

引用格式:李轶锴,吴庆宪,丁晟辉,等.基于TLD和fDSST的长时间目标跟踪算法[J].电光与控制,2019,26(4):44-48,70. LI Y K, WU Q X, DING S H, et al. A long-term tracking algorithm based on TLD and fDSST[J]. Electronics Optics & Control, 2019, 26(4):44-48, 70.

基于TLD和fDSST的长时间目标跟踪算法

李轶锴, 吴庆宪, 丁晟辉, 胡 鲲
(南京航空航天大学自动化学院, 南京 211100)

摘要:为加强fDSST算法在目标快速运动、快速形变、目标消失情况下的跟踪精度,提出了一种基于TLD和fDSST的长时间目标跟踪算法。在fDSST算法的基础上,加入了检测器和学习器对跟踪结果进行修正和学习,并利用检测器和学习器的正负样本对跟踪结果进行置信度评估,从而解决了在跟踪失败情况下的错误参数学习问题。实验表明,基于TLD和fDSST的长时间目标跟踪算法,不但解决了fDSST算法由于目标快速运动、形变甚至消失而使跟踪失败,难以进行长时间持续跟踪的问题,且很大程度上增强了TLD算法的跟踪精度。

关键词: 目标跟踪; 长时间跟踪; TLD; fDSST; 跟踪精度; 再次识别

中图分类号: TP391.41 **文献标志码:** A **doi:**10.3969/j.issn.1671-637X.2019.04.009

A Long-Term Tracking Algorithm Based on TLD and fDSST

LI Yi-kun, WU Qing-xian, DING Sheng-hui, HU Kun

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China)

Abstract: In order to improve the tracking precision of the fDSST algorithm when the target is moving fast, deforms or even disappears, a long-term tracking algorithm based on TLD and fDSST is proposed. Based on the fDSST algorithm, a detector and an online learner are used to modify and learn from the tracking results. To solve the problem of learning wrong parameters when the fDSST algorithm fails, the positive and negative samples of the detector and the learner are used to evaluate the tracking results. The experiment results indicate that the long-term tracking algorithm based on TLD and fDSST solves the problem that the fDSST algorithm fails to track the target for a long time when the target is in fast moving, deforms or even disappears. This approach could also improve the tracking precision of the TLD algorithm.

Key words: target tracking; long-term tracking; TLD; fDSST; tracking precision; re-recognition

0 引言

目标跟踪一直是机器视觉领域的重要研究方向,其可以广泛应用于医疗急救、智能监控、工业自动化及自动驾驶等领域。经过各国学者的深入研究,近几十年跟踪算法层出不穷,从传统的光流法、MeanShift方法、SVM(Support Vector Machine)方法、基于检测和在线学习的方法(如TLD算法^[1])、相关滤波方法(如KCF算法^[2]、DCF算法^[3])及其改进算法^[4],到近年来兴起的基于深度学习的方法^[5],这些方法在跟踪速度、跟踪精度、抗干扰能力和抗遮挡能力方面都各有优缺点。由于算法效果易受到复杂场景、目标外形变化、目标快速运动、遮挡及目标消失等因素影响,因此需进一

步研究长时间目标跟踪算法。

其中,近期兴起的基于深度特征学习的方法^[6]跟踪效果非常好,但计算量大,很难做到对目标实时跟踪。TLD算法是近年来较受欢迎的一种长时间跟踪方法,可有效适应目标外形小幅度变化、目标被部分遮挡等情况。文献[7]用MeanShift算法代替TLD算法中的跟踪器,加快了TLD算法运算速度。文献[8]提出了用卡尔曼滤波器和马尔可夫模型改进TLD算法,提高了算法速度和跟踪精度。文献[9]提出了基于MIL的TLD跟踪算法,提高了算法稳定性和速度。fDSST(fast Discriminative Scale Space Tracking)算法是相关滤波方法的一种,它利用一个二维位置滤波器和一个一维尺度滤波器来实现对目标的跟踪并利用快速傅里叶变换(FFT)加快运算速度^[10]。

本文提出了一种基于TLD和fDSST的长时间目标跟踪算法。利用TLD算法检测器、学习器修正fDSST算法跟踪结果,提高了在目标大幅抖动、外形快速变

收稿日期:2018-03-09

修回日期:2018-05-26

作者简介:李轶锴(1994—),男,河北宁晋人,硕士,研究方向为计算机视觉。

化、快速移动、消失情况下fDSST的跟踪性能。与TLD算法相比,也较大程度提高了跟踪精度。

1 相关算法简介

1.1 TLD算法

TLD算法是由KALAL等提出的一种对未知目标进行长时间跟踪的算法^[1]。该算法分为3个模块:跟踪模块、检测模块和学习模块,算法架构如图1所示。

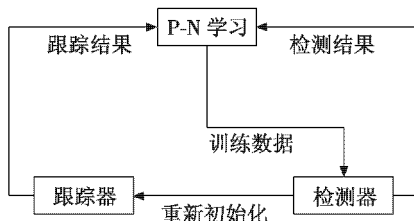


图1 TLD算法架构

Fig.1 Structure of TLD algorithm

跟踪器采用中值流算法,预测物体位于图像序列中位置变化。检测器由3个级联分类器组成,用于区分前景和背景。结合检测器和跟踪器结果,得到目标在图像时间序列中的位置,达到跟踪目的。同时,检测器和跟踪器产生的正负样本将送到学习模块,用于优化跟踪器和检测器,使得算法结果更加稳定。关于TLD算法的详尽原理可参见文献[1]。

1.2 fDSST算法

fDSST算法是由DANELLJAN提出的基于相关滤波器改进的一种短时间跟踪算法^[10]。该算法设计了位置滤波器和尺度滤波器两个滤波器,分别估计目标位置和尺度。同时为加快运算速度,通过对维数压缩、滤波结果插值方式进行加速,在没有牺牲实时效果的前提下,扩大目标检测范围,使得算法的鲁棒性进一步提升。

1.3 fDSST算法和TLD算法的缺陷

fDSST算法是一种短时间跟踪算法,它计算效率高,短时间内跟踪效果非常好,但对于目标短时间内快速形变和快速移动的情况跟踪效果并不好。因为fDSST是一种模板类方法,快速形变的目标使得基于HOG梯度的模板跟踪失败,因此文献[10]先用余弦窗函数对原始样本进行滤波,再用于训练。另外,由于检测器检测区域与运算量呈正相关,因此不可能采取很大的检测区域对目标进行检测,这也使得该算法对于快速移动的目标跟踪效果不好。TLD算法的跟踪模块采用中值流跟踪算法,其本质是两次光流法,这种方式运算量很大,且精度不高,在检测器和学习器的辅助下才可以完成对目标长时间跟踪。因此,本文提出一种基

于TLD和fDSST的长时间跟踪算法来克服这两种算法存在的缺陷。

2 基于TLD框架的fDSST目标跟踪

本文用fDSST算法对TLD算法中的跟踪器进行替换,从而在算法效率略有提升的情况下改进TLD算法的跟踪精度。基于TLD框架的fDSST跟踪算法流程结构如图2所示,首先接收1帧图像,分别送入fDSST跟踪器、检测器和学习模块。其中,fDSST跟踪器处理过程为:视频帧样本送入位置滤波器和尺度滤波器分别预测目标位置和尺度;然后根据初始产生的正负样本对跟踪结果置信度进行评估,若置信度低于某一个阈值则认为跟踪失败;上述对fDSST跟踪器的阈值检测方法避免了滤波器学习错误的图像信息,以免永久性跟踪失败。当检测器重新检测到正确目标之后,重新激活跟踪器,使其继续工作。检测器将视频帧图像由方差分类器筛选前景概率大的区域,经级联随机森林分类器分类后,由KNN分类器进一步检验输出目标的检测结果。视频的正负样本图像与跟踪、检测结果传送到学习模块学习训练,用于对跟踪器和检测器结果的评估。最后比较跟踪结果与检测结果的置信度,得出目标位置。

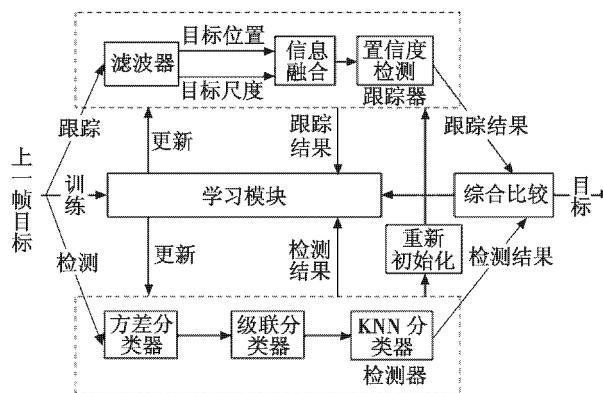


图2 改进算法流程结构

Fig.2 Structure of the improved algorithm

本文算法以fDSST算法为基础,短时期内跟踪效果非常好,跟踪精度高,又有检测器的检测结果与其融合,对比置信度、方差等性能指标,修正和加强跟踪器的结果。且当目标消失时,停止训练滤波器参数,避免学习错误的目标信息。当检测器重新检测到目标时,重新激活跟踪器,让其重新工作,从而解决了再次识别的问题。

2.1 fDSST跟踪器

DSST(Discriminative Scale Space Tracking)算法^[10]设计了位置和尺度两种滤波器分别用来预测目标位置和尺度。fDSST算法是DSST算法的加速算法,通过对滤

波结果的插值、维数压缩进行加速。加速后的算法具有很高的帧率 (Frames Per Second, FPS)。

fDSST 算法利用循环矩阵方法由第 1 帧选定目标生成一系列训练样本,用于训练最优相关滤波器估计下一帧中目标的位置。并选用了多通道 HOG^[11] 特征,训练多通道相关滤波器,加强了算法的跟踪精度。fDSST 算法原理如下所述。

首先,设初始图像中选定的正样本为 f , f 由 d 维特征向量 $f(n) \in \mathbf{R}^d$ 组成, d 为选取特征的维数, $f^l, l \in \{1, 2, \dots, d\}$ 表示第 l 通道的特征。每个通道对应的滤波器为 h^l , 总的位置滤波器表示为 h 。文献[10]选取以正样本中心为高斯函数中心的二维高斯函数作为期望输出样本 g , 使得

$$\varepsilon = \left\| g - \sum_{l=1}^d h^l * f^l \right\|^2 + \lambda \sum_{l=1}^d \| h^l \|^2 \quad (1)$$

最小。式中: $*$ 表示卷积; λ 表示一个权重系数, 引入这一项的目的一方面是避免求解滤波器频域参数过程中分母为零, 另一方面也可避免滤波器过拟合。

为加速运算, 可将其转换到复频域采用帕萨瓦尔定理求解得到

$$H^l = \frac{\bar{G}F^l}{\sum_{k=1}^d \bar{F}^k F^k + \lambda} \quad l = 1, 2, \dots, d \quad (2)$$

式中, H^l, F^l, G 为 h^l, f^l, g 经过离散傅里叶变换 (DFT) 得到的对应变量, \bar{G} 表示 G 的共轭转置。

为了保证滤波器 h^l 的鲁棒性, 将不同时间片的样本加入到对滤波器的训练中, 于是式(1)中的误差变为统计 f_1, \dots, f_t 的平均值, 其中 t 为时间。由文献[12]可知, 线性最小二乘问题可以由 DFT 进行块对角化, 于

是通过解决 N 个 $d \times d$ 的线性系统可以解得滤波器 H , 其中 N 为滤波器 h^l 的个数。但这种方法的计算量太大, 于是采用一个训练样本来更新滤波器参数, 为保证鲁棒性能, 采用

$$\begin{cases} A_t^l = (1 - \eta)A_{t-1}^l + \eta \bar{G}F_t^l & l = 1, 2, \dots, d \\ B_t = (1 - \eta)B_{t-1} + \eta \sum_{k=1}^d \bar{F}_t^k F_t^k \end{cases} \quad (3)$$

来更新。式中: A_t^l 和 B_t 为滤波器 H_t^l 对应的一个训练样本 f_t 的分子和分母; η 为学习率。

设 z_t 为中心对应目标中心位置的图像片, Z_t^l 为其经离散傅里叶变换得到的变量, 即选定的正样本区域, 则输出 y_t 的复频域表示形式为

$$Y_t = \frac{\sum_{l=1}^d \bar{A}_{t-1}^l Z_t^l}{B_{t-1} + \lambda} \quad (4)$$

y_t 可以根据离散傅里叶反变换 (IDFT) 求得。位置滤波器估计目标位置的思路如图 3 所示。

尺度滤波器是为适应目标尺度变化所设计的一维滤波器, 用于估计目标尺度, 为构造训练样本 $f_{t, \text{scale}}$, 用图像中心大小不一的图像片来分离特征, 设 $P \times R$ 为当前帧目标大小, S 为尺度滤波器大小, 对于每个 $n \in \left\{ \left[-\frac{S-1}{2} \right], \dots, \left[\frac{S-1}{2} \right] \right\}$, 分离出一个以目标为中心、大小为 $a^p P \times a^r R$ 的图像片 I_n , 其中, a 为尺度系数, 用于控制图像片的大小变化。尺度滤波器采用和位置滤波器一样的设计方法, 不同的是采用一维高斯函数作为期望输出, 推导尺度滤波器并利用 DFT 加速, 最后用测试样本通过求得的滤波器可以得到期望的最佳尺度。尺度滤波器估计目标尺度的思路如图 3 所示。

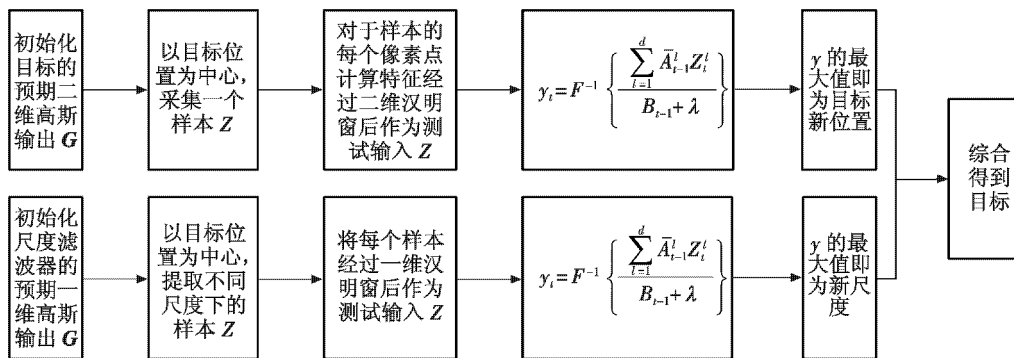


图 3 fDSST 位置滤波器和尺度滤波器目标位置估计示意图

Fig. 3 Illustration of position filter and scale filter in fDSST

2.2 检测器

检测器采用 3 个分类器级联而成, 结构如图 4 所示。首先在待检测图像片上产生一系列滑动窗口, 滑动窗

口将经过 3 个分类器的筛选, 区分其为正样本还是负样本。方差分类器采用积分图^[13]的方式筛选高方差区域, 确定目标出现概率较大的滑窗。级联分类器由随

机分类器^[14]级联而成,分类速度快,经过 Boosting^[15]之后可将一些弱分类器变成一个强分类器,达到较好的分类效果。KNN(K-Nearest Neighbors)分类器^[16]计算图像片与K个训练样本之间的相似度,判断其为正样本还是负样本。

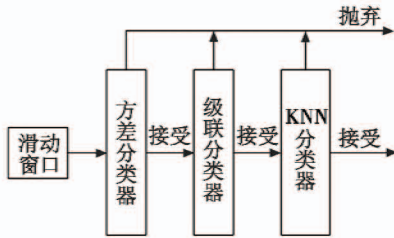


图4 检测器示意图

Fig.4 Illustration of detector

2.3 学习模块

P-N学习模块有两个“专家”,即P-expert和N-expert,分别负责选出错误的正、负样本。结合跟踪模块、检测模块结果产生有效训练样本,完成对检测器的更新,消除检测误差,加强算法的鲁棒性。学习器结构如图5所示,更多P-N学习的内容可参考文献[17]。

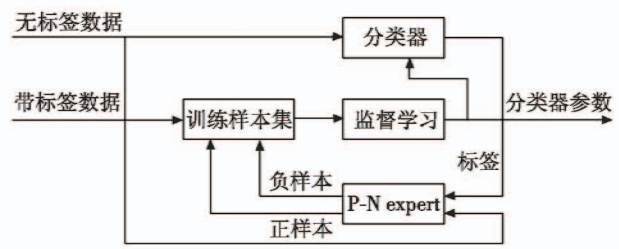


图5 P-N学习器示意图

Fig.5 Illustration of P-N learner

3 实验结果及分析

将本文算法与TLD算法和fDSST算法进行了对比实验。在CPU为i5-3210M 2.5 GHz,4 GiB内存的笔记本电脑上,使用Linux系统进行实验,对上述fDSST算法、TLD算法和本文算法进行对比验证。

3.1 测试视频及说明

本文将基于TLD算法和fDSST算法的长时间跟踪算法在以下严峻情况下进行测试,实验使用了OTB50数据集和文献[1]使用的数据集对3种算法的跟踪效果进行测试,如图6所示。



图6 跟踪目标测试截图

Fig.6 Snapshots of test video sequences

3.2 测试结果及分析

本文采用3种评价指标对以上方法的跟踪结果进行评估,分别为:1) OP(Overlap Precision)表示预测跟踪结果与人为标定结果的IOU(Intersection Over Union)大

于0.5的帧数占总视频帧数的百分比;2) DP(Distance Precision)表示预测跟踪结果与人为标定结果的中心像素差值小于20的帧数占总视频帧数的百分比;3) FPS表示每秒运行多少帧,用来评估算法跟踪速度。

表1 测试视频跟踪精度结果

Table 1 Result of tracking precision in test video

视频属性	光照变化	尺度变化	遮挡	形变	运动模糊	快速运动	面内旋转	面外旋转	超出视野	背景混杂	低分辨率	ALL	
fDSST算法	69.83	60.59	65.04	55.49	60.76	57.03	69.74	66.11	57.40	65.22	56.54	65.86	
OP	TLD算法	39.39	55.19	52.59	43.16	53.39	50.69	49.66	45.36	57.35	43.74	52.04	
	本文算法	74.87	72.01	73.42	69.44	69.45	69.37	72.48	72.42	74.43	73.57	73.92	75.68
DP	fDSST算法	72.59	64.20	68.89	59.36	64.71	61.17	73.07	69.60	61.14	68.79	60.05	69.28
	TLD算法	43.65	58.53	56.11	46.41	56.99	53.93	53.17	49.35	60.77	47.38	66.37	55.34
	本文算法	77.04	74.72	76.22	72.52	72.63	72.59	75.00	74.95	77.40	76.25	76.32	78.28

3.2.1 跟踪精度

本文选取的评价指标中,OP和DP是表征算法跟踪精度的指标,3种算法在不同视频属性下的跟踪精度结果如表1所示。其中最后一列表示在整体测试集上的测试精度,可以看出本文算法跟踪精度远高于改

进前两种算法的跟踪精度,且fDSST算法的跟踪精度远高于TLD算法的跟踪精度。将每种算法在某种情况下跟踪精度远低于其自身在整体测试集上测试精度的OP和DP值加粗表示,可以看出,fDSST算法在目标大尺度变化、形变、运动模糊以及快速运动、超出视野

范围及低分辨率情况下跟踪精度大大下降,而 TLD 算法在光照变化、目标形变、平面外旋转、背景混杂情况下跟踪精度低于整体值,和前文分析相符合,而本文算法将两者优势互补,大大提高了整体跟踪精度。图 7 选取了 2 个视频比较 3 种算法的结果。

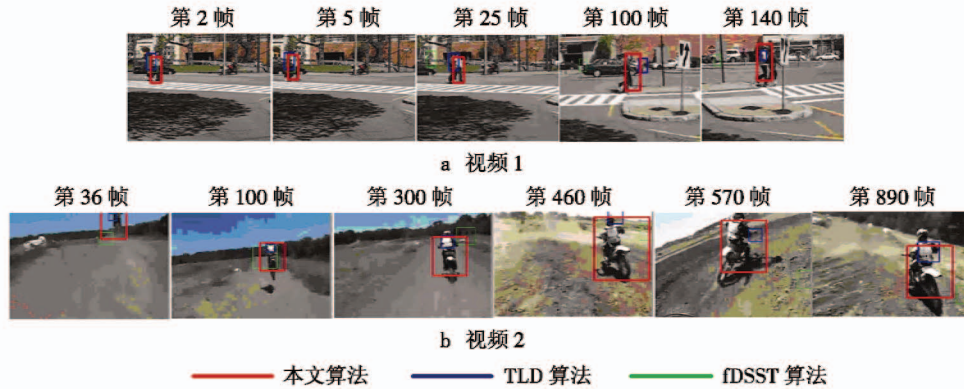


图 7 三种算法实验结果比较

Fig. 7 Comparison of the experimental results of the three algorithms

3.2.2 跟踪速度

对上述测试集测试得到的跟踪速度为: fDSST 算法为 32.27 帧/s; TLD 算法为 10.74 帧/s; 本文算法为 15.66 帧/s。

可以看出, fDSST 算法的处理速度最快, 本文算法次之, 传统 TLD 算法最慢, 由于 TLD 算法的跟踪器采用两次光流法, 属于特征点跟踪方法, 处理速度较慢, 且光流法的实时性本来就很差, 计算量大。本文算法将其替换为 fDSST 算法, 减少了计算量, 加快了处理速度。

4 结束语

本文针对 TLD 算法和 fDSST 算法的局限性提出基于 TLD 算法和 fDSST 算法的长时间跟踪方法, 在保证短时间跟踪效果和速度的前提下, 由检测器和学习器的结果对跟踪结果进行修正得出置信度最高的结果。经实验证明, 本文算法在光照变化、目标大面积遮挡、快速移动、形变甚至消失等严峻情况下均可完成对任务目标的跟踪。

参考文献

- [1] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [2] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [3] DANELLJAN M, HAGER G, KHAN F S, et al. Accurate

经实验发现, TLD 在测试时跟踪器正确处理的帧数不多, 多依靠检测器重新初始化跟踪器完成跟踪, 而 fDSST 算法在跟踪失败之前的视频帧均可正确处理, 说明 fDSST 算法只是需要一个错误更正机制, 所以本文算法在跟踪精度上大有提高。

- scale estimation for robust visual tracking [C]//Proceedings of British Machine Vision Conference, 2014: 1-11.
- [4] 忽晓伟, 陈娟. 融合颜色特征的核相关滤波器目标跟踪[J]. 电光与控制, 2017, 24(6): 43-46.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [6] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking [C]//European Conference on Computer Vision, 2016: 472-488.
- [7] 肖庆国, 叶庆卫, 周宇, 等. 基于 Mean-Shift 优化的 TLD 视频长时间跟踪算法[J]. 计算机应用研究, 2015, 32(3): 925-928.
- [8] 周鑫, 钱秋朦, 叶永强, 等. 改进后的 TLD 视频目标跟踪方法[J]. 中国图象图形学报, 2013, 18(9): 1115-1123.
- [9] 高帆, 吴国平, 刑晨, 等. TLD 目标跟踪算法研究[J]. 电视技术, 2013, 37(11): 70-74, 202.
- [10] DANELLJAN M, HAGER G, KHAN F S, et al. Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [11] WANG X Y, HAN T X, YAN S C. An HOG-LBP human detector with partial occlusion handling [C]//IEEE 12th International Conference on Computer Vision, 2009: 32-39.
- [12] GALOOGAHI H K, SIM T, LUCEY S. Multi-channel correlation filters [C]//Proceedings of the IEEE International Conference on Computer Vision, 2013, 3072-3079.

(下转第 70 页)

器后只需相应增减方程中的量测值,重新设计迭代方程,为全源导航打下了基础。

参 考 文 献

- [1] 雷浩然. 弹载 SINS/GNSS 组合导航系统研究[D]. 南京:南京理工大学,2014.
- [2] 管斌. MIMU/GPS/磁强计组合导航技术研究[D]. 长沙:国防科学技术大学,2010.
- [3] GREWAL M S, WEILL L R, ANDREWS A. Global positioning systems, inertial navigation, and integration[M]. New York:John Wiley & Sons, 2007.
- [4] SMITH D, SINGH S. Approaches to multisensor data fusion in target tracking: a survey[J]. IEEE Transactions on Knowledge & Data Engineering, 2006, 18(12):1696-1710.
- [5] 陈颖,马忠孝,贺峻峰. 全源定位与导航技术发展概况和应用展望[C]//中国惯性技术学会第七届学术年会,2015:271-273.
- [6] 王慧哲. 基于多信息融合的无人机全源导航关键技术研究[D]. 南京:南京航空航天大学,2017.
- [7] KUMMERLE R, GRISETTI G, STRASDAT H, et al. G2o: a general framework for graph optimization[C]//IEEE International Conference on Robotics and Automation, 2011:3607-3613.
- [8] WANG Y, LI X. Graph-optimization-based ZUPT/UWB fusion algorithm[J]. International Journal of Geo-Information, 2018, 7(1):18-40.
- [9] KSCHISCHANG F R, FREY B J, LOELIGER H A. Factor graphs and the sum-product algorithm[J]. IEEE Transactions on Information Theory, 2001, 47(2):498-519.
- [10] TA D N, KOBILAROV M, DELLAERT F. A factor graph approach to estimation and model predictive control on unmanned aerial vehicles[C]//International Conference on Unmanned Aircraft Systems, 2014:181-188.
- [11] 王慧哲,曾庆化,刘建业,等. 基于因子图的无人机全源导航关键技术研究[J]. 导航与控制,2017,16(2):1-5.
- [12] CHIU H P, WILLIAMS S, DELLAERT F, et al. Robust vision-aided navigation using sliding-window factor graphs[C]//IEEE International Conference on Robotics & Automation, 2013:46-53.
- [13] FALLON M F, KAESS M, JOHANSSON H, et al. Efficient AUV navigation fusing acoustic ranging and side-scan sonar[C]//Proceedings of IEEE International Conference on Robotics and Automation, 2011:2398-2405.
- [14] GRISETTI G, KUMMERLE R, STACHNISS C, et al. A tutorial on graph-based SLAM[J]. IEEE Intelligent Transportation Systems Magazine, 2010,2(4):31-43.
- (上接第 48 页)
- [13] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001:1-9.
- [14] BOSCH A, ZISSERMAN A, MUNOZ X. Image classification using random forests and ferns[C]//IEEE 11th International Conference on Computer Vision, 2007:1-8.
- [15] OZA N C. Online bagging and boosting[C]//International Conference on Systems, Man and Cybernetics, 2005:2340-2345.
- [16] ZHANG H, BERG A C, MAIRE M, et al. SVM-KNN: discriminative nearest neighbor classification for visual category recognition[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006:2126-2136.
- [17] KALAL Z, MATAS J, MIKOLAJCZYK K. P-N learning: bootstrapping binary classifiers by structural constraints[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2010:49-56.