

## 基于目标运动模型的无人机对地视觉目标跟踪

苏昂<sup>1,2\*</sup>, 陆伟康<sup>1,2</sup>, 张仕林<sup>1</sup>, 李璋<sup>1,2</sup><sup>1</sup>国防科技大学空天科学学院, 湖南 长沙 410073;<sup>2</sup>图像测量与视觉导航湖南省重点实验室, 湖南 长沙 410073

**摘要** 视觉目标跟踪是无人机对地面瞬时和时敏目标执行侦察、定位和打击任务的关键,面临着成像平台运动、严重遮挡、目标从视场中消失等困难。针对该问题,提出一种基于运动模型的无人机对地视觉目标跟踪方法。首先,采用基于稠密逆向搜索的快速光流算法计算相邻两帧之间的单应变换,将目标位置从历史帧映射到当前参考帧,以解耦成像平台运动。然后,在参考帧上采用线性运动模型对目标运动建模,当遮挡发生时利用运动模型预测目标位置。最后,采用短时和长时跟踪器结合的方法解决部分遮挡目标样本误更新跟踪器导致的跟踪漂移问题。以相关滤波跟踪器为基准,在收集的无人机视频上开展实验,结果表明所提方法能够显著提高对成像平台运动和严重遮挡的适应性,并且能够方便集成到其他目标跟踪方法中。

**关键词** 目标跟踪; 运动模型; 相关滤波; 光流; 无人机

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1415022

## Visual Ground Target Tracking of Unmanned Aerial Vehicle Based on Target Motion Model

Su Ang<sup>1,2\*</sup>, Lu Weikang<sup>1,2</sup>, Zhang Shilin<sup>1</sup>, Li Zhang<sup>1,2</sup><sup>1</sup>College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, Hunan, China;<sup>2</sup>Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, Hunan, China

**Abstract** Visual target tracking is crucial for an unmanned aerial vehicle (UAV) to conduct a strike, location, and reconnaissance against moving and time-sensitive ground targets; however it is hindered by imaging platform motion, severe occlusion, and target disappearance from the field of vision. A visual ground target tracking approach based on a motion model for UAVs is proposed to enhance the robustness for these challenges. First, a fast optical flow algorithm based on the dense inverse search is used to compute the homography transformation between two consecutive frames, and the target position is mapped from the historical frame to the present reference frame to decouple the motion of the imaging platform. The target motion on the reference frame is then modeled using a linear motion model, which is used to predict the target position when occlusion occurs. Finally, short-term and long-term trackers are combined to solve the tracking drift generated by the false update of the tracker for partially occluded target samples. Based on the discriminative correlation filter, experiments were conducted on the collected UAV videos. The findings reveal that the proposed approach can substantially improve the adaptation to the imaging platform motion and severe occlusion, and can be easily combined with other target tracking approaches.

**Key words** target tracking; motion model; correlation filtering; optical flow; unmanned aerial vehicle

## 1 引言

近年来无人机在军事和民用领域均逐渐得到广泛

应用,在各行各业发挥着越来越重要的作用。在最近的反恐战争和低烈度局部战争中,察打一体无人机取得了丰富战果,例如,美国MQ-9“死神(Reaper)”无人

收稿日期: 2022-05-05; 修回日期: 2022-05-31; 录用日期: 2022-06-05

基金项目: 国家自然科学基金(11727804)

通信作者: \*suang2008@126.com

机、装备雷神 MTS-B 光电吊舱和 AGM-114“地狱火 (Hellfire)”激光制导导弹。作战过程中,依靠光电吊舱发现和跟踪目标,然后用激光指示器照射目标,通过照射的激光光斑引导导弹命中目标。无人机光电吊舱中视觉目标长时稳定跟踪对无人机侦察、监视、定位和精确打击目标起着决定性作用。在民用方面,无人机执行搜救、监视、定位及跟随伴飞任务时,也需要依靠视觉对地面目标进行持续稳定跟踪。因此研究稳定可靠的视觉目标跟踪方法对无人机的应用具有重要意义。

随着深度学习技术的发展及其在目标跟踪领域的应用,跟踪算法逐渐发展为传统方法和深度学习方法两大类<sup>[1]</sup>。传统方法主要采用 Histogram of Oriented Gradient (HOG)、Color Names (CN)、其他手工设计特征及在线更新的判别模型进行目标跟踪,包括 TLD<sup>[2]</sup>、Struck<sup>[3]</sup>、KCF<sup>[4]</sup>等经典跟踪方法。其中,相关滤波类方法由于计算效率高和跟踪效果好等优势成为传统方法的主流,发展出了解决尺度问题的 DSST<sup>[5]</sup>、解决边界效应的 SRDCF<sup>[6]</sup>、面向长时目标跟踪的 LCT<sup>[7]</sup>和 MUSTer<sup>[8]</sup>、提升跟踪效率的 fDSST<sup>[9]</sup>、采用空间和通道可靠性加权的 CSR-DCF<sup>[10]</sup>、基于空间和时域正则化的 STRCF<sup>[11]</sup>和 AutoTrack<sup>[12]</sup>、面向快速运动目标的自适应跟踪方法<sup>[13]</sup>等。传统目标跟踪方法具有计算简单、速度较快的特点,但是由于手工设计特征表达能力不足,影响了跟踪性能进一步提升。

最初的深度学习跟踪方法采用卷积神经网络提取特征,再结合传统跟踪方法的在线更新判别模型进行跟踪,比如 DeepSRDCF<sup>[14]</sup>、C-COT<sup>[15]</sup>、ECO<sup>[16]</sup>、UPDT<sup>[17]</sup>等。其中 DeepSRDCF 基于 SRDCF 采用深度特征发展而来,C-COT 和 ECO 采用连续的相关滤波跟踪框架,UPDT 在 ECO 基础上自适应融合深度特征和手工特征进行跟踪。随后产生了端到端的深度学习跟踪方法,具有代表性的是 MDNet<sup>[18]</sup>,该方法在线更新神经网络模型,计算复杂度高,影响了跟踪速度。然后出现的孪生网络 (siamese network) 跟踪方法直接采用卷积神经网络提取特征并离线训练匹配函数,无需在线更新神经网络,因此跟踪效果好、速度快。经典的代表性方法有 SINT<sup>[19]</sup>、SiameFC<sup>[20]</sup>、CFNet<sup>[21]</sup>、SiamRPN<sup>[22]</sup>、DaSiamRPN<sup>[23]</sup>、SiamRPN++<sup>[24]</sup>、SiamFC++<sup>[25]</sup>、Ocean<sup>[26]</sup>等。SINT 和 SiameFC 奠定了孪生网络目标跟踪基本框架,SiamRPN、DaSiamRPN 和 SiamRPN++ 引入了区域生成网络 (RPN) 进行跟踪,SiamFC++ 和 Ocean 采用了更简洁的无锚框目标表示方法。此外还产生了面向红外目标<sup>[27]</sup>和快速运动目标<sup>[28]</sup>的孪生网络跟踪方法。虽然视觉目标跟踪算法已取得了众多研究进展,但是面对无人机对地面目标跟踪时成像平台运动导致的目标在图像中的快速移动、严重的目标遮挡甚至从视场中消失等困难时依然存在跟踪性能不足的问题。

本文针对无人机对地面目标跟踪的问题,研究一种基于目标运动模型的跟踪方法,增强了对成像平台运动导致的目标快速运动、严重目标遮挡等问题的适应性。提出一种基于解耦成像平台运动的目标运动建模方法,采用基于稠密逆向搜索的快速光流算法计算相邻两帧之间的单应变换,将目标位置从历史帧映射到当前参考帧,然后在参考帧上采用线性运动模型对目标运动建模;提出一种长短时相关滤波跟踪器方法,通过短时和长时跟踪器的配合,解决无法准确判断目标遮挡开始时间导致部分遮挡目标样本错误更新跟踪器而产生的跟踪漂移问题。

## 2 基本原理

### 2.1 相关滤波目标跟踪

相关滤波目标跟踪方法通过实时更新训练一个相关滤波器实现跟踪。假设在输入图像上提取的  $d$  维特征为  $\mathbf{f}$ ,令  $i$  表示特征维度,则  $i \in \{1, \dots, d\}$ ,其中每个维度的特征  $\mathbf{f}^i$  均为二维矩阵形式。针对特征  $\mathbf{f}^i$  优化计算一个相关滤波器  $\mathbf{h}^i$ ,使得目标函数最小,目标函数表达式为

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

式中:  $\mathbf{g}$  表示期望的相关滤波输出响应,通常假设为以目标位置为中心的二维高斯函数;  $\lambda$  为正则化参数。根据文献[5],该目标函数的最优解为

$$\mathbf{H}^i = \frac{\overline{\mathbf{G}} \cdot \mathbf{F}^i}{\sum_{k=1}^d \overline{\mathbf{F}}^k \cdot \mathbf{F}^k + \lambda}, \quad (2)$$

式中:  $\mathbf{H}^i$ 、 $\mathbf{F}^i$ 、 $\mathbf{G}$  分别表示  $\mathbf{h}^i$ 、 $\mathbf{f}^i$ 、 $\mathbf{g}$  经傅里叶变换后的矩阵;  $\overline{\mathbf{G}}$  和  $\overline{\mathbf{F}}^k$  分别表示  $\mathbf{G}$  和  $\mathbf{F}^k$  的复共轭。其中的矩阵相乘均为点乘。跟踪过程中,根据当前目标位置对当前时刻  $t$  的滤波器  $\mathbf{H}_t^i$  进行更新,通常对其分子  $\mathbf{A}_t^i$  和分母  $\mathbf{B}_t^i$  分别更新,表示为

$$\mathbf{A}_t^i = (1 - \eta) \mathbf{A}_{t-1}^i + \eta \overline{\mathbf{G}}_t \mathbf{F}_t^i, \quad (3)$$

$$\mathbf{B}_t^i = (1 - \eta) \mathbf{B}_{t-1}^i + \eta \sum_{k=1}^d \overline{\mathbf{F}}_t^k \mathbf{F}_t^k, \quad (4)$$

式中:  $\eta$  为学习率参数,作为控制滤波器更新的权重。假设在下一帧待跟踪图像上提取的特征矩阵为  $\mathbf{z}$ ,其傅里叶变换结果设为  $\mathbf{Z}$ ,那么跟踪响应值  $\mathbf{r}$  的计算公式为

$$\mathbf{r} = \text{IFFT} \left\{ \frac{\sum_{i=1}^d \overline{\mathbf{A}}_t^i \mathbf{Z}^i}{\mathbf{B}_t + \lambda \mathbf{I}} \right\}, \quad (5)$$

式中:  $\text{IFFT} \{ \cdot \}$  表示傅里叶逆变换。然后在二维跟踪响应矩阵上寻找最大值,得到目标跟踪的位置。特征提取部分与文献[5]类似,采用 PCA-HOG 特征<sup>[29]</sup>以提高计算效率,主要是利用主成分分析法 (PCA) 降低 HOG 特征的维度,通过 PCA 将 KCF 跟踪算法<sup>[4]</sup>中使用的 31 维 HOG 特征转为 18 维的特征。

## 2.2 稠密逆向搜索光流

无人机对地面目标成像时,目标在图像中的运动叠加了成像平台运动和地面目标的运动,导致运动模式具有较强的随机性,不易直接进行运动建模,需要分离成像平台运动,再对目标在地面的运动进行建模。无人机从高空对地面场景成像,通常可以假设地面近似为一个平面,那么前后帧图像之间的变换关系可以用单应矩阵来表示。将  $t$  和  $t-1$  时刻图像上的二维图像坐标写成齐次形式,分别为  $\mathbf{x}_t$  和  $\mathbf{x}_{t-1}$ ,记  $\mathbf{H}'_{t-1}$  为单应变换矩阵,那么两幅图像之间的单应变换关系表示为  $\mathbf{x}_t = \mathbf{H}'_{t-1} \mathbf{x}_{t-1}$ 。

为了求解相邻两帧图像之间的单应变换,采用一种基于稠密逆向搜索(DIS)的快速光流算法<sup>[30]</sup>计算两幅图像之间的密集光流场,根据阈值分割出背景和运动目标,然后仅使用背景区域光流求解两帧之间的单应变换矩阵。假设  $t$  时刻图像  $I_t$  为参考帧,在参考帧上以坐标  $\mathbf{x} = [x, y]^T$  为中心、分辨率为  $\theta_{ps} \times \theta_{ps}$  的图像块  $T$  为模板,需要在  $t-1$  时刻从图像  $I_{t-1}$  上找到一个同样分辨率的图像块与模板  $T$  实现最佳匹配,该过程的优化目标函数为

$$\mathbf{u} = \arg \min_{\mathbf{u}'} \sum_{\mathbf{x}} [I_{t-1}(\mathbf{x} + \mathbf{u}') - T(\mathbf{x})]^2, \quad (6)$$

式中:向量  $\mathbf{u} = [u, v]^T$  即为待求的光流矢量。求解该目标函数式(6)需要采用非线性迭代。迭代求解过程中,假设当前光流估计值为  $\mathbf{u}$ ,每次迭代时需要计算相对于当前光流估计值  $\mathbf{u}$  的最优改进量  $\Delta \mathbf{u}$ 。对于该问题,经典的求解目标函数<sup>[31]</sup>的表达式为

$$\Delta \mathbf{u} = \arg \min_{\Delta \mathbf{u}'} \sum_{\mathbf{x}} [I_{t-1}(\mathbf{x} + \mathbf{u} + \Delta \mathbf{u}') - T(\mathbf{x})]^2. \quad (7)$$

每次迭代时都需要根据当前光流估计值  $\mathbf{u}$  对待匹配图像  $I_{t-1}$  进行亚像素插值变换,得到  $I_{t-1}(\mathbf{x} + \mathbf{u})$ ,并针对变换后图像重新计算 Hessian 矩阵。为了提高计算效率,DIS 光流法采用的目标函数为

$$\Delta \mathbf{u} = \arg \min_{\Delta \mathbf{u}'} \sum_{\mathbf{x}} [T(\mathbf{x} - \mathbf{u}) - I_{t-1}(\mathbf{x} + \Delta \mathbf{u}')]^2, \quad (8)$$

避免了每次迭代都要重新计算  $I_{t-1}(\mathbf{x} + \mathbf{u})$  以及 Jacobian 和 Hessian 矩阵。

## 2.3 目标运动建模

根据求解的图像序列帧间单应变换关系,将目标历史位置映射到当前参考帧,如图 1 所示。此时,成像平台运动被分离,在参考帧图像坐标系下仅包含目标自身运动,针对目标运动的具体状态,可以对目标运动进行建模。常用的机动目标运动模型有匀速运动、匀加速运动、转弯运动等,可以采用时间的一阶或二阶函数对不同类型的目标运动建模。

针对无人机对地面目标跟踪的任务,根据视频序列中比如车辆、行人等目标的运动特点,采用最常见的匀速直线运动假设。设目标的图像坐标为  $(p_x, p_y)$ ,建



图 1 根据单应变换将目标历史位置映射到当前参考帧的结果  
Fig. 1 Result of mapping the historical position of the target to the current reference frame according to homography transformation

立的线性运动模型为

$$\begin{cases} p_x = a_{11}t + a_{12} \\ p_y = a_{21}t + a_{22} \end{cases}, \quad (9)$$

根据映射到当前帧图像坐标系的目标历史位置坐标,求解目标运动模型的四个参数  $a_{11}$ 、 $a_{12}$ 、 $a_{21}$  和  $a_{22}$ 。再根据求解的目标运动模型,即可预测在当前帧上目标可能出现的位置。

## 2.4 长短时相关滤波跟踪器

相关滤波目标跟踪过程中,需要使用当前帧目标样本对滤波器  $\mathbf{h}$  进行更新。当目标被遮挡时需要及时暂停更新,否则被部分遮挡的目标样本将污染跟踪滤波器,导致跟踪漂移,因此准确判断目标是否被遮挡极为关键。通常根据相关滤波器的跟踪响应最大值  $r_{\max}$  判断目标是否被遮挡,然而这种判断并不十分准确。目标被遮挡往往是从部分遮挡到完全遮挡的渐变过程,当目标刚开始被部分遮挡时,跟踪响应分数并不会立即下降,存在一定的滞后,此时部分遮挡的目标样本将误更新跟踪滤波器,从而造成滤波器污染。针对以上问题,提出一种长短时相关滤波器的方法,设计一个长时滤波器和一个短时滤波器,其中短时滤波器逐帧更新,用于目标未被遮挡时的目标跟踪,长时滤波器间隔  $N$  帧更新(比如  $N$  取 25),用于目标被遮挡后的目标重捕获。基于长短时滤波器的目标跟踪算法的整体工作流程如图 2 所示。

首先,采用 DIS 快速光流算法解耦摄像机运动,将目标历史位置映射到当前帧图像坐标系中对目标运动建模,预测目标可能出现的位置;然后,当目标被遮挡时,根据短时滤波器的跟踪响应判断目标被遮挡,此时逐帧更新的短时滤波器会被污染,而长时滤波器由于采用间隔更新策略还未被污染;那么即可根据目标运动模型预测的目标可能位置,采用长时滤波器进行搜索,当重新捕获目标后,根据长时滤波器重启短时滤波器,对目标进行持续跟踪。

在判断目标是否被遮挡和目标是否被重捕获时,根据跟踪响应最大值  $r_{\max}$  直接设定阈值来判断在实际操作中不够方便和稳定,随着每次跟踪目标的不同,相关滤波跟踪响应值的大小存在一定的差异。因此,采

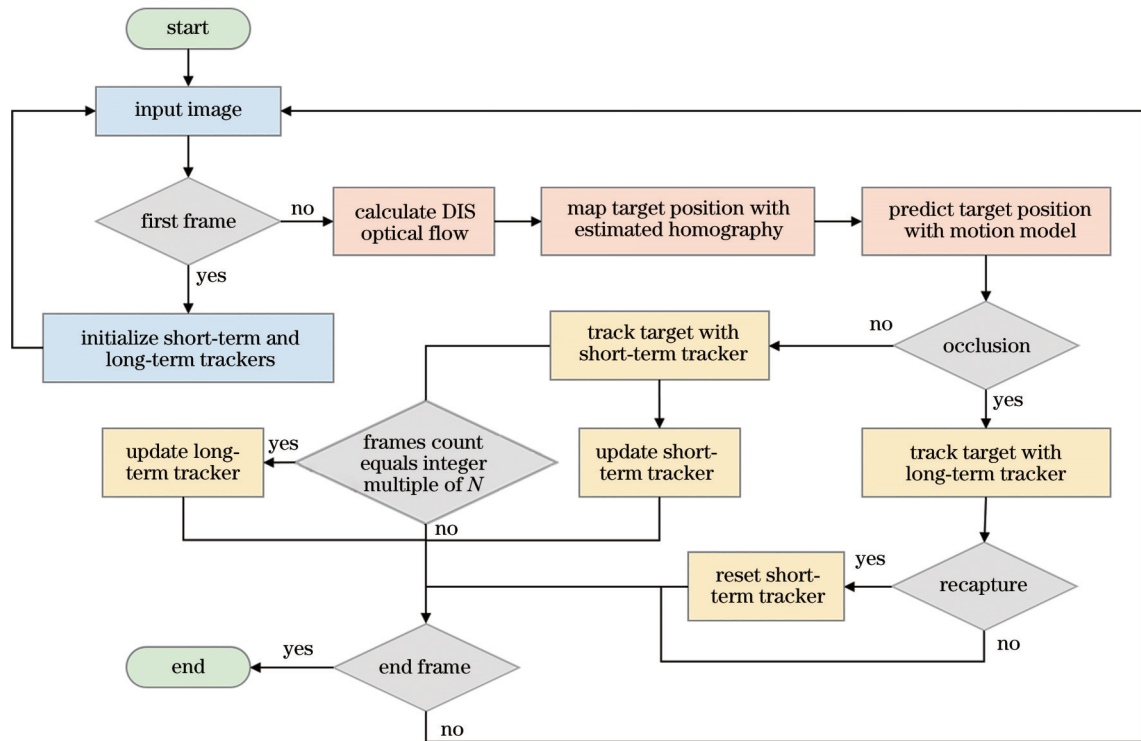


图 2 目标跟踪算法流程

Fig. 2 Flowchat of the target tracking algorithm

用当前跟踪响应最大值  $r_{\max}$  与跟踪响应历史平均值  $r_{\text{mean}}$  的比例系数  $\gamma = r_{\max}/r_{\text{mean}}$  进行判断,比例系数  $\gamma$  的判断阈值可以设置为 0.7。

### 3 实验结果分析

#### 3.1 实验配置

为了对所提基于运动模型的目标跟踪方法进行实验测试,收集了无人机实际拍摄的 11 段具有挑战性的视频序列,共计约 4850 帧,目标大小约 25 pixel,包含成像平台突然运动导致的目标快速移动、严重的目标遮挡以及目标在视场中消失再出现等无人机对地面目标跟踪时普遍面临的困难因素。此外,还从公开的无人机单目标跟踪数据集 VisDrone2019<sup>[32]</sup>中选择了 11 个视频序列开展实验,包含公园、广场、草地、公路、操场等不同场景。实验测试时,目标跟踪部分在文献[4]和文献[5]提供的 Matlab 开源代码基础上开发,帧间单应估计部分采用 Opencv 3.4.1 提供的 DIS 光流算法,具体实现时使用 mexopencv,在 Matlab 中可直接调用相关函数。实验使用的笔记本电脑配置 2.30 GHz Inter CPU 和 8 GB 内存并安装 Windows 10 操作系统。

#### 3.2 实拍数据集实验结果定性分析

对无人机实拍的 11 段视频序列人工标注目标框,作为真值。首先,对所提方法进行定性分析,根据每段视频序列最后一帧跟踪器得到的目标中心位置与真值中心位置的距离是否大于 10 个像素为标准来判定跟踪是否失败。实验对比了 KCF<sup>[4]</sup>、fDSST<sup>[9]</sup>、ECO<sup>[16]</sup>、

TLD<sup>[2]</sup>、MUSTer<sup>[8]</sup>和 LCT<sup>[7]</sup>等经典的传统跟踪方法,结果如表 1 所示,其中符号  $\checkmark$  表示跟踪成功,符号  $\times$  表示跟踪失败。

所提跟踪方法标记为 Ours 和 Ours\_450,两者的区别是提取 DIS 光流时,Ours 方法采用目标附近  $300 \times 300$  像素,而 Ours\_450 方法采用附近  $450 \times 450$  像素。使用更大图像区域计算 DIS 光流获取单应矩阵时会更稳定,但会影响计算效率。Seq 10 和 Seq 11 视频序列中背景纹理特征不明显,使得 DIS 光流法无法准确地计算帧间光流,未能有效消除成像平台运动,从而导致 Ours 方法跟踪失败。Ours\_450 方法通过扩大光流的计算范围得到了更稳定的帧间单应估计,从而能够成功跟踪。在 Seq 4 视频序列中,由于目标被遮挡的时间较长而且在遮挡的同时伴随着成像平台的快速运动和旋转,此时单应性变换误差累计较大,目标运动模型建模不准确,导致预测的目标位置偏差较大。当目标重新出现时,所提方法的长时滤波器无法实现目标重捕获,导致跟踪失败。TLD 方法对 Seq 4 能够成功跟踪的原因是该方法设计了独立的目标检测模块,用于发生遮挡时在全图搜索进行目标重捕获。MUSTer 方法对 Seq 4 能够成功跟踪的原因是该方法在相关滤波模块之外设计了一个基于帧间关键点匹配的跟踪模块,对成像平台导致的目标运动具有较好适应性。但是这两种方法设计的目标检测模块和关键点匹配模块会消耗较大的计算资源,导致跟随速度较慢。

实验测试使用的 11 个视频序列是在无人机实拍

表 1 无人机实拍视频目标跟踪结果

Table 1 Target tracking results for real videos captured by unmanned aerial vehicle

Sequence	KCF	fDSST	ECO	TLD	MUSTer	LCT	Ours	Ours_450
Seq 1 (400 frames)	✓	✓	×	✓	✓	✓	✓	✓
Seq 2 (400 frames)	×	✓	×	✓	✓	✓	✓	✓
Seq 3 (500 frames)	×	×	✓	✓	✓	×	✓	✓
Seq 4 (900 frames)	×	×	×	✓	✓	×	×	×
Seq 5 (300 frames)	✓	✓	✓	✓	✓	✓	✓	✓
Seq 6 (400 frames)	×	✓	✓	✓	✓	×	✓	✓
Seq 7 (500 frames)	×	×	✓	✓	✓	×	✓	✓
Seq 8 (300 frames)	×	×	×	✓	×	×	✓	✓
Seq 9 (250 frames)	✓	✓	✓	✓	×	✓	✓	✓
Seq 10 (300 frames)	×	×	×	✓	✓	×	×	✓
Seq 11 (650 frames)	×	×	×	×	×	×	×	✓
Success times	3	5	5	10	8	4	8	10

的长视频中分段截取的,采用原始长视频中的帧序号对视频序列中的图像编号,对部分视频序列的跟踪结果进行展示和分析。视频序列 Seq 3 对应原始视频的第 6301 帧到第 6800 帧,部分帧的跟踪结果如图 3 所示。此时成像平台存在突然的大幅运动,导致目标小车在图像中快速移动,其中 KCF、fDSST 和 LCT 方法跟踪失败,其他方法跟踪成功。目标快速运动超过了 KCF、fDSST 和 LCT 方法预设的跟踪搜索范围,导致跟踪失败。虽然 LCT 方法设计目标重捕获机制,但是由于目标快速运动,算法无法判断合适的目标重捕获搜索范围,因此无法正确重捕获;相对而言,ECO 方法预设了更大的搜索范围,因此能够跟踪成功;TLD 方法跟踪时利用了稀疏光流,MUSTer 方法跟踪时采用了关键点匹配模块,两种方法对目标快速运动具有一

定适应能力,从而均跟踪成功;Ours 通过消除成像平台运动的不利影响,也能够实现对目标的稳定跟踪。

视频序列 Seq 8 对应原始视频的第 10001 帧到第 10300 帧,部分帧的跟踪结果如图 4 所示。目标小车下穿立交桥被完全遮挡时,只有 Ours 和 TLD 方法能够跟踪成功,其他方法均跟踪失败。当目标被完全遮挡从视野中消失时,由于无法准确判断发生遮挡的开始时刻,在目标被逐渐遮挡过程中,KCF、fDSST 和 ECO 方法会在更新跟踪器时引入部分遮挡目标样本的干扰,产生跟踪漂移,最终导致跟踪失败;LCT 方法虽然设计了独立的目标检测模块,但是无法预测目标可能出现的位置,导致目标重捕获失败;MUSTer 方法虽然设计了目标关键点匹配模块,但是却无法适应这种目标被完全遮挡的情况,从而导致跟踪失败;TLD 方法

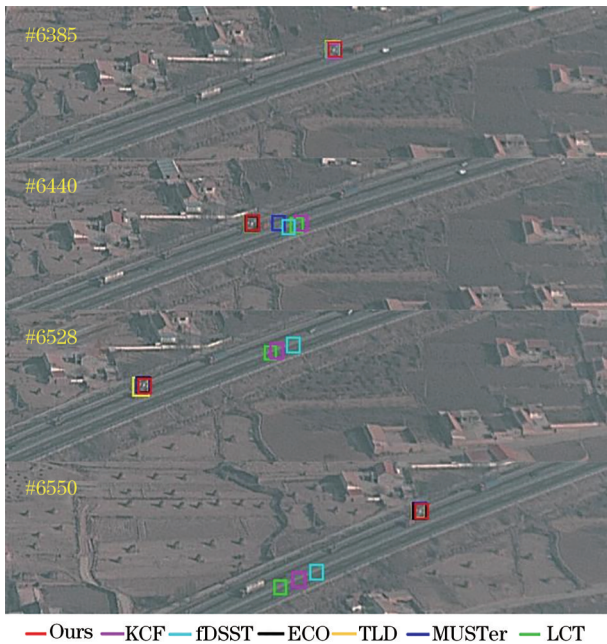


图 3 视频序列 Seq 3 的目标跟踪结果

Fig. 3 Object tracking results of video sequence Seq 3

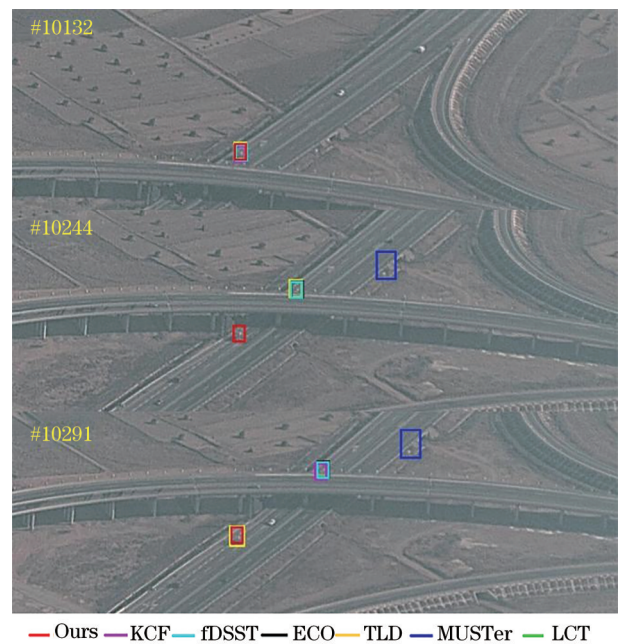


图 4 视频序列 Seq 8 的目标跟踪结果

Fig. 4 Object tracking results of video sequence Seq 8

具有全图搜索的独立目标检测模块,因此当目标重新出现时可以重新捕获目标,最终实现成功跟踪;所提方法在判断目标发生遮挡后,停止长时和短时滤波器更新,根据目标运动模型预测目标可能出现的位置,在该位置使用长时滤波器进行目标重捕获,当目标重新出现时,长时滤波器捕获目标,重启短时滤波器并开启长时和短时滤波器更新,因此能够适应目标被完全遮挡的情况,最终成功跟踪目标。

### 3.3 实拍数据集实验结果定量分析

根据人工标注的目标框真值,采用成功率和准确率对所提跟踪方法进行定量分析。其中,成功率定义

为跟踪目标框与真值目标框的重叠率不低于一定阈值条件的跟踪帧数与视频序列总帧数之比;准确率定义为跟踪目标框中心位置与真值目标框中心位置的偏差 not 高于一定阈值条件的跟踪帧数与视频序列总帧数之比。

实验同样对比了 KCF、fDSST、ECO、TLD、MUSTer 和 LCT 等经典跟踪方法,结果如图 5 所示。实验结果表明:Ours 的平均成功率和平均准确率分别达 0.745 和 0.824,显著优于其他跟踪方法;Ours\_450 方法通过扩大光流场计算范围,将平均成功率和平均准确率分别提高到 0.891 和 0.969。

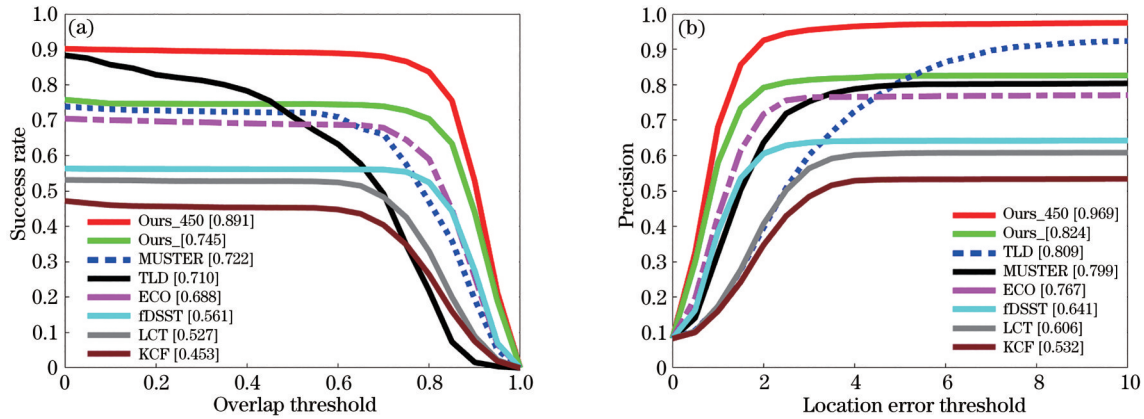


图 5 实拍数据集中目标跟踪成功率和准确率的实验结果。(a)成功率;(b)准确率

Fig. 5 Experiment results of object tracking success rate and precision on collected dataset. (a) Success rate; (b) precision

实验测试了各跟踪方法的运行效率,结果如表 2 所示。

实验结果表明,Ours 和 Ours\_450 跟踪速度分别达 52.3 frame/s 和 30.2 frame/s。由于所提方法增加了基于快速光流法的帧间单应估计步骤,相对于 KCF 和

fDSST 等相关滤波跟踪方法,在跟踪速度方面存在较大的下降,但是依然能够基本满足无人机对地面目标视觉跟踪的实时性要求。而 TLD 和 MUSTer 方法虽然跟踪效果也较好,但是由于设计了独立的目标检测模块和关键点匹配模块,消耗了大量计算资源,跟踪速度很低。

表 2 各算法的跟踪速度

Table 2 Tracking speed of different algorithms

Parameter	KCF	fDSST	ECO	TLD	MUSTer	LCT	Ours	Ours_450
Speed / (frame · s <sup>-1</sup> )	240.3	139.0	50.5	3.8	0.7	32.4	52.3	30.2

### 3.4 公开数据集实验结果分析

使用公开数据集 VisDrone2019 单目跟踪任务中 11 个包括摄像机晃动、地面目标运动和遮挡的视频序列开展实验,同样对比了 KCF、fDSST、ECO、TLD、MUSTer 和 LCT 等经典跟踪方法,结果如图 6 所示。在该数据集中只使用 Ours\_450 方法进行测试,实验结果表明,Ours\_450 依然表现突出,平均成功率和平均准确率分别达 0.675 和 0.964,其中跟踪平均准确率指标显著优于其他跟踪方法。但是,由于所提方法没有设计尺度估计模块,因此在公开数据集上目标重叠率方面表现欠佳,跟踪平均成功率指标落后于 ECO 跟踪方法。

公开数据集 VisDrone2019 中三个具有代表性和挑战性的视频序列目标跟踪结果如图 7 所示。序列 1

跟踪的目标为运动的女孩,包含目标遮挡、相似目标干扰、运动模糊及无人机视角变化等跟踪挑战。第 424 帧前后,目标与相似目标发生交会,只有 ECO 和所提方法能成功跟踪目标,其余跟踪方法均发生了误检,将经过的相似干扰物当作目标。序列 2 中跟踪的目标为运动的小女孩,包含光照强度低、背景快速移动等跟踪挑战。第 365 帧前后,帧间的背景运动变化较为剧烈,导致其他算法跟踪失败,只有 ECO 和所提方法能成功跟踪目标。序列 3 中跟踪的目标为运动的小绵羊,包含小目标跟踪、目标外观变化、背景快速移动等跟踪挑战。第 705 帧前后,无人机镜头转动过快导致背景和 target 发生快速移动,其他方法跟踪失败,只有 TLD 和所提方法能够准确地跟踪目标。总体来说,相比于其他跟踪方法,所提方法采用帧间单应估计消除摄像机

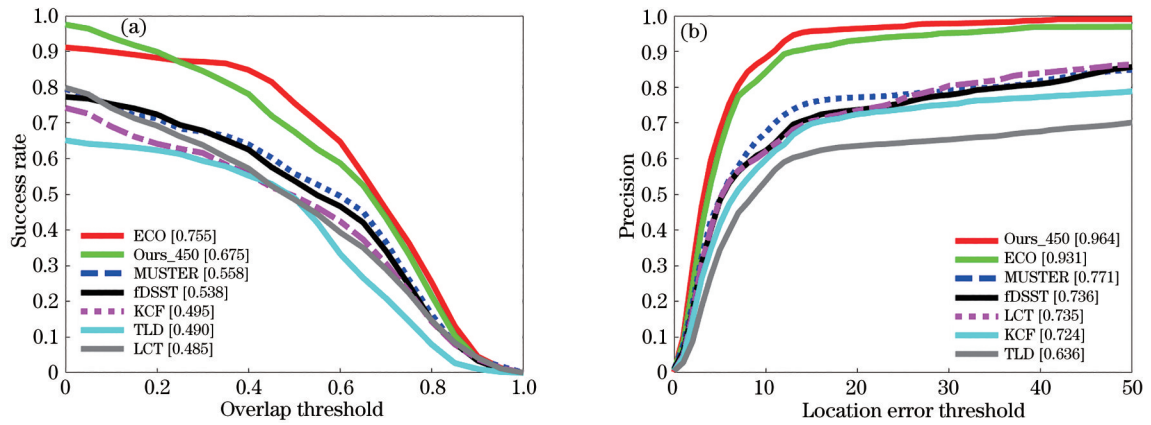


图 6 VisDrone2019 部分视频序列中目标跟踪成功率和准确率的实验结果。(a)成功率;(b)准确率  
Fig. 6 Experiment results of object tracking success rate and precision on VisDrone2019 partial video sequences. (a) Success rate; (b) precision

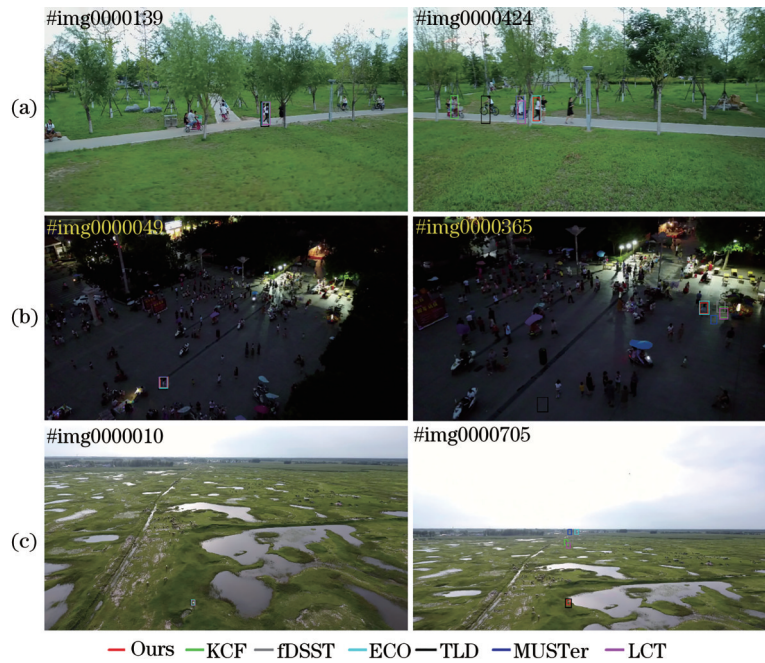


图 7 VisDrone2019 数据集中视频序列目标跟踪结果。(a)序列 1;(b)序列 2;(c)序列 3  
Fig. 7 Object tracking results of video sequence in VisDrone2019 dataset. (a) Seq 1; (b) Seq 2; (c) Seq 3

运动带来的背景和目标快速移动影响,并且在分离摄像机运动后对目标运动进行建模,从而能够准确估计目标可能出现的位置,使得所提跟踪方法能够有效适应背景和目标快速运动、严重的目标遮挡及目标附近有相似目标干扰等问题。

## 4 结 论

针对无人机对地视觉目标跟踪问题,提出了一种基于运动模型的目标跟踪方法,该方法能够适应无人机成像平台运动导致的目标在图像中快速移动、严重的目标遮挡、目标从视场中消失等问题,并且可以方便地集成到其他跟踪方法中。相对于传统相关滤波跟踪方法,所提方法利用基于稠密逆向搜索的快速光流法解耦成像平台运动,通过设计长短时滤波器解决目标

逐渐被遮挡过程中的相关滤波器漂移问题,最终跟踪速度能够满足实时性要求。在无人机实拍视频上开展的实验测试表明,所提方法能够实现稳定可靠的实时目标跟踪,具备较强的工程实用性。

## 参 考 文 献

- [1] 李玺, 查宇飞, 张天柱, 等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报, 2019, 24(12): 2057-2080. Li X, Zha Y F, Zhang T Z, et al. Survey of visual object tracking algorithms based on deep learning[J]. Journal of Image and Graphics, 2019, 24(12): 2057-2080.
- [2] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [3] Hare S, Torr P H S. Struck: structured output tracking with kernels[C]//2011 IEEE International Conference on

- Computer Vision (ICCV), November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 263-270.
- [4] 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.
- [5] Danelljan M, Häger G, Khan F S, et al. Accurate scale estimation for robust visual tracking[C]//Proceedings of the British Machine Vision Conference 2014, September 1-5, 2014, Nottingham. London: British Machine Vision Association, 2014.
- [6] Danelljan M, Häger G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking [C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4310-4318.
- [7] Ma C, Yang X K, Zhang C Y, et al. Long-term correlation tracking[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5388-5396.
- [8] Hong Z B, Chen Z, Wang C H, et al. MUlti-Store Tracker (MUSTer): a cognitive psychology inspired approach to object tracking[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 749-758.
- [9] Danelljan M, Häger G, Khan F S, et al. Discriminative scale space tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [10] Lukežič A, Vojtíř T, Zajc L Č, et al. Discriminative correlation filter tracker with channel and spatial reliability [J]. International Journal of Computer Vision, 2018, 126 (7): 671-688.
- [11] Li F, Tian C, Zuo W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4904-4913.
- [12] Li Y M, Fu C H, Ding F Q, et al. AutoTrack: towards high-performance visual tracking for UAV with automatic spatio-temporal regularization[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11920-11929.
- [13] 刘宗达, 董立泉, 赵跃进, 等. 视频中快速运动目标的自适应模型跟踪算法[J]. 光学学报, 2021, 41(18): 1815001.
- Liu Z D, Dong L Q, Zhao Y J, et al. Adaptive model tracking algorithm for fast-moving targets in video[J]. Acta Optica Sinica, 2021, 41(18): 1815001.
- [14] Danelljan M, Häger G, Khan F S, et al. Convolutional features for correlation filter based visual tracking[C]//2015 IEEE International Conference on Computer Vision Workshop, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 621-629.
- [15] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9909: 472-488.
- [16] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6931-6939.
- [17] Bhat G, Johnander J, Danelljan M, et al. Unveiling the power of deep tracking[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11206: 493-509.
- [18] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4293-4302.
- [19] Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1420-1429.
- [20] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.
- [21] Valmadre J, Bertinetto L, Henriques J, et al. End-to-end representation learning for correlation filter based tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5000-5008.
- [22] Li B, Yan J J, Wu W, et al. High performance visual tracking with siamese region proposal network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8971-8980.
- [23] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11213: 103-119.
- [24] Li B, Wu W, Wang Q, et al. SiamRPN: evolution of siamese visual tracking with very deep networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4277-4286.
- [25] Xu Y D, Wang Z Y, Li Z X, et al. SiamFC++ : towards robust and accurate visual tracking with target estimation guidelines[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12549-12556.



- [26] Zhang Z P, Peng H W, Fu J L, et al. Ocean: object-aware anchor-free tracking[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12366: 771-787.
- [27] 李畅, 杨德东, 宋鹏, 等. 基于全局感知孪生网络的红外目标跟踪[J]. 光学学报, 2021, 41(6): 0615002.  
Li C, Yang D D, Song P, et al. Global-aware siamese network for thermal infrared object tracking[J]. Acta Optica Sinica, 2021, 41(6): 0615002.
- [28] 郑军松, 郭浩, 李阿标, 等. 基于孪生网络快速运动弱目标实时跟踪[J]. 激光与光电子学进展, 2022, 59(4): 0410011.  
Zheng J S, Guo H, Li A B, et al. Real-time tracking of fast moving weak object based on siamese network[J]. Laser & Optoelectronics Progress, 2022, 59(4): 0410011.
- [29] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [30] Kroeger T, Timofte R, Dai D, et al. Fast optical flow using dense inverse search[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9908: 471-488.
- [31] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//Proceedings of the 7th International Joint Conference on Artificial Intelligence, August 24-28, 1981, Vancouver, BC, Canada. New York: ACM Press, 1981: 674-679.
- [32] Du D, Zhu P, Wen L, et al. VisDrone-SOT2019: the vision meets drone single object tracking challenge results [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), October 27-28, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 199-212.