关系进行求解,满足:

$$kX_1/d = \tan(\theta_{\max}) \quad (11)$$

$$k = d \tan(\theta_{\max}) / X_1 \quad (12)$$

式中, X_1 为图像横向像素值的一半, θ_{\max} 为摄像头正方向最大拍摄角。利用 C 和 D 处求解出的夹角,即可计算出摄像头旋转角度,满足:

$$\theta_{\text{cam}} = \theta_D - \theta_C \quad (13)$$

且图像中任意情况下都满足:

$$kX_D/d = \tan(\theta_{\text{cam}} + \theta_C) \quad (14)$$

将式(10)和式(12)代入后,得:

$$X_D = X_1 \frac{\tan(\theta_{\text{cam}} + \arctan(\tan(\theta_{\max}) \frac{X_C}{X_1}))}{\tan(\theta_{\max})} \quad (15)$$

式中, X_D 即为旋转后目标在图像中的横坐标。如图3所示,世界坐标系下旋转后的目标纵坐标满足:

$$\tilde{Y}_B = \tilde{D}_{BO} \cos(\theta_A + \theta_{\text{cam}}) = \tilde{D}_{AO} \cos(\theta_A + \theta_{\text{cam}}) \quad (16)$$

式中, \tilde{D}_{AO} 和 \tilde{D}_{BO} 分别表示世界坐标系下 A 位置和 B 位置到摄像头距离,且 \tilde{D}_{AO} 的计算式满足:

$$\tilde{D}_{AO} = \|f(x_A, y_A)\| \quad (17)$$

式中, $\|\cdot\|$ 表示欧几里得范数。将 \tilde{Y}_B 再转换到图像中即可得到目标在图像中的纵坐标 Y_B 。通过上述步骤即可求得摄像头转向情况下任意目标 N 的补偿位置 $\hat{L}_t^N(X, Y)$ 。

2.2.2 目标速度变化

将目标速度 v_t^N 分解为沿着栅格横向纵向两个方向的方向向量 v_X 和 v_Y 。通过转换公式 $f(x, y)$,可以得到目标在世界坐标系下的速度分量 \tilde{v}_X 和 \tilde{v}_Y ,以及方向角 $\tilde{\theta}_v$ 。在旋转 θ_{cam} 之后,速度方向为 $\tilde{\theta}_v + \theta_{\text{cam}}$,并得到旋转后目标所在位置沿栅格方向的速度分量 \tilde{v}_X' 和 \tilde{v}_Y' 。重新转换到像素坐标系下后,得到了目标在摄像头转向情况下的补偿速度 \hat{v}_t^N 。流程示意图如图5所示。

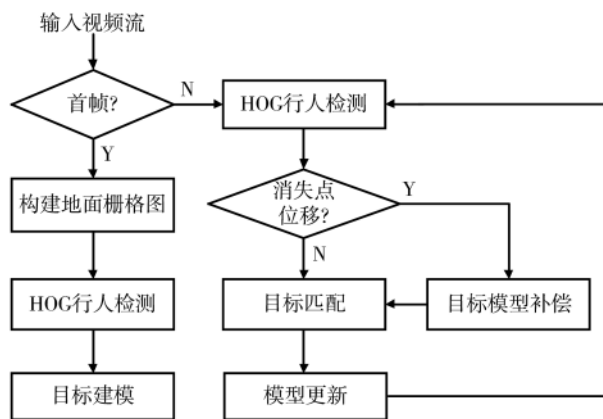


图5 流程示意图

3 实验结果

本文所提出的系统是针对移动场景,对摄像头方向角变化敏感。算法将会在ETH移动场景数据集上进行

《电子技术应用》2020年第46卷第8期

计算机技术与应用 Computer Technology and Its Applications

验证。部分摄像头外参将自行拟定。检测算法方面仅使用了 OpenCV 自带的 HOG 行人检测算法, 实验结果将更倾向于比较算法跟踪性能。实验将与 LEE K H^[18]、JUN K S^[19]所提出的方法进行比较。

所有的实验将在 i7 2.70 GHz CPU 和 4 GB+16 GB DDR 的个人计算机上运行, 所使用的编程环境为 VS 12.0 下配置的 OpenCV 3.3.0。

3.1 评价标准

对于跟踪结果, 将用 Score 重叠分数比率来计算。Rect_T 表示跟踪算法结果, Rect_{GT} 表示测试集标准结果, 定义式如下:

$$\text{Score} = \frac{\text{area}(\text{Rect}_T \cap \text{Rect}_{GT})}{\text{area}(\text{Rect}_T \cup \text{Rect}_{GT})} \quad (18)$$

在参数指标方面, 将有: 多目标跟踪准确度(MOTA)、多目标跟踪精度(MOTP)、目标混淆次数(IDsw)、正确跟踪轨迹占 80% 以上的百分比(MT)、正确跟踪的轨迹占 20% 以上的百分比(ML)。

3.2 参数分析

参数 c 与参数 ε 分别用于判定目标移动置信区间与消失点移动判断。为验证取值, c 将从离散序列 {20, 25, 30, 35, 40} 中选取, ε 将从离散序列 {3, 5, 7, 9} 中选取。数据结果保存在表 1、表 2 中。

通过实验数据, c 的取值为 30, ε 取值为 5。

3.3 定性比较

文献[18]主要改进了贪婪算法, 加快算法收敛。文献[19]则是在 CNN 神经网络的基础上, 将随机蕨引入其

表 1 参数 c 在 MOPT 指标中的作用

参数	数值				
c	20	25	30	35	40
MOPT/%	70.7	73.1	74.2	72.6	69.7

表 2 参数 ε 在 MOPT 指标与处理速度中的作用

参数	数值			
ε	3	5	7	9
MOPT/%	74.0	74.2	74.1	73.6
处理速度/(f/s)	2.82	2.84	2.86	2.89

中, 为每个跟踪目标选择部分最佳蕨类, 提高算法运行效率。为公平比较, 所有算法都将使用 HOG 行人检测作为探测器。图 6 为各个算法在运行过程中的视频截图, 表 3 展示了所有跟踪算法在实验中的运行结果。

与文献[18]、[19]相比, 本文所提出的算法用到了目标 3D 信息, 使得本文提出的算法在目标遮挡以及跟踪精确率上都有更好的表现, 且减少了发生目标之间误报的情况。

3.4 讨论

本文所提出的算法将会进行地面估计, 在运行过程之中, 可以根据检测到的目标作出地面占用估计, 通过这样的一种三维信息估计, 使得算法在处理目标遮挡方面具有一定鲁棒性。此外, 算法会对视频序列中非正常消失的目标单独建立颜色模型, 每检测到一个新目标的时候, 会与先前建立的目标模型比较以确定是否是同一



图 6 视觉跟踪效果

表 3 各算法在 ETHMS 数据集中的表现比较

视频集	方法	MOTA/%	MOTP/%	IDsw	MT/%	ML/%
BAHNHOF	本文算法	44.5	86.2	11	42.5	23.3
	文献[18]	44.0	82.5	13	41.1	23.3
	文献[19]	43.7	79.6	39	26.0	27.4
LOEWENPLATZ	本文算法	42.2	82.3	21	32.9	38.6
	文献[18]	41.4	83.1	17	31.8	39.8
	文献[19]	38.3	75.8	50	20.5	52.2

计算机技术与应用 Computer Technology and Its Applications

个目标。然而,算法还是存在部分限制。首先,算法所采用的探测器仅仅是 OpenCV 自带的 HOG 算法,在准确度上有所不足。其次,为了更好地估算目标,算法要求得到视频序列中的消失点或移动平台的移动变化情况,且针对摄像机的情况需要预先配置部分参数。

4 结论

本文主要提出了一种地面建模方式以及摄像头镜头偏移时的目标补偿方法。在普通摄像机的基础上,通过 HOG 检测目标位置,其次根据地面栅格估计目标 3D 信息,使得算法拥有较好的遮挡鲁棒性。在未检测到跟踪目标的情况下,算法将通过消失点变化情况重新估算目标位置,相较于通过目标前后帧之间的运动状态估算目标位置,拥有更好的鲁棒性,且估算算法运行复杂度较低,可以适用于大部分的平台。

参考文献

- [1] VERMA M, RAMAN B. Local neighborhood difference pattern: a new feature descriptor for natural and texture image retrieval[J]. *Multimedia Tools and Applications*, 2018, 77(10): 11843–11866.
- [2] HERBERT B, ANDREAS E, TINNE T, et al. Speeded-up robust features[J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346–359.
- [3] Chu Chunte, HWANG J N, PAI H I. Tracking human under occlusion based on adaptive multiple kernels with projected gradients[J]. *IEEE Transactions on Multimedia*, 2013, 15(7): 1602–1615.
- [4] 段沛沛, 李辉, 雒明世. 基于联合字典稀疏表达的目标识别算法研究[J]. *电子技术应用*, 2019, 45(4): 18–21.
- [5] 蒋兆军, 成孝刚, 彭雅琴, 等. 基于深度学习的无人机识别算法研究[J]. *电子技术应用*, 2017, 43(7): 84–87.
- [6] Hao Zhaohui, Liu Guixi, Zhang Haoyang. Correlation filter-based visual tracking via adaptive weighted CNN features fusion[J]. *IET Image Processing*, 2018, 12(8): 1423–1431.
- [7] RAMIREZ-QUINTANA J A, CHACON-MURGUIA M I. Self-adaptive SUM-CNN neural system for dynamic object detection in normal and complex scenarios[J]. *Pattern Recognition*, 2015, 48(4): 1137–1149.
- [8] HYEONSEOB N, BOHYUNG H. Learning multi-domain convolutional neural networks for visual tracking[C]. *CVPR*, 2016: 4293–4302.
- [9] Li Peihua. An efficient particle filter-based tracking method using graphics processing unit (GPU)[J]. *Journal of Signal Processing Systems*, 2012, 68(3): 317–332.
- [10] POJALA C, SOMNATH S. Rough-set-theoretic fuzzy cues-based object tracking under improved particle filter framework[J]. *IEEE Transactions on Fuzzy Systems*, 2016, 24(3): 695–707.
- [11] OTSUKA K, MUKAWA N. Multiview occlusion analysis for tracking densely populated objects based on 2-D visual angles[C]. *CVPR*, 2004: 90–97.
- [12] MITTAL A, DAVIS L S. M2tracker: a multi-view approach to segmenting and tracking people in a cluttered scene[J]. *International Journal of Computer Vision*, 2003, 51(3): 189–203.
- [13] UTASI A, BENEDEK C. A 3-D marked point process model for multi-view people detection[C]. *CVPR*, 2011: 20–25.
- [14] BENEDEK C. A bayesian approach on people localization in multicamera systems[J]. *Circuits and Systems for Video Technology*, 2013, 23(1): 105–115.
- [15] FLEURET F, BERCLAZ J, LENGAGNE R, et al. Multi-camera people tracking with a probabilistic occupancy map[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 267–282.
- [16] LIN Y S, LO K H, CHEN H T, et al. Vanishing point-based image transforms for enhancement of probabilistic occupancy map-based people localization[J]. *IEEE Transactions on Image Processing*, 2014, 23(12): 5586–5598.
- [17] JEROME B, FRANCOIS F, ENGIN T, et al. Multiple object tracking using k-shortest paths optimization[J]. *TPAMI*, 2011, 33(9): 1806–1819.
- [18] LEE K H, HWANG J N. Ground-moving-platform-based human tracking using visual SLAM and constrained multiple kernels[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(12): 3602–3612.
- [19] JUN K S, JAE-YEAL N, CHUL K B. Online tracker optimization for multi-pedestrian tracking using a moving vehicle camera[J]. *IEEE Access*, 2018(6): 48675–48687.

(收稿日期: 2019-07-02)

作者简介:

何嘉(1993-), 男, 硕士研究生, 主要研究方向: 计算机视觉。

奚峥皓(1981-), 通信作者, 男, 博士, 讲师, 主要研究方向: 机器视觉、图像处理、路径规划、智能认知学习与控制, E-mail: zhenghaoxi@sues.edu.cn。

阚秀(1983-), 女, 博士, 副教授, 主要研究方向: 智能控制、路径规划、网络化系统建模、多指标滤波/控制/状态估计。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所