

引用格式: SUN Mengyu, WANG Peng, XU Junqi, et al. Adaptive Information Selection for Infrared Object Tracking with Variable Scale Correlation Filter[J]. Acta Photonica Sinica, 2023, 52(12): 1210003

孙梦宇, 王鹏, 徐均琪, 等. 自适应信息选择的变尺度相关滤波红外目标跟踪算法[J]. 光子学报, 2023, 52(12): 1210003

# 自适应信息选择的变尺度相关滤波红外目标跟踪算法

孙梦宇<sup>1</sup>, 王鹏<sup>2</sup>, 徐均琪<sup>1</sup>, 李晓艳<sup>2</sup>, 郜辉<sup>2</sup>, 邱若海<sup>2</sup>

(1 西安工业大学 光电工程学院, 西安 710021)

(2 西安工业大学 电子信息工程学院, 西安 710021)

**摘要:** 针对红外目标跟踪中红外图像模糊、存在噪声、目标特征少的问题, 提出了自适应信息选择的变尺度相关滤波红外目标跟踪算法。首先, 在提取的灰度特征中重新提取了梯度信息, 用于增强特征感受野大小, 丰富目标特征信息; 其次, 稀疏滤波器系数, 减少滤波器信息冗余, 并将其结合至滤波器训练过程中, 不同通道下的空间信息保留程度不同, 有效提高滤波器表达能力; 最后, 在原有保留尺度更新的基础上, 加入变尺度滤波器, 构建边界框比例不同的尺度池, 有效应对边界框比例变化的情况。在 LSOTB-TIR 数据集和 PTB-TIR 数据集上做了对比和消融实验以验证算法的有效性, 结果表明该算法在 LSOTB-TIR 数据集上精确度和成功率分别达到 71.3% 和 59.4%, 在使用手工特征的算法中获得了更好的表现。

**关键词:** 红外与夜视技术; 目标跟踪; 相关滤波; 红外图像处理; 变尺度滤波器; 稀疏化表示

中图分类号: TP391.41

文献标识码: A

doi: 10.3788/gzxb20235212.1210003

## 0 引言

视觉目标跟踪任务作为机器视觉领域中的重要环节近年来受到了广泛的关注<sup>[1-4]</sup>。红外目标跟踪是基于热红外图像对目标进行跟踪的视觉任务, 根据热红外图像的成像原理, 在目标受到部分掩体遮挡时, 依然能够获取目标信息, 且在黑夜条件下, 红外相机能够对目标进行观测, 而可见光无法获取目标图像。但红外图像相较于可见光图像而言缺少纹理信息, 并且红外图像中目标边界模糊且存在噪声, 使得目标与背景区分难度更高<sup>[5-8]</sup>。

依据在线训练的优势, 相关滤波算法具备良好的鲁棒性<sup>[9-12]</sup>, 同时基于手工特征的相关滤波算法无需图形处理器(Graphics Processing Unit, GPU)或神经网络处理单元(Neural-Network Processing Unit, NPU), 适用于部署在计算能力弱的嵌入式硬件平台上。由于相关滤波算法训练无关图像模态, 因此较多学者以基于可见光图像的滤波算法<sup>[13-16]</sup>作为基础开展研究, 但直接将其应用难以获取较好的表现。GUNDOGDU E 等<sup>[17]</sup>训练了红外卷积网络用于特征提取。YU Tianwen 等<sup>[18]</sup>提出了基于连续卷积和自适应特征融合的鲁棒性红外跟踪器(Robust TIR tracker based on the Continuous Correlation Filters and adaptive feature fusion, RCCF-TIR), 采用相关峰值作为权重实现响应图的融合。ZHANG He 等<sup>[19]</sup>提出了基于水平集的 ECO 算法(Efficient Convolution Operators based on the Level Set method, ECO\_LS), 额外加入图像分割器, 将分割结果与跟踪结果融合。YUAN Di 等<sup>[20]</sup>提出了自适应多特征融合模型(Adaptively Multi-Feature fusion Tracker, AMFT), 根据 Kullback-Leibler 散度衡量各个特征对应的响应图。上述滤波算法更多关注于融合

**基金项目:** 国家自然科学基金(No. 62171360), 陕西省科技厅重点研发计划(No. 2022GY-110), 国家重点研发计划(No. 2022YFF0604900), 2022年度陕西高校青年创新团队项目, 山东省智慧交通重点实验室(筹)项目

**第一作者:** 孙梦宇, sunmengyu@st.xatu.edu.cn

**通讯作者:** 徐均琪, jqxu2210@163.com

**收稿日期:** 2023-04-28; **录用日期:** 2023-06-26

<http://www.photon.ac.cn>

深度特征与手工特征,运算负担较大,而ECO\_LS仅采用手工特征,获得了较好的表现,但其采用的并行结构同样使其无法完成实时运算。

近年来深度学习的目标跟踪算法收到了广泛的关注<sup>[21-22]</sup>。目标感知的深度跟踪算法(Target-Aware Deep Tracking, TADT)<sup>[23]</sup>利用回归损失和排序损失提高网络对前景的分辨能力;结合分割的孪生网络跟踪算法(SiamMask)<sup>[24]</sup>将分割结合作为头部加入跟踪网络中,获得更为精确的目标状态信息;交叉模态蒸馏的判别预测模型(Discriminative Model Prediction with Cross-Modal Distillation, CMD-DiMP)<sup>[25]</sup>利用知识蒸馏将可见光信息转化为红外信息,获取更高的跟踪精度;HUANG Yueping等<sup>[7]</sup>提出利用无监督方法优化特征提取网络,采用多通道滤波融合深度与浅层特征,获得更好的目标表达能力;LI Yingchao等<sup>[26]</sup>将红外图像转换至希尔伯特空间,并级联多个通道注意力机制获取更好的跟踪精度。但基于深度学习的网络需要大量的计算资源以及存储资源的支持,须具备较好的硬件平台才能达到实时表现。

针对上述问题,本文提出了自适应信息选择的变尺度相关滤波红外目标跟踪算法,为增强红外目标特征表达能力,在原有提取的手工特征上,重新提取尺度为2的梯度信息,丰富目标信息,并在响应图融合阶段加入权值,获取更高的跟踪精度;在训练过程中,将各个通道信息独立看待,对每个通道做空间选择,从而获取更为准确的表达能力;为适应目标边界框比例的变换,在传统的尺度估计后,构建变尺度滤波器,调整边界框比例,获取更为准确的边界框。

### 1 基于通道与空间信息选择的变尺度红外目标跟踪算法

提出的红外目标跟踪算法的整体结构如图1所示,红色连接线表示样本帧特征流向,绿色连接线表示当前帧特征流向,黄色箭头表示跟踪过程。 $w_t$ 表示本帧的滤波器系数, $w_{t-1}$ 表示上一帧求解的滤波器系数, $f$

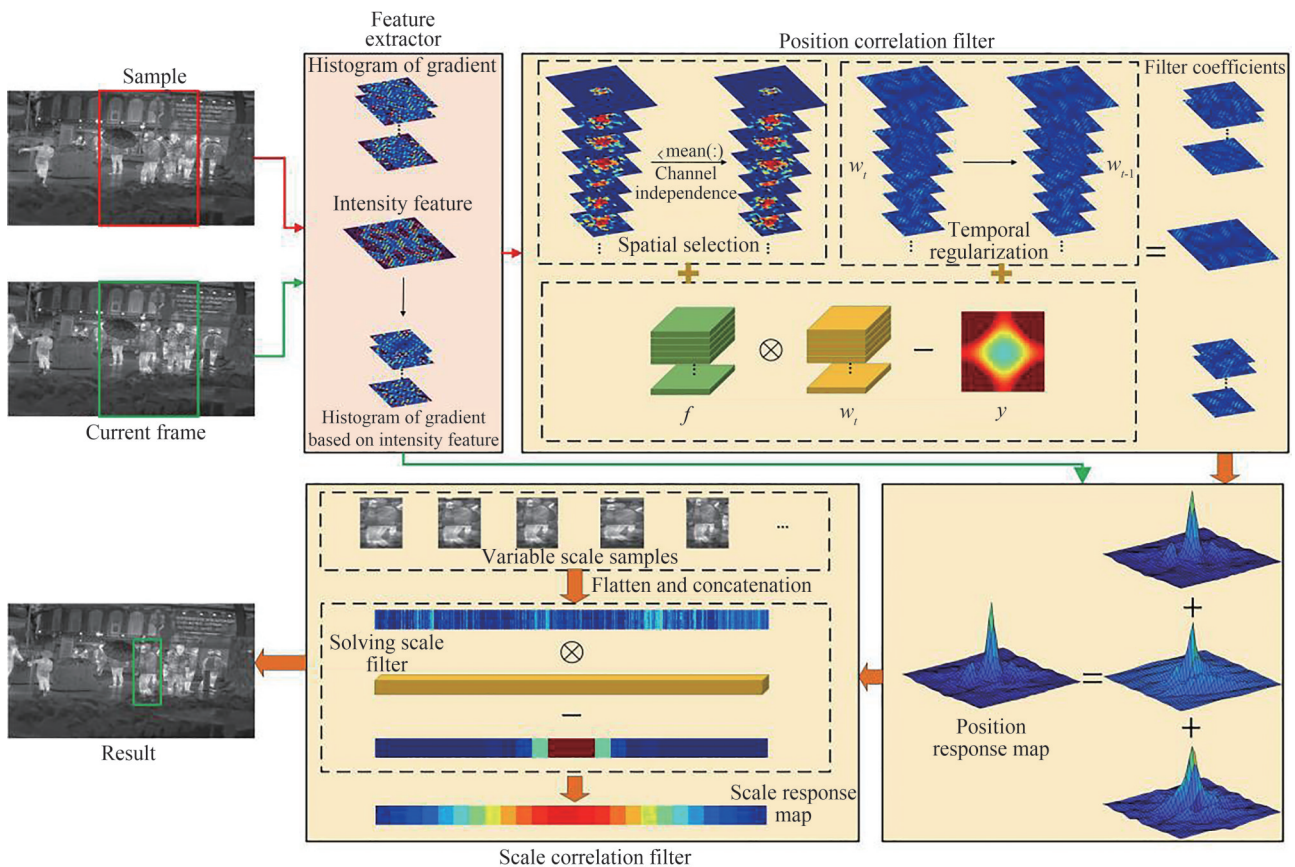


图1 算法整体框架

Fig.1 Overall algorithm architecture

表示样本特征, $y$ 表示理想响应函数。

算法共分为三个部分,分别是特征提取器、定位滤波器以及尺度滤波器。首先,在样本帧和当前帧中选取提取区域,将其送入特征提取器中提取特征信息,在特征提取器中,采用梯度直方图以及灰度特征作为基础特征,提取灰度特征后,在其基础上额外提取新的梯度直方图<sup>[27]</sup>,用于补充目标特征;然后,样本帧特征信息将会输入至定位滤波器中,在学习定位滤波器的阶段,加入时间正则项以及空间信息选择,以通道作为单位学习每个通道的空间权重,减少滤波器系数;其次,学习到的定位滤波器系数将会与当前帧的特征信息进行卷积运算,将多种特征对应的响应图加权融合,获取目标位置;最后,从获取的目标位置提取变尺度样本,根据所求的尺度滤波系数进行卷积,获取对应的尺度因子,从而确定边界框比例,获取最终目标的边界框。

### 1.1 特征提取器

由于红外图像分辨率低,可能会出现图像模糊、背景区分度不高等情况,如图2所示,这会导致原始提取到的手工特征难以表示目标信息,并且模糊部分过多,原始的手工特征感受野过小,所提取到的特征与原有特征难以匹配。为此,在提取的灰度特征图中重新提取梯度直方图,其具备更大的感受野,并且与原始的梯度直方图不同,能够提取到不同表示的梯度信息,增强对目标的表达能力。



图2 数据集图像  
Fig.2 The images in datasets

图3展示了部分特征可视化结果,为了方便展示,将特征均重新缩放至样本大小,其中梯度直方图部分与基于灰度特征的梯度直方图部分均是选用了同一梯度方向通道的可视化结果,可以看到虽然两种特征对应的梯度方向一致,但其所获得信息不同。由此完成了特征提取,所提取的特征将用于构建定位滤波器。

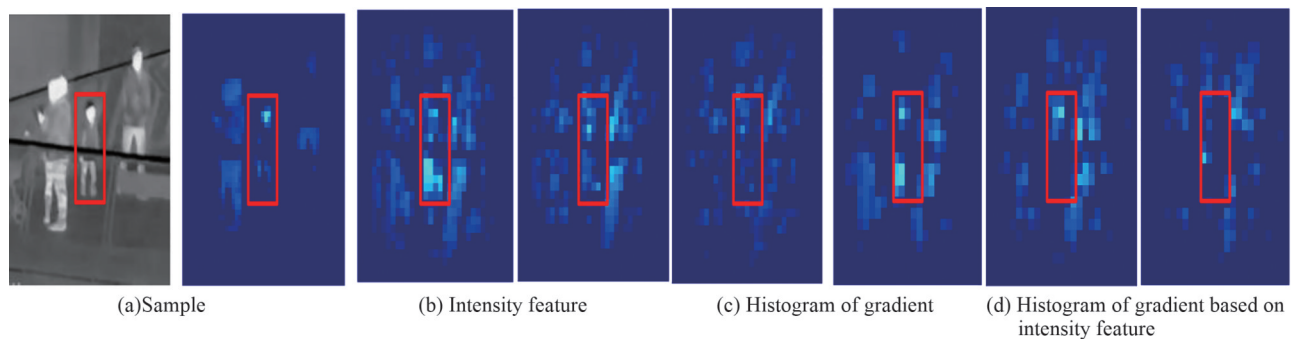


图3 特征可视化  
Fig.3 The visualization of features

### 1.2 定位滤波器

由于目标跟踪为时序任务,目标在整个跟踪过程中是运动的,其背景同样会出现变化,为了有效提高滤波器时序表达能力,在构建定位滤波器过程中加入时间正则项,有效应对目标运动、部分遮挡等情况带来的特征变化明显的问题。同时在图3中可以看到,并不是所有特征均能对目标有较好的表达能力,部分特征对目标表达能力较弱,因此在定位滤波器中加入了通道独立的空间信息选择,由此保留对目标定位有效的信息,从而获的更高的定位精度。

假设目标样本特征为  $f \in R^{H \times W \times N}$ , 其中  $H$  和  $W$  分别表示特征空间大小,  $N$  表示通道数量, 滤波器系数为  $w \in R^{H \times W \times N}$ , 理想响应函数为  $y \in R^{H \times W \times N}$ , 则定位滤波器公式为

$$L(w) = \arg \min_w \left\| \sum_{i=1}^N f^i \otimes w^i - y \right\|_F^2 + \lambda_1 \|w\|_F + \lambda_2 \|w - w_{t-1}\|_F^2 \quad (1)$$

式中, 上角标  $i$  表示第  $i$  个特征通道,  $\|\cdot\|_F$  表示 F 范数,  $\otimes$  表示循环卷积运算,  $w_{t-1}$  表示前一帧保存的滤波器系数,  $\lambda_1$  和  $\lambda_2$  分别表示空间信息选择项和时间正则项的正则化系数。  $\|w\|_F$  用于稀疏化滤波器系数, 时间正则项  $\|w - w_{t-1}\|_F^2$  则是在训练过程中, 让本帧训练出的滤波器系数接近上一帧的滤波器系数, 搭建时序的联系。

用增广拉格朗日对式(1)求解, 定义松弛变量  $w' = w$ , 式(1)重写为

$$L(w, w', \Gamma, \mu) = \arg \min_{w, w'} \left\| \sum_{i=1}^N f^i \otimes w^i - y \right\|_F^2 + \lambda_1 \|w'\|_F + \lambda_2 \|w - w_{t-1}\|_F^2 + \frac{\mu}{2} \left\| w - w' + \frac{\Gamma}{\mu} \right\|_F^2 \quad (2)$$

式中,  $\Gamma$  为拉格朗日乘子,  $\mu$  为惩罚系数。

将式(2)拆分, 分别求解  $w, w'$ , 则

$$\begin{cases} L(w, w', \Gamma, \mu) = \arg \min_w \left\| \sum_{i=1}^N f^i \otimes w^i - y \right\|_F^2 + \lambda_2 \|w - w_{t-1}\|_F^2 + \frac{\mu}{2} \left\| w - w' + \frac{\Gamma}{\mu} \right\|_F^2 \\ L(w, w', \Gamma, \mu) = \arg \min_{w'} \lambda_1 \|w'\|_F + \frac{\mu}{2} \left\| w - w' + \frac{\Gamma}{\mu} \right\|_F^2 \end{cases} \quad (3)$$

由于式(3)中含有卷积运算, 为了加快对  $w$  的求解速率, 将其转换至频域, 由此式(3)中的上式改写为

$$\arg \min_w \left\| \sum_{i=1}^N \hat{f}^i \odot \hat{w}^i \right\|_F^2 + \lambda_2 \|\hat{w} - \hat{w}_{t-1}\|_F^2 + \frac{\mu}{2} \left\| \hat{w} - \hat{w}' + \frac{\hat{\Gamma}}{\mu} \right\|_F^2 \quad (4)$$

式中, 符号  $\hat{\cdot}$  为变量的频域表示。

式(4)具备解析解, 根据 Sherman-Morrison 公式, 频域下滤波器系数  $\hat{w}$  可表示为

$$\hat{w} = \frac{1}{\lambda_2 + \mu} \left( \mathbf{I} - \frac{\hat{f}\hat{f}^T}{\lambda_2 + \frac{\mu}{2} + \hat{f}^T \hat{f}} \right) \left( \hat{f}\hat{y} + \frac{\mu}{2} \hat{w}' + \lambda_2 \hat{w}_{t-1} - \frac{\hat{\Gamma}}{2} \right) \quad (5)$$

根据式(5)求解出滤波器系数  $w$  后, 将其带入  $w'$  的求解式(3)中, 对  $w'$  继续求解。式(3)存在解析解, 因此,  $w'$  可表示为

$$w' = \max \left( 0, 1 - \frac{\lambda_1}{\mu \left\| w + \frac{\Gamma}{\mu} \right\|_F} \right) \left( \left\| w + \frac{\Gamma}{\mu} \right\|_F \right) \quad (6)$$

随后更新拉格朗日乘子以及惩罚因子, 更新公式为

$$\begin{cases} \Gamma = \Gamma + \mu(w - w') \\ \mu = \min(\alpha\mu, \mu_{\max}) \end{cases} \quad (7)$$

式中,  $\alpha$  为尺度因子, 用于平滑更新过程。

由此整个定位滤波器求解完成, 将其迭代两次, 获取滤波器系数  $w$  的近似解。在求解过程中, 获得第一次迭代结果后, 对求解的变量取均值, 将小于均值的变量置 0, 将其带入下一次迭代过程。

随后,在提取下一帧的图像特征后,利用求解的滤波器系数与特征进行卷积,获取位置滤波响应图为

$$S = f \otimes w \quad (8)$$

提取了3种特征,共计算获取3个位置滤波响应图,将其缩放至同一尺度,进行加权求和,获取最终响应图 $S_f$ 为

$$S_f = S_{\text{gray}} + S_{\text{hog}} + \beta S_{\text{gray\_hog}} \quad (9)$$

式中, $S_{\text{gray}}$ 、 $S_{\text{hog}}$ 、 $S_{\text{gray\_hog}}$ 分别表示灰度特征、梯度直方图以及基于灰度特征的梯度直方图的响应图, $\beta$ 为权重,设为0.01。基于灰度特征的梯度直方图用于补充目标特征信息,增强目标表达能力,但由于数据集中存在较多小目标,如图2(b),此特征具有较大的感受野,包含了较多的背景信息,使得目标信息淹没,无法发挥作用,因此将其权重设置较低。

### 1.3 尺度滤波器

目标在运动过程中,其尺度不断变化,边界框比例同样可能产生变化,但原有的目标尺度估计方式无法应对边界框比例变化问题,仅通过在定位阶段选择多个固定比例尺度采集尺度样本,对每个尺度样本均进行定位,选择响应值最高的结果作为估计尺度。

在原有尺度估计后,重新构建了用于变尺度估计的尺度滤波器,以原有的尺度估计结果作为基础,在其基础上重新采集变尺度样本,并将其展平为1维特征进行拼接。

假设, $s=1,2,\dots,M$ 表示边界框变化比例,在采集变尺度样本时,以本帧估计的目标位置作为中心点,设 $c \times c$ 为目标基础空间尺寸,变尺度采集的空间尺度为 $a^K c \times a^K c$ , $a$ 表示尺度因子,其中 $K = \left\lfloor \frac{2s - M - 1}{2} \right\rfloor$ , $\lfloor \cdot \rfloor$ 表示向下取整,在此将 $M$ 设置为5,但由于采集的是变尺度样本,实际样本数量为 $M^2$ ,即25个变尺度样本,长和宽分别对应5个尺度变化,并对其进行排列组合,将 $s$ 设置为1.01。

由于灰度特征无法用于准确判别样本的尺度变换,因此在尺度样本提取的过程中,仅利用窗口大小为4的梯度直方图作为尺度特征。在获取到样本后,构建尺度滤波器,公式为

$$L(w_s) = \|w_s \otimes f_s - y_s\|_F^2 + \lambda \|w_s\|_F^2 \quad (10)$$

下角标 $s$ 表示尺度滤波,将式(10)转换至频移,其具有解析解,解析解形式为

$$\hat{w}_s = \frac{\hat{f}_s \hat{y}_s}{\hat{f}_s \hat{f}_s^T + \lambda} \quad (11)$$

求解后对所获得的变尺度样本特征进行卷积,即

$$\hat{S}_s = \hat{f}_s \otimes \hat{w}_s \quad (12)$$

由此,获得多个变尺度下的尺度响应图,选择其中响应值最大的对应的尺度,即为所估计的尺度。

在原有的尺度估计基础上进行二次变尺度估计,构建的变尺度估计滤波器虽具备在边界比例变换的情况下的尺度,但单独使用并不能够有效提高算法精度,主要原因有:

1) 尺度滤波器若单独工作,在定位阶段将提取单尺度样本,缺乏尺度信息,从而使得定位滤波器精度下降,此现象与尺度自适应多特征跟踪器(Scale Adaptive with Multiple Features tracker, SAMF)<sup>[28]</sup>和判别式尺度空间跟踪器(Discriminative Scale Space Tracker, DSST)<sup>[29]</sup>两种算法相同,两种算法采用相同的基础定位滤波器,但SAMF在定位阶段采集多尺度样本,在获取目标边界框估计能力的同时,也提升了算法定位精度,DSST则设计了额外的固定边界框比例的尺度滤波器,其定位精度低于SAMF。

2) 目标在整个运动过程中前后两帧的边界框比例变化不大,采用定尺度估计的方法可以判别目标框整体变化情况,本文所设计的变尺度滤波器则进行进一步微调,在整体估计后的基础上,重新采集变尺度样本,能够有效减小变尺度滤波器中样本采集的数量,降低变尺度滤波器的计算负担。

所提算法的伪代码如表1所示。

表1 算法伪代码  
Table 1 The pseudo-code of algorithm

---

Algorithm 1 The tracking process for one time-step  $t$

---

- 1: Inputs: last frame weights  $w_{t-1}$ , ideal response map  $y$ , ideal response map for scale estimation  $y_s$ , current frame feature  $f$ , current frame feature for scale estimation  $f_s$ , last frame feature for scale estimation  $f_s^{t-1}$ , penalty parameter  $\mu$
- 2: if  $i = 1$  then
- 3:  $w_{t-1} \leftarrow$  zeros matrix,  $\mathbf{F} \leftarrow$  zeros matrix
- 4: for  $i=1:N_{\text{iteration}}$  do
- 4:  $w \leftarrow$  equation 5( $f, w_{t-1}, y, \mathbf{F}$ )
- 5:  $w' \leftarrow$  equation 6( $w, y, \mathbf{F}$ )
- 5: update  $\mu$ , update  $\mathbf{F}$
- 6: end for
- 7: else
- 8:  $S \leftarrow w \otimes f$
- 9:  $S_s \leftarrow w_s \otimes f_s$
- 10: for  $i=1:N_{\text{iteration}}$  do
- 11:  $w \leftarrow$  equation 5( $f, w_{t-1}, y, \mathbf{F}$ )
- 12:  $w' \leftarrow$  equation 6( $w, y, \mathbf{F}$ )
- 13: update  $\mu$ , update  $\mathbf{F}$
- 14: end for
- 15:  $w_s \leftarrow$  equation 11( $y_s, f_s$ )
- 16: end if

---

## 2 实验

为了验证算法的有效性,选择LSOTB-TIR数据集<sup>[30]</sup>与PTB-TIR数据集<sup>[31]</sup>作为测试集。LSOTB-TIR数据集共包含了120个测试序列,序列最大长度为2110帧,最小长度为105帧,平均长度为684帧,共计82000帧图像,包含了22种目标类别,12种类别属性,对序列采集场景进行了标注,类别与标注信息如表2和表3所示。

PTB-TIR数据集是红外行人跟踪数据集,目标尺寸变化较小,包含60个测试序列,9种类别属性,共

表2 场景定义  
Table 2 The definition of scenarios

Scenario	Definition
VS	The videos come from a surveillance camera
HH	The videos are shot from a hand-held camera
DS	The videos are captured from a drone-mounted camera
VM	The videos come from a vehicle-mounted camera

表3 序列属性  
Table 3 The attributes of sequences

Attribute	Definition
TC	Two same intensity targets cross each other
IV	The target intensity is changed during tracking
DIS	Existing the intra-class object near the target
BC	The background has a similar appearance to the target
DEF	The target is deformable during tracking
OCC	The target is partly or fully occluded during tracking
OV	The target partly or fully leaves the image
SC	The ratio of the target size is out of the range [0.5, 2]
FM	The target moves more than 20 pixels
MB	The target is blurry due to the target or camera motion
LR	The target size is lower than 800 pixels
ARV	The target aspect ratio is out of the range [0.5, 2]

30 029 帧图像,最小序列长度为 50 帧,最大序列长度为 1 792 帧。共选取了 8 个算法作为对比算法,用于展现本文所提算法的优越性,其中有效卷积算子跟踪器(Efficient Convolution Operators for tracking, ECO)、自适应学习判别式相关滤波(Learning Adaptive Discriminative Correlation Filters, LADCF)、自适应通道选择跟踪器(Adaptive Channel Selection Mechanism for Discriminative Correlation Filters, ACSDCF)、基于空间正则项的判别式相关滤波器(Spatially Regularized Discriminative Correlation Filters, SRDCF)<sup>[32]</sup>为相关滤波算法,仅使用了手工特征,即梯度直方图以及灰度特征,TADT、SiamMask、多域卷积神经网络跟踪算法(Multi-Domain convolutional neural Networks for visual tracking, MDNet)<sup>[33]</sup>和 SiamFC 为深度学习算法。LADCF 与 ACSDCF 的结果均采用作者公布的原始代码在数据集中测试获得,其余对比算法的结果均来自于数据集官方公布的测试结果。实验均在 MATLAB2018a 上进行,实验平台处理器为 Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz 2.81 GHz,内存为 8 G。

## 2.1 数据集对比实验

### 2.1.1 LSOTB-TIR 数据集测试结果

依据 LSOTB-TIR 数据集官方的测评指标,共采用 3 种指标对算法进行对比,分别为精确度、归一化精确度以及成功率。其中精确度与归一化精确度均以中心位置误差进行衡量,中心位置误差计算公式为

$$d = \sqrt{(x_g - x_p)^2 + (y_g - y_p)^2} \quad (13)$$

式中, $x$  和  $y$  分别表示中心点  $x$  坐标和  $y$  坐标,下角标  $g$  和  $p$  分别表示真值和预测值。精确度和归一化精确度则是根据不同的中心位置误差作为阈值,统计全部序列结果计算获得。

成功率根据重叠率(Intersection of Union, IoU)计算获取,计算公式为

$$\text{IoU} = \frac{B_g \cap B_p}{B_g \cup B_p} \quad (14)$$

式中, $B$  表示边界框。与精确度和归一化精确度类似,成功率同样依据不同 IOU 阈值计算获取。

3 种指标的曲线如图 4 所示。可以看到本文所提算法在 3 个指标中均取得了最优的表现。与相关滤波

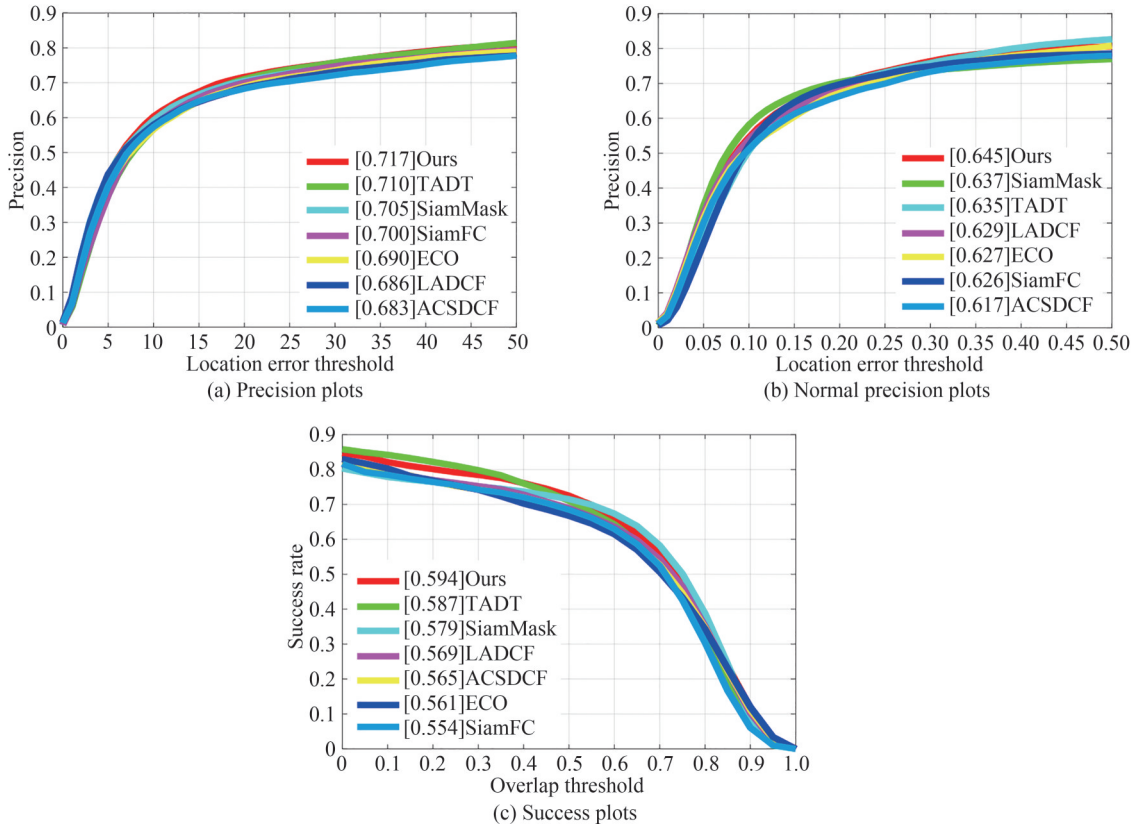


图4 LSOTB-TIR 数据集测试结果

Fig.4 The evaluation results of LSOTB-TIR dataset

算法对比,本文算法在精确度上超过第二名的ECO算法2.7%,相较于LADCF和ACSDCF分别提高了3.1%和3.4%;在归一化精确度中,本文算法相较于第二名的LADCF算法提高了1.6%;在成功率曲线中,本文算法相较于LADCF和ACSDCF分别提高了2.5%和2.9%。证明了进行通道独立的空间信息选择策略的有效性,同时额外提取的基于灰度特征的梯度直方图信息能够进一步提供目标信息,变尺度滤波器使得边界框估计效果得到了提升。

与深度学习算法对比,本文算法相较于TADT在3种指标下分别提高了0.7%、1%和0.7%,比SiamMask分别高出了1.2%、0.8%和1.5%,该算法在仅利用手工特征的基础上,可以超过部分深度学习算法,同时,算法无需训练,在线推理过程中也仅利用CPU即可,所需计算资源较少。

### 2.1.2 PTB-TIR数据集测试结果

PTB-TIR数据集采用了精确度和成功率作为评价指标,测试结果曲线如图5所示。

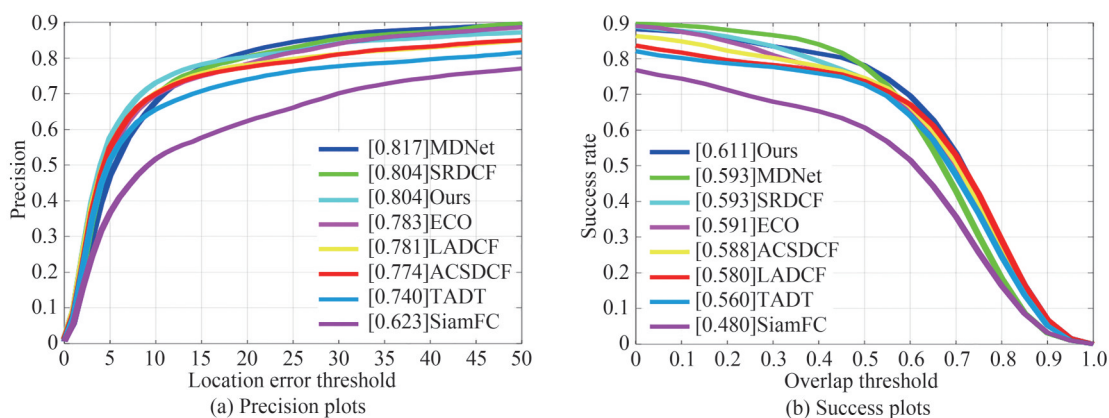


图5 PTB-TIR数据集测试结果

Fig.5 The evaluation results of PTB-TIR dataset

可以看到,本文所提算法在成功率指标中取得了优异的表现。与相关滤波算法相比,本文算法与SRDCF在精确度指标中取得了同样的表现,而成功率提高了1.8%。相较于LADCF和ACSDCF,本文算法的精确度指标分别提高了2.3%和3.0%,成功率分别提高了2.3%和3.1%。由于PTB-TIR数据集中目标的边界框比例变化较小,且比例变换属性下的序列数量少,因此变尺度滤波器对算法提高较小,数据集中更注重前景与背景的区别,因此所提的特征提取与信息选择策略发挥了更为重要的作用。

相较于深度学习算法而言,MDNet的精确度高于本文算法1.3%,MDNet依据于所获取的初始样本,在线对网络中的头部进行更新,速率较慢,但能够取得更高的分类精度,而本文算法的成功率相较于MDNet提高了0.7%,说明MDNet虽对前景与背景区分效果更好,但边界框回归效果较差。本文算法在仅利用手工特征的基础上,可以获取更为准确的边界框。

## 2.2 消融学习

为了验证各个模块的有效性,在LSOTB-TIR数据集中进行实验,以LADCF和ACSDCF作为基础算法,本文所提算法和这两种算法所采用的优化过程相似,LADCF仅考虑空间信息,对每个通道采用同样的空间采样模板,而ACSDCF则认为通道应该独立看到,仅对通道做了选择,本文算法则是将在各个通道中对空间信息做不同的选择方式。消融实验结果如表4所示。

从表4中可以看到,采用通道独立的空间信息选择的方法,可以达到0.707的精确度和0.589的成功率,相较于LADCF和ACSDCF分别高出了0.21、0.24的精确度以及0.2、0.24的成功率,同时每秒帧数(Frames Per Second, FPS)均高于两者,证明了本文采用的信息选择方式更为有效,所学习到的滤波器更为稀疏,加快整体优化过程;而在加入变尺度估计后,精确度提高了0.06,成功率提高了0.02,FPS仅下降了1.49,说明变尺度滤波器可以在原有尺度滤波的基础上进一步优化边界框估计效果;加入基于灰度特征的梯度直方图后,精确度提高至0.717,成功率提高至0.594,FPS下降了3.49,说明加入的新特征可以丰富目标表达的能力,有效提高跟踪精度,并且本文所提算法最终可达34.85的FPS,达到实时处理效果。



表 4 消融实验结果  
Table 4 The results of ablation experiments

Name	Information selection	Variable scale estimation	Feature extractor	Precision	Norm. precision	Success rate	FPS
LADCF				0.686	0.629	0.569	34.43
ACSDCF				0.683	0.617	0.565	35.96
Ours_1	✓			0.707	0.635	0.589	39.83
Ours_2	✓	✓		0.713	0.639	0.591	38.34
Ours	✓	✓	✓	0.717	0.645	0.594	34.85

### 2.3 超参实验

为了验证超参数对本文算法的影响,以 LSOTB-TIR 数据集作为测试集,对定位融合权重 $\beta$ 、尺度样本数量 $M$ 两个超参数进行实验分析。

#### 2.3.1 定位融合权重

对定位融合权重进行消融实验时,将尺度样本数量固定为 5,其余参数不变,结果如图 6 所示。可以看到当定位融合权重 $\beta$ 为 0.001 时效果最好。当权重增大时,曲线下面积 (Area Under the Curve, AUC) 降低,分析认为是所提取的基于灰度特征的梯度直方图虽能补充目标信息,但由于其感受野较大,使得其所提取特征中包含了部分背景信息,致使响应图主峰周围存在较多的次峰,过多融合相应的响应图,使得主峰不再突出,降低了目标定位的精度。而当权重进一步降低时,基于灰度特征的梯度直方图所能提供的信息过少,无法达到较好的补充效果,因此与未加入基于灰度特征的梯度直方图的表现无明显区别。

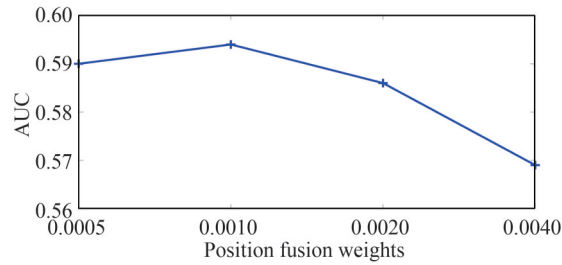


图 6 定位融合权重消融实验结果  
Fig.6 The ablation result of location fusion weight

#### 2.3.2 尺度样本数目

对尺度样本数量进行消融实验时,将定位融合权重固定为 0.001,其余参数不变,结果如图 7 所示。可以看到,当尺度样本数量 $M$ 选择为 5 时,效果最好。当尺度样本数目设置过多时,尺度池中存在长宽比例变化较为严重的尺度,在目标定位位置存在一定偏差或仅定位到目标部分信息时,易对目标的部分区域进行边界框估计,从而获取错误的估计结果,因此导致 AUC 下降。而当尺度样本数目较少时,无法适应全部的变化情况,边界框估计结果不准确。

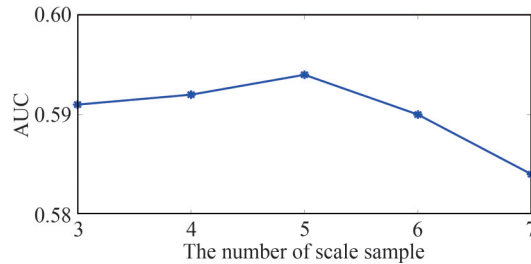


图 7 尺度样本数目消融实验结果  
Fig.7 The ablation result of the number of scales

## 2.4 定性分析

在LSOTB-TIR数据集中选择了4个序列,将跟踪结果进行了展示。图8中,蓝色框为本文算法跟踪结果,红色框为ACSDCF跟踪结果,绿色框为LADCF跟踪结果,黑色框为ECO跟踪结果。

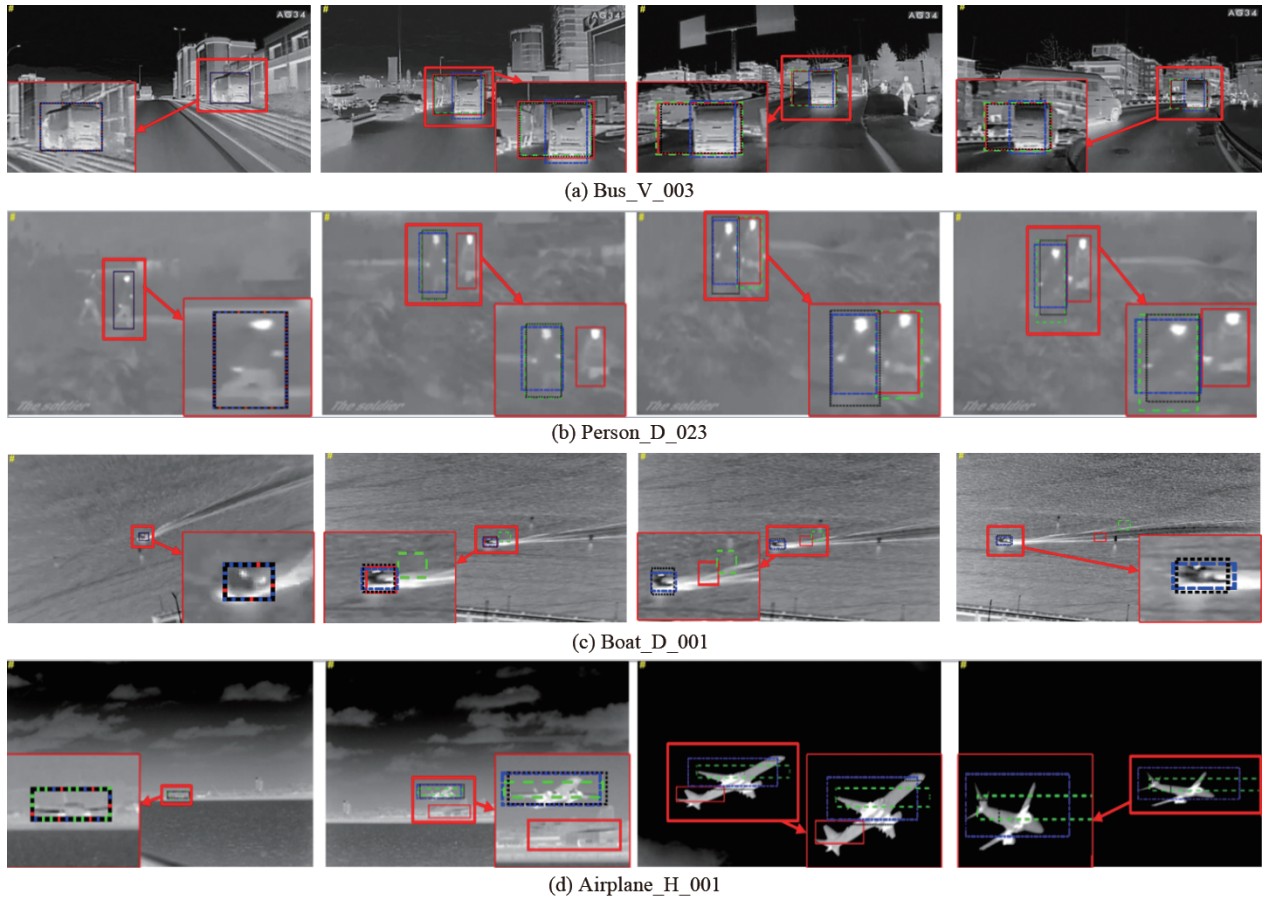


图8 4个序列结果展示

Fig.8 The tracking results on four sequences

在bus\_V\_003序列中,目标车辆处于运动状态,其视角在整个运动过程中发生了变化,从而导致目标车辆在图像中的边界框比例发生了变化,可以看到本文算法可以适应目标车辆的比例变化情况。在第2、3、4列的图像中,其余算法虽然依然可以跟踪到目标车辆,但其边界框估计结果保留了原始长宽比例,本文算法则可根据目标车辆的比例变化改变所估计的边界框长宽比例,其跟踪效果更好。

在person\_D\_023序列中,目标周围存在同类型目标,且在运动过程中两者在图像中出现了重叠。在第2列的图像中,相似物距离目标较近,ACSDCF跟踪产生漂移,跟踪错误目标;在第3列相似物与目标在图像中重叠后分开,LADCF跟踪错误,而在后续中,LADCF重新跟踪目标,可以看到本文算法和ECO可一直准确跟踪目标;但在第4列的结果中,本文算法的尺度估计更为准确。

在boat\_D\_001序列中,目标快速移动,并在运动过程中背景存在干扰物。在第2列和第3列图像中可以看到,目标快速移动,同时镜头发生位移,导致目标运动距离较大,ACSDCF和LADCF两种算法均跟踪失败,而本文算法和ECO算法跟踪成功。

在airplane\_H\_001序列中,目标在起飞阶段与背景区分度较低,而在飞行过程中其尺度发生变化。可以看到在第2列的图像中,ACSDCF未能准确区分背景与前景,导致跟踪漂移,在后续序列中重新跟踪到目标;在第3列中,目标尺度发生快速变化,所有算法虽能成功跟踪目标,但尺度估计效果一般,本文算法和ECO算法估计效果更好,ACSDCF由于尺度估计错误未能跟踪目标主体;在第4列中,可以看到ACSDCF

由于尺度估计偏差严重,在后续序列中未能跟踪目标,其余算法均可以跟踪目标,但可以看到本文算法的尺度估计效果更好。

本文算法通过在灰度特征图中额外提取梯度信息,增强目标表达能力,并且在定位滤波器训练过程中,设计通道独立的空间信息选择,使得所设计出的定位滤波器更具判别能力,能够在目标快速运动、背景存在同类物体或相似目标时准确区分目标,如序列 person\_D\_023 和 boat\_D\_001;设计的变尺度滤波器能够有效应对目标由于势角变换等情况导致的边界框比例变换的问题,能够自适应调整比例变换,如序列 bus\_V\_003 和 airplane\_H\_001,证明了本文算法改进的有效性。

### 3 结论

本文提出了自适应信息选择的变尺度相关滤波红外目标跟踪算法,能够在红外跟踪序列下准确跟踪到目标。在提取红外图像特征的过程中,为了弥补目标信息少以及图像模糊、分辨率较低等带来目标特征不明显的问题,在原有的灰度特征基础上重新提取了梯度直方图信息,丰富目标表现能力;在滤波器训练中则是加入时间正则项与通道独立的空间信息选择,将每个通道信息独立看待,同时对空间信息进行独立选择,有效保留判别信息,增强滤波器判别能力;在原有尺度滤波的基础上,将目标的长宽信息分别看待,设计独立的变尺度滤波器,采集多尺度样本信息,从而适应目标在运动过程中比例变化的问题。

本文算法在 LSOTB-TIR 数据集中跟踪效果显著,精确度和成功率分别达到了 71.7% 和 59.4%,相较于同样进行信息选取的滤波器算法 LADCF 和 ACSDCF,精确度分别提高了 3.1% 和 3.4%,成功率分别提高了 2.5% 和 2.9%,同时超过了部分深度学习跟踪算法的表现,仅在 CPU 上运行,达到了 34.85 帧/s 的运行速率,具备良好的实时性以及跟踪精度。

本文算法依然存在一定的局限性。算法虽能够应对尺度比例变化问题,但当结合尺度的定位置信度较高时,变尺度滤波器并不能完全发挥其作用,仅能保持原有目标比例;在背景干扰物较多的情况下,传统手工特征难以准确区分目标,与基于红外的深度网络所提取的特征相比存在一定差异;算法中存在较多的超参,虽大部分参数在不同数据集下敏感度低,无需变化,但需要人为设定,依赖经验值和实验数据。

### 参考文献

- [1] ONDRASOVIC M, TARABEK P. Siamese visual object tracking: a survey[J]. IEEE Access, 2021: 110149-110172.
- [2] XU Guoxia, WANG Hao, ZHAO Meng, et al. Learning the distribution-based temporal knowledge with low rank response reasoning for UAV visual tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 1(1): 1-11.
- [3] YUAN Yubin, WU Yiquan, ZHAO Langyue, et al. Research progress of UAV aerial video multi-object detection and tracking based on deep learning [J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(18): 6-36.  
苑玉彬, 吴一全, 赵朗月, 等. 基于深度学习的无人机航拍视频多目标检测与跟踪研究进展[J]. 航空学报, 2023, 44(18): 6-36.
- [4] LI Gong, ZHAO Wei, LIU Peng, et al. Smooth-IoU loss for bounding box regression in visual tracking [J]. Acta Automatica Sinica, 2023, 49(2): 288-306.  
李功, 赵巍, 刘鹏, 等. 一种用于目标跟踪边界框回归的光滑 IoU 损失[J]. 自动化学报, 2023, 49(2): 288-306.
- [5] LU Jian, DENG Bo, QUE Longcheng. Stable object tracking method for complex infrared ground environment[J]. Acta Photonica Sinica, 2019, 48(10): 1010001.  
吕坚, 邓博, 阙隆成. 复杂红外地面环境下的稳定目标跟踪方法[J]. 光子学报, 2019, 48(10): 1010001.
- [6] LIU Hui, HE Yong, HE Boxia, et al. Infrared target tracking algorithm based on multiple feature fusion and region of interest prediction[J]. Acta Photonica Sinica, 2019, 48(7): 0710004.  
刘辉, 何勇, 何博侠, 等. 基于多特征融合与 ROI 预测的红外目标跟踪算法[J]. 光子学报, 2019, 48(7): 0710004.
- [7] HUANG Yueping, HE Yujie, LU Ruitao, et al. Thermal infrared object tracking via unsupervised deep correlation filters [J]. Digital Signal Processing, 2022: 103432.
- [8] GUNDOGDU E, OZKAN H, DEMIR H S, et al. Comparison of infrared and visible imagery for object tracking: Toward trackers with superior IR performance[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015: 1-9.

- [9] WANG Fasheng, YIN Shuangshuang, MBELWA J T, et al. Context and saliency aware correlation filter for visual tracking[J]. *Multimedia Tools and Applications*, 2022, 81(19): 27879-27893.
- [10] ZHANG Jianming, YUAN Tingyu, HE Yaoqi, et al. A background-aware correlation filter with adaptive saliency-aware regularization for visual tracking[J]. *Neural Computing and Applications*, 2022, 34(8): 6359-6376.
- [11] BOURAFFA T, YAN L, FENG Z, et al. Context-aware correlation filter learning toward peak strength for visual tracking[J]. *IEEE Transactions on Cybernetics*, 2021, 51(10): 5105-5115.
- [12] WANG Yuhui, XU Zhiyuan, YE Demao. A target tracking algorithm based on local feature segmentation[J]. *Laser Technology*, 2019, 43(4): 569-573.  
王宇慧,徐志远,叶德茂.一种基于局部部特征分块的目标跟踪算法[J].*激光技术*, 2019, 43(4): 569-573.
- [13] 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(CVPR)*, 2017: 6931-6939.
- [14] XU Tianyang, FENG Zhenhua, WU Xiaojun, et al. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking [J]. *IEEE Transactions on Image Processing*, 2019, 28(11): 5596-5609.
- [15] LIANG Yanjie, LIU Yi, YAN Yan, et al. Robust visual tracking via spatio-temporal adaptive and channel selective correlation filters[J]. *Pattern Recognition*, 2020, 112: 107738.
- [16] ZHU Xuefeng, WU Xiaojun, XU Tianyang, et al. Robust visual object tracking via adaptive attribute-aware discriminative correlation filters[J]. *IEEE Transactions on Multimedia*, 2021(24):15457-15466.
- [17] GUNDOGDU E, KOC A, SOLMAZ B, et al. Evaluation of feature channels for correlation-filter-based visual object tracking in infrared spectrum [C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, USA, 2016: 290-298.
- [18] YU Tianwen, MO Bo, LIU Fuxiang. Robust thermal infrared object tracking with continuous correlation filters and adaptive feature fusion[J]. *Infrared Physics & Technology*, 2019,(98): 69-81.
- [19] ZHANG He, YIN Zhaoyuan, ZHANG Hanling. Thermal infrared object tracking using correlation filters improved by level set[J]. *Signal, Image and Video Processing*, 2023(17): 791-797.
- [20] YUAN Di, SHU Xiu, LI Qiao, et al. Robust thermal infrared tracking via an adaptively multi-feature fusion model[J]. *Neural Computing and Application*. 2023(35): 3423-3434.
- [21] CAO Ziang, FU Changhong, YE Junjie, et al. Hift: hierarchical feature transformer for aerial tracking[C].*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 15457-15466.
- [22] GAO Shenyuan, ZHOU Chunluan, MA Chao, et al. Aiatrack: attention in attention for transformer visual tracking[C]. *17th European Conference on Computer Vision*, 2022: 146-164.
- [23] LI Xin, MA Chao, WU Baoyuan, et al. Target-aware deep tracking[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 1369-1378.
- [24] WANG Qiang, ZHANG Li, BERTINETTO L, et al. Fast online object tracking and segmentation: a unifying approach [C].*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 1328-1338.
- [25] SUN Jingxian, ZHANG Lichao, ZHA Yufei, et al. Unsupervised cross-modal distillation for thermal infrared tracking [C]. *Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 2262-2270.
- [26] LI Yingchao, YANG Shuai. Infrared small object tracking based on Att-Siam network[J]. *IEEE Access*, 2022, 10: 133766-133777.
- [27] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2005: 886-893.
- [28] LI Yang, ZHU Jianke. A scale adaptive kernel correlation filter tracker with feature integration [C]. *13th European Conference on Computer Vision(ECCV)*, 2015: 254-265.
- [29] DANELLJAN M, KHAN F S, FELSBERG M, et al. Accurate scale estimation for robust visual tracking[C]. *British Machine Vision Conference 2014(BMVC)*, 2014: 1-11.
- [30] LIU Qiao, LI Xin, HE Zhenyu, et al. LSOTB-TIR: a large-scale high-diversity thermal infrared object tracking benchmark[C].*Proceedings of the 28th ACM International Conference on Multimedia*, 2020: 1-14.
- [31] LIU Qiao, HE Zhenyu, LI Xin, et al. PTB-TIR: a thermal infrared pedestrian tracking benchmark[J]. *IEEE Trans. Multimedia*, 2019, 22(3): 666-675.
- [32] DANELLHAN M, HAGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking [C]. *2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015: 4310-4318.
- [33] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking [C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, 2016: 4293-4302.

## Adaptive Information Selection for Infrared Object Tracking with Variable Scale Correlation Filter

SUN Mengyu<sup>1</sup>, WANG Peng<sup>2</sup>, XU Junqi<sup>1</sup>, LI Xiaoyan<sup>2</sup>, GAO Hui<sup>2</sup>, DI Ruohai<sup>2</sup>  
(1 School of Opto-electronical Engineering, Xi'an Technological University, Xi'an 710021, China)  
(2 Electronic Information Engineering, Xi'an Technological University, Xi'an 710021, China)

**Abstract:** While there is less information contained in the infrared image, it still has problems with image blur and noise. It caused many difficulties in infrared visual target tracking. The discriminative correlation filter is a reliable method for tracking objects that can be trained online on an embedded hardware platform. But the variable scale of the object cannot be efficiently solved. We proposed an adaptive information selection for infrared object tracking with a variable scale correlation filter to solve the problem mentioned earlier. The proposed algorithm is divided into three parts, one for feature extractor, one for location filter, and one for scale filter. Three features are extracted by the feature extractor to represent the object. The insensitive feature and the histogram of gradient are chosen as the based features. The insensitive feature represents the intensity of the object at its current position in the infrared image. The information extracted from the gradient histogram, which contains more information about the shape of the target, represents the regions with sudden changes in the infrared image such as the edge and corner of the target. To enhance the object's representation, we generate a new histogram of gradients based on the insensitive feature, which has various representations of the object. After that, the position filter will receive the sample frame feature information. We train three position filters depend on three types of feature, respectively. In the training phase of the position filter, temporal regularization, spatial regularization and spatial information selection are added during optimization. The temporal regularization is used to enforce correlation filter more similarity to the last coefficients, which makes correlation filter more fit on the variation of the object. The spatial regularization is limited the correlation filter concentrate on the region of the object when training. For reducing redundancy of the object's information and speeding up the training process, we set different thresholds for coefficients. When the coefficients are lower than the threshold, we set it to zero to reduce the number of coefficients. The spatial coefficient of each channel is treated individually. Then, the position filter will be convolved with the feature of the current frame. We compute three response maps according to the different features. The response map corresponding to various features will be weighted and sum together to find the max value position which is the estimated object position. The weights of the various features are set to the hyperparameter. The weight of the new feature is set smaller than the other base features, because of the larger receptive field. For better representation of the object's scale change, we extract the variable scale sample from the estimated object position. The ratio of the height and width of the variable scale sample is different. We flatten the features of all variable scale samples and concatenate them together along with the spatial dimension as the scale training features. The corresponding scale factor is obtained by convolution according to the obtained scale filter coefficient, so as to determine the boundary ratio and obtain the boundary of the object. We choose LSOTB-TIR dataset and PTB-TIR dataset as the test dataset. The proposed algorithm can reach 34.85 frames per seconds on the central processing unit platform. On the LSOTB-TIR data set, the accuracy and success rate of the proposed algorithm reached 71.3% and 59.4% on the LSOTB-TIR dataset, and 80.4% and 61.1% on the PTB-TIR dataset, respectively. We analysis each component in our algorithm and find the spatial selection contributes more for precision. We also visualize the result of our method, which demonstrates our method can track the object consistently and can adapt the change of boundary ratio.

**Key words:** Infrared night vision technology; Target tracking; Correlation filter; Infrared image processing; Variable scale filter; Sparse representation

**OCIS Codes:** 100.4999; 100.4997; 040.1880