

卫星视频中的单目标分割和跟踪^①



王丽黎, 张 慧

(西安理工大学 自动化与信息工程学院, 西安 710048)

通信作者: 张 慧, E-mail: 2359498987@qq.com

摘 要: 对于卫星视频图像中存在的目标与背景对比性低、缺乏目标特征信息等问题, 提出一种结合目标运动信息、时空背景和外观模型的目标分割和跟踪方法. 根据首帧定位得到目标区域, 首先对目标使用方向梯度直方图方法提取特征利用核相关滤波器得到目标跟踪区域 1; 接着利用颜色空间特征建立目标与其周围区域上下文信息的空间模型得到目标跟踪区域 2; 然后利用视觉背景提取算法以像素为单位在目标区域上检测运动目标得到单目标的分割区域 3; 最后分别对 3 个区域进行相关计算得到最优区域作为最终目标跟踪位置和模板更新样本. 实验结果表明, 本文算法与 KCF 算法相比, 跟踪的成功率和准确率有很大的提高, 同时实现了单目标分割.

关键词: 目标跟踪; 目标分割; 核相关滤波; 视觉背景提取; 时空上下文; 特征提取

引用格式: 王丽黎, 张慧. 卫星视频中的单目标分割和跟踪. 计算机系统应用, 2023, 32(2): 406-411. <http://www.c-s-a.org.cn/1003-3254/8928.html>

Single Target Segmentation and Tracking in Satellite Video

WANG Li-Li, ZHANG Hui

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: To address the problems of low contrast between target and background and lack of target feature information facing satellite video images, this study proposes a target segmentation and tracking method combining target motion information, spatio-temporal background, and appearance model. After the target area is obtained by positioning in the first frame, the histogram of oriented gradient method is employed to extract the features of the target, and the kernel correlation filter (KCF) is utilized to obtain the target tracking area 1. Subsequently, color and spatial features are used to build a spatial model of the context information about the target and its surrounding area and thereby obtain the target tracking area 2. Then, the visual background extraction algorithm is applied to detect the moving target in the target area in pixels and further obtain the segmentation area 3 of the single target. Finally, the correlation of the three areas is calculated, respectively, to obtain the optimal area as the final target tracking position and the template update sample. The experimental results show that compared with the KCF algorithm, the proposed algorithm obtains a significantly higher tracking success rate and accuracy and also achieves single target segmentation.

Key words: target tracking; target segmentation; kernel correlation filtering; visual background extraction; spatio-temporal context; feature extraction

随着遥感卫星技术的发展, 通过视频卫星实现对目标的检测和跟踪在智能交通、自然灾害监测和军事侦察等方面具有重要的作用. 近年来, 基于相关滤波的跟踪方法凭借较快的计算速度和较高的精度成为目标

跟踪领域的主流算法之一. Bolme 等^[1] 首先将相关滤波方法用在跟踪领域, 提出了最小平方误差输出和 (minimum output sum of squared error, MOSSE) 滤波算法. Henriques 等^[2] 利用循环移位进行样本采样在

① 收稿时间: 2022-06-20; 修改时间: 2022-07-18; 采用时间: 2022-08-08; csa 在线出版时间: 2022-09-23

CNKI 网络首发时间: 2022-11-15

MOSSE 的基础上提出了核循环结构 (circulant structure of tracking-by-detection with kernels, CSK) 算法. Danelljan 等^[3] 在 CSK 的基础上利用颜色空间 (color name, CN) 特征, 提高了算法有效性. Henriques 等^[4] 将 CSK 算法的灰度特征换成方向梯度直方图 (histogram of oriented gradient, HOG) 特征, 提出了核相关滤波 (kernel correlation filter, KCF) 跟踪算法. 上述算法忽略了目标空间关系, Zhang 等^[5] 通过贝叶斯框架对目标及其周围区域建立时空关系提出了时空上下文 (spatio-temporal context, STC) 算法.

近年来, 目标跟踪算法主要是针对传统视频中运动目标进行研究, 而针对卫星视频中运动目标的研究较少. Shao 等^[6] 在核相关滤波算法中使用光流和梯度直方图作为特征进行目标跟踪; Xuan 等^[7] 先利用卡尔曼预测目标区域, 再进行核相关滤波操作减少边界效应问题; Wang 等^[8] 在 CSK 算法的基础上加入 Gabor 滤波器增加目标的特征实现更稳健的跟踪; 韩鸣飞等^[9] 以 KCF 算法为框架, 融合目标灰度特征和 HOG 特征提高目标跟踪的鲁棒性; Zhang 等^[10] 在相关滤波算法的基础上加入卷积网络预测模型实现目标跟踪.

对于卫星视频运动目标存在相对于背景尺寸很

小、目标缺少清晰的特征信息的问题. 本文结合目标的运动信息、时空背景和外观模型进行跟踪, 以 KCF 算法为基本框架增加了基于颜色特征的时空上下文模型分别训练两个跟踪器, 并且加入了改进的基于运动目标像素的检测分割算法, 同时得到 3 个目标候选区域对目标进行最终相关计算得到目标最优位置.

1 目标分割和跟踪方法

本文分别利用 HOG 特征、CN 特征和像素值信息训练得到独立的两个跟踪器和一个检测分割器. 首先提取目标的 HOG 特征利用岭回归方法和核函数训练得到第 1 个目标跟踪器; 接着利用 CN 特征对目标及其局部背景区域的上下文信息建立空间模型得到第 2 个目标跟踪器; 然后利用首帧图像初始化样本模型, 对候选区域进行像素计算得到检测分割器; 最后根据两个目标跟踪器和检测分割器分别得到两个目标跟踪区域和一个检测分割区域. 实验发现, 利用 3 个候选区域进行特征融合得到的结果没有分别计算响应值进行最大值比较得到目标位置的效果好, 于是本算法利用目标的互补信息对 3 个候选区域分别求取最大响应值, 将得到的 3 个最大响应值进行比较得到最优目标位置. 如图 1.

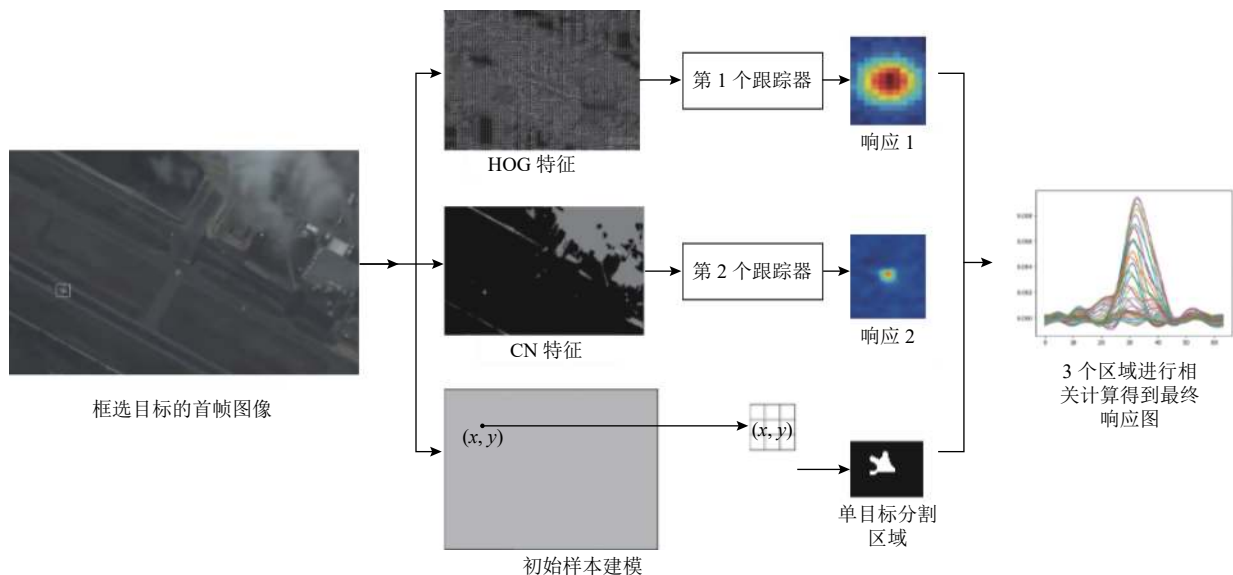


图 1 目标预测流程图

1.1 第 1 个目标跟踪器

根据首帧框选的目标区域提取 HOG 特征, 进行循环移位得到大量正负样本训练非线性回归器, 然后通过相关计算得到响应矩阵的最大值位置, 此位置就是

目标中心位置. 将样本使用岭回归方法对函数 $f(z) = w^T z$ 进行训练, 以获得参数 w :

$$\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2 \quad (1)$$

其中, x_i 为循环移位样本, y_i 表示满足高斯分布的样本输出, λ 是正则化参数.

对式 (1) 求导, 并扩展到复数域得: $w = (X^H X + \lambda I)^{-1} X^H y$. 将循环矩阵 X 在傅里叶空间中角化, 矩阵 X 可写成 $X = F \text{diag}(\hat{x}) F^H$, 由 w 和 X 联合求得:

$$\hat{w} = \frac{\hat{x} \odot \hat{y}}{\hat{x} \odot \hat{x}^* + \lambda} \quad (2)$$

其中, \hat{x} 和 \hat{y} 是原始矩阵的傅里叶变换.

通过引入高斯核函数, 将原始非线性问题转换成线性问题, 并在傅里叶域转换得到核函数回归解:

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (3)$$

根据训练好的滤波器对目标窗口区域进行滤波计算, 得到输出置信图.

$$\hat{f}(z) = \hat{k}^{xz} \odot \hat{\alpha} \quad (4)$$

其中, k^{xx} 和 k^{xz} 均按式 (5) 计算:

$$k^{xx'} = \exp\left(-\frac{1}{\sigma^2} (\|x\|^2 + \|x'\|^2 - 2F^{-1}(\hat{x}^* \odot \hat{x}'))\right) \quad (5)$$

完成视频一帧的目标跟踪后, 需要更新跟踪器的模板参数 α 和样本参数 x , 即:

$$\alpha = (1 - \varepsilon)\alpha + \varepsilon \text{new}_\alpha \quad (6)$$

$$x = (1 - \varepsilon)x + \varepsilon \text{new}_x \quad (7)$$

1.2 第 2 个目标跟踪器

由于卫星视频中目标只占少量像素, 很少的目标信息无法实现良好的跟踪. 于是我们利用目标及局部区域背景的 CN 颜色特征建立空间关系增强跟踪器的性能. 将首帧目标区域与其局部区域背景建立一个空间上下文模型, 对图像进行卷积操作, 获得置信图, 置信图最大的位置作为最佳目标位置.

假定当前帧目标位置为 x^* , 则图像特征集合为 $X^c = \{c(z) = (I_{c_i}(z), z) | z \in \Omega_c(x^*)\}$, $I_{c_i}(z)$ 表示位置 z 处的图像对应的 11 维 CN 特征, $\Omega_c(x^*)$ 是目标的局部上下文区域. 则目标的置信图公式为:

$$c(x) = P(x|o) = \sum_{c(z) \in X^c} P(x, c(z)|o) = \sum_{c(z) \in X^c} P(x|c(z), o) P(c(z)|o) \quad (8)$$

其中, $P(x|c(z), o)$ 是对目标与其上下文信息的空间关系建模的条件概率, $P(c(z)|o)$ 是对各个点 z 的局部上下文信息建模的先验概率, 具体形式分别为式 (9) 和式 (10):

$$P(x|c(z), o) = h^{sc}(x - z) \quad (9)$$

$$P(c(z)|o) = I_{c_i}(z) w_\sigma(z - x^*) \quad (10)$$

其中, $h^{sc}(x - z)$ 是目标位置与局部区域点之间的相对距离和方向关系的函数, $w_\sigma(z) = a e^{-\frac{|z|^2}{\sigma^2}}$ 是一个离中心越近权重越大的加权函数.

目标置信图描述了区域各点为目标的概率大小, 公式定义为:

$$c(x) = P(x|o) = b e^{-\left|\frac{x-x^*}{\alpha}\right|^\beta} \quad (11)$$

其中, b 为归一化常数; β 为形状控制参数; α 为尺度参数.

结合上下文先验模型和置信图公式得到时空上下文模型表达式为:

$$h_{t,c_i}^{sc}(x) = \mathcal{F}^{-1} \left(\frac{\mathcal{F} \left(b e^{-\left|\frac{x-x^*}{\alpha}\right|^\beta} \right)}{\mathcal{F} \left(I_{c_i}(x) w_\sigma(x - x^*) \right)} \right) \quad (12)$$

然后更新得到 $t+1$ 帧的时空上下文模型为:

$$H_{t+1,c_i}^{stc} = (1 - \rho) H_{t,c_i}^{stc}(x) + \rho h_{t,c_i}^{sc}(x) \quad (13)$$

其中, H_{t+1,c_i}^{stc} 和 $H_{t,c_i}^{stc}(x)$ 分别为 $t+1$ 帧和 t 帧基于 CN 特征对应的时空上下文模型, ρ 为模型更新率.

利用得到的模型与图像进行卷积得到目标置信图, 目标在 $t+1$ 帧的位置由置信图中最大值决定, 即式 (14):

$$x_{t+1}^* = \arg \max_{x \in \Omega_c(x_t^*)} c_{t+1}(x) = \arg \max_{x \in \Omega_c(x_t^*)} \mathcal{F}^{-1} \left(\sum_{i=0}^{11} \mathcal{F} \left(H_{t+1,c_i}^{stc}(x) \right) \odot \mathcal{F} \left(\tilde{I}_{t+1,c_i}(x) w_\sigma(x - x_t^*) \right) \right) \quad (14)$$

其中, \tilde{I}_{t+1,c_i} 分别对应第 $t+1$ 帧相互独立的 11 维 CN 特征.

1.3 目标检测分割

考虑到卫星视频中背景相对静止, 于是利用改进的 ViBe 算法对运动目标检测分割. ViBe 算法^[11]是一种像素级前景检测算法, 不关注运动目标的特征信息, 所以此算法对于卫星视频中的运动目标检测有较好的效果. 由于 ViBe 算法是通过初始化背景模型, 对每一帧图像的所有像素值进行差值计算得到动态区域, 导致检测速度较低、容易产生鬼影现象, 针对原 ViBe 算法存在的问题, 本文利用视频帧的时间关系, 对每一帧的目标区域与背景模型相应位置进行像素值计算, 实现单目标的检测分割, 从原理上降低了 ViBe 算法产生的鬼影、动态背景干扰问题, 提高了运动目标检测分割的效果和速度.

改进的 ViBe 算法利用视频首帧图像进行背景建模,对首帧图像的每个像素点及其 8 个相关邻域随机选取 20 个像素组成该像素的背景模型,通过循环执行此步骤,将首帧图像的所有像素值完成建模操作。 x 像素点处的背景样本集可以表示为:

$$M(x) = \{p_1, p_2, \dots, p_N\} \quad (15)$$

其中, p_1, p_2, \dots, p_N 表示每一层样本的 x 点对应的像素值。

然后将每帧的目标区域代替原始整张图像与背景模型对应的位置进行像素值计算,设定半径阈值 R 和匹配阈值 A ,将像素差值与半径阈值 R 进行对比,如式 (16) 所示,将大于阈值 R 的进行计数,数值超过匹配阈值 A ,则表示像素变化较大,定义该点为前景像素;数值小于匹配阈值 A ,则表示像素变化较小,定义该点为背景像素,如式 (17) 所示。

$$abs(p - p_i) \geq R \quad (16)$$

$$\{S_R(p(x)) \cap \{p_1, p_2, \dots, p_N\}\} \geq A \quad (17)$$

将检测区域判断为背景的每个像素点有 $1/\varphi$ 的概率更新自己及其邻域的背景模型样本值。其中像素点连续 B 次被检测为前景,则将其更新为背景点。

1.4 最终目标位置

本文利用两个跟踪器和检测分割器得到独立的 3 个候选区域,对候选区域进行相关计算分别得到对应候选区域的最大响应值 r_1, r_2, r_3 ,选取其中的最大值 r_{\max} ,得到最终目标位置。如式 (18) 所示。

$$r_{\max} = \max(r_1, r_2, r_3) \quad (18)$$

2 实验结果与分析

2.1 实验环境及配置

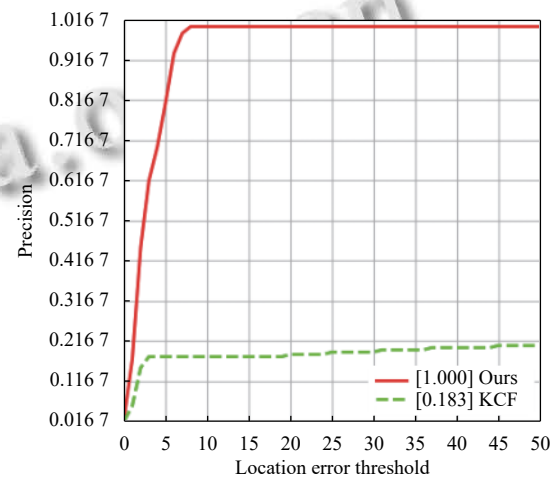
本文算法的实验是在 i5-6500 CPU,主频 3.2 GHz,内存 4 GB 的台式电脑上进行的,实现本文算法的平台为 PyCharm。实验中所用参数参考 ViBe 算法设置:半径阈值 R 为 20,匹配阈值 A 为 2,随机因子 φ 为 16, B 为 50。选取 3 组 SatSOT^[12] 卫星视频序列数据集作为本文实验的测试序列。为了验证本文算法的鲁棒性和有效性,本文采取成功率和准确率作为跟踪评价标准,将 KCF 算法与本文算法进行对比。

2.2 算法结果比较分析

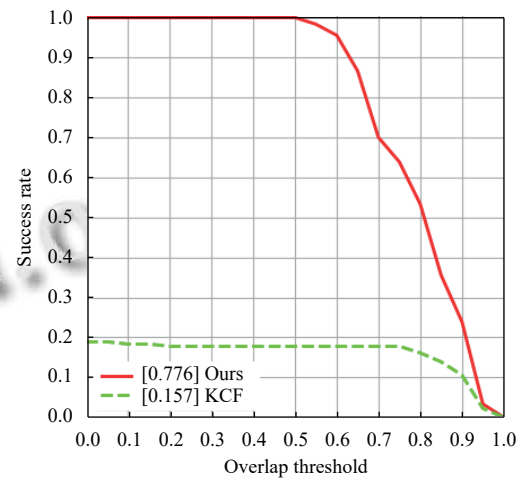
2.2.1 定量分析

为了验证算法的跟踪性能,在 3 组卫星视频数据集上测试了 KCF 算法和本文算法。本文算法在 plane_

03 序列中的准确率和成功率分别为 1 和 0.776, KCF 算法在 plane_03 序列中的准确率和成功率分别为 0.183 和 0.157; 本文算法在 plane_06 序列中的准确率和成功率分别为 1 和 0.739, KCF 算法在 plane_06 序列中的准确率和成功率分别为 0.097 和 0.073; 本文算法在 ship_01 序列中的准确率和成功率分别为 1 和 0.443, KCF 算法在 ship_01 序列中的准确率和成功率分别为 0.505 和 0.152。如图 2、图 3 和图 4 所示。



(a) 准确率



(b) 成功率

图 2 plane_03 序列中的准确率和成功率对比图

2.2.2 定性分析

为了更直观评估算法的目标分割效果,将本文算法与 ViBe 算法在 plane_03 和 plane_06 视频序列上进行了对比实验,原 ViBe 算法只能对整幅图像进行检测分割,存在严重的鬼影和动态背景干扰问题,改进算法大大缓解了鬼影和动态背景干扰问题,而且实现了对单个目标的检测分割。如图 5 和图 6 所示。

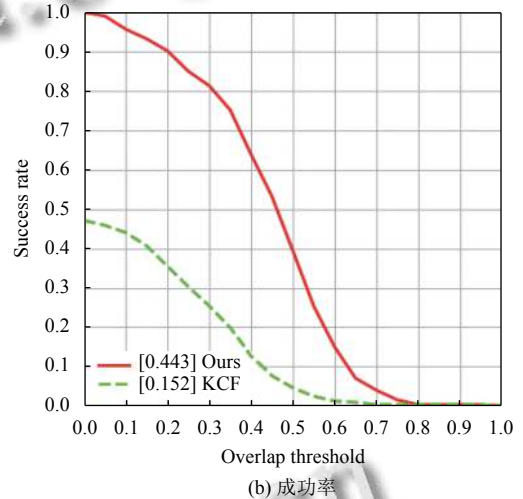
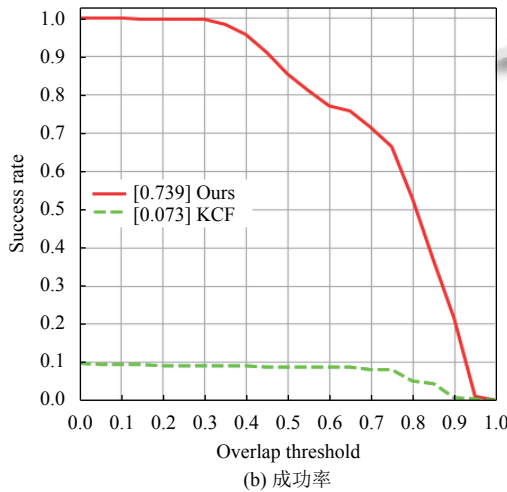
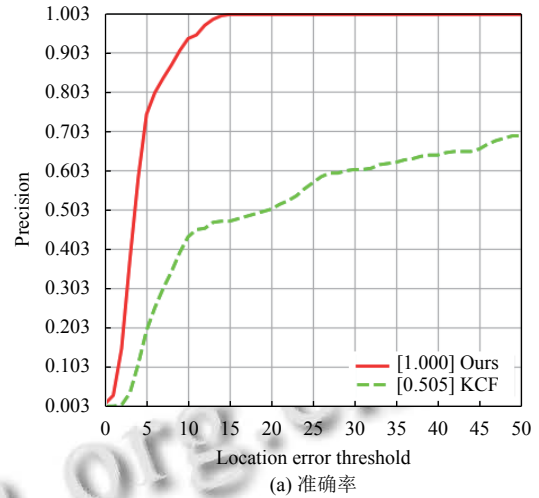
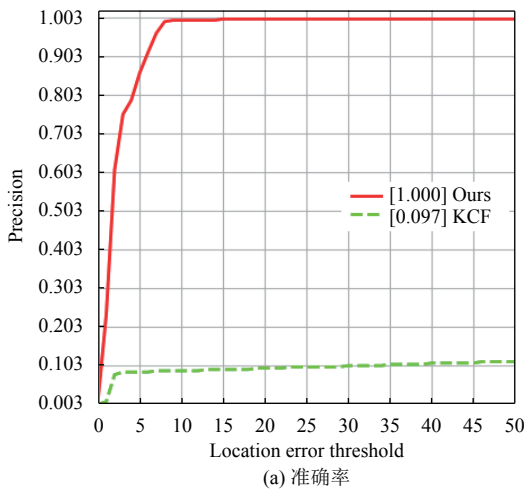


图3 plane_06 序列中的准确率和成功率对比图

图4 ship_01 序列中的准确率和成功率对比图

为了更直观评估算法的跟踪性能,将本文算法与KCF 算法分别在 plane_03、plane_06 和 ship_01 视频序列上进行了对比实验,其中红色框代表本文算法跟踪结果,蓝色框代表 KCF 算法跟踪结果.本文算法相较于 KCF 算法在卫星视频中可以实现更加稳定准确的跟踪.如图 7、图 8 和图 9 所示.

3 结论

针对卫星视频中的小目标,本文提出一种简单的方法可以同时实现对单目标的检测分割和跟踪两个任务.分别结合特征信息、运动信息和背景信息训练得到两个跟踪器和一个检测分割器,并将跟踪结果和检测结果分别进行相关计算得到最优目标位置.实验结果表明,算法可以成功实现单目标的分割和跟踪,提高了卫星视频中目标的跟踪性能.



图5 ViBe 算法与改进算法在 plane_03 序列分割结果对比



图6 ViBe 算法与改进算法在 plane_06 序列分割结果对比



图7 两种算法在 plane_03 序列跟踪结果



图8 两种算法在 plane_06 序列跟踪结果



图9 两种算法在 ship_01 序列跟踪结果

参考文献

- Bolme DS, Beveridge JR, Draper BA, *et al.* Visual object tracking using adaptive correlation filters. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010. 2544–2550. [doi: [10.1109/CVPR.2010.5539960](https://doi.org/10.1109/CVPR.2010.5539960)]
- Henriques JF, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels. Proceedings of the 12th European Conference on Computer Vision. Florence: Springer, 2012. 702–715.
- Danelljan M, Khan FS, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 1090–1097. [doi: [10.1109/CVPR.2014.143](https://doi.org/10.1109/CVPR.2014.143)]
- Henriques JF, Caseiro R, Martins P, *et al.* High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583–596. [doi: [10.1109/TPAMI.2014.2345390](https://doi.org/10.1109/TPAMI.2014.2345390)]
- Zhang KH, Zhang L, Liu QS, *et al.* Fast visual tracking via dense spatio-temporal context learning. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 127–141.
- Shao J, Du B, Wu C, *et al.* Can we track targets from space? A hybrid kernel correlation filter tracker for satellite video. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8719–8731. [doi: [10.1109/TGRS.2019.2922648](https://doi.org/10.1109/TGRS.2019.2922648)]
- Xuan SY, Li SY, Han MF, *et al.* Object tracking in satellite videos by improved correlation filters with motion estimations. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(2): 1074–1086. [doi: [10.1109/TGRS.2019.2943366](https://doi.org/10.1109/TGRS.2019.2943366)]
- Wang YM, Wang TY, Zhang G, *et al.* Small target tracking in satellite videos using background compensation. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10): 7010–7021. [doi: [10.1109/TGRS.2020.2978512](https://doi.org/10.1109/TGRS.2020.2978512)]
- 韩鸣飞, 李盛阳, 万雪, 等. 融合特征的卫星视频车辆单目标跟踪. 中国图象图形学报, 2021, 26(11): 2741–2750.
- Zhang WH, Jiao LC, Liu F, *et al.* MBLT: Learning motion and background for vehicle tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 4703315. [doi: [10.1109/TGRS.2021.3109028](https://doi.org/10.1109/TGRS.2021.3109028)]
- Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences. IEEE Transactions on Image Processing, 2011, 20(6): 1709–1724. [doi: [10.1109/TIP.2010.2101613](https://doi.org/10.1109/TIP.2010.2101613)]
- Zhao MQ, Li SY, Xuan SY, *et al.* SatSOT: A benchmark dataset for satellite video single object tracking. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5617611. [doi: [10.1109/TGRS.2022.3140809](https://doi.org/10.1109/TGRS.2022.3140809)]

(校对责编: 孙君艳)