

# 多模态深层次高置信度融合跟踪算法<sup>①</sup>



高伟<sup>1</sup>, 薛杉<sup>1</sup>, 胡秋霞<sup>1</sup>, 李嘉琦<sup>1</sup>, 田杰<sup>1</sup>, 饶晔<sup>1</sup>, 杨举<sup>2</sup>

<sup>1</sup>(西安航空学院 计算机学院, 西安 710077)

<sup>2</sup>(中国电子科技集团第十五研究所 西安研发中心, 西安 710005)

通信作者: 薛杉, E-mail: xueshan@xaau.edu.cn

**摘要:** 为解决单目标跟踪中因目标外观及环境变化导致的跟踪失败问题, 提出一种多模态深层次高置信度融合跟踪算法. 首先构建目标颜色模型和基于双线性插值 HOG 特征形状模型的高维度多模态模型, 之后对候选目标利用粒子滤波进行搜索. 针对模型融合的难点, 通过准确量化形状和颜色模型多种置信度并设计高置信度融合准则, 以实现该多模态模型中不同置信度的深层次自适应加权平衡融合. 最后针对模型更新参数固定的问题, 设计非线性分级平衡更新策略. 经过在 OTB-2015 数据集上的测试, 发现该算法的平均 *CLE* 和 *OS* 在所有参照算法表现中均表现最佳, 其值分别为 30.57 和 0.609. 此外, 其 FPS 为 15.67, 满足了跟踪算法在一般情况下的实时性要求. 在某些常见的特定场景中, 其精确率、成功率指标在多数情况下的表现也超过了同类顶尖算法.

**关键词:** 视觉目标跟踪; 多模态; 置信度融合; 深层次加权; 分级平衡更新

引用格式: 高伟, 薛杉, 胡秋霞, 李嘉琦, 田杰, 饶晔, 杨举. 多模态深层次高置信度融合跟踪算法. 计算机系统应用, 2024, 33(9): 153-163. <http://www.c-s-a.org.cn/1003-3254/9633.html>

## Multi-modal Deep-level High-confidence Fusion Tracking Algorithm

GAO Wei<sup>1</sup>, XUE Shan<sup>1</sup>, HU Qiu-Xia<sup>1</sup>, LI Jia-Qi<sup>1</sup>, TIAN Jie<sup>1</sup>, RAO Ye<sup>1</sup>, YANG Ju<sup>2</sup>

<sup>1</sup>(School of Computer Science, Xi'an Aeronautical University, Xi'an 710077, China)

<sup>2</sup>(Research and Development Center of Xi'an, the 15th Research Institute of China Electronics Technology Group, Xi'an 710005, China)

**Abstract:** This study proposes a multi-modal deep-level high-confidence fusion tracking algorithm in response to the tracking failure issues caused by changes in target appearance and environment in single-target tracking applications. First, a high-dimensional multi-modal model is constructed utilizing the target's color model combined with a shape model based on bilinear interpolation HOG features. Then, candidate targets are searched using particle filtering. The challenge posed by model fusion is addressed by scrupulously quantifying a range of confidences in shape and color models. This is followed by the introduction of a high-confidence fusion criterion, which enables a deeply-adaptive, weighted, and balanced fusion with different confidence levels in the multi-modal model. To counter the issue of static model update parameters, a nonlinear, graded balanced update strategy is designed. Upon testing on the OTB-2015 dataset, this algorithm's average *CLE* and *OS* metrics demonstrated superior performance compared to all reference algorithms, with values of 30.57 and 0.609, respectively. Moreover, with an FPS of 15.67, the algorithm fulfills the real-time operation requirements inherent in tracking algorithms under most conditions. Notably, in some common specific scenarios, the accuracy and success rate of the algorithm also outperform the top-tier algorithms in most cases.

**Key words:** visual object tracking; multi-modal; confidence fusion; deep-level weighting; hierarchical balanced updating

① 基金项目: 陕西省自然科学基金面上项目 (2023JCYB194, 2024JCYBMS169); 西安航空学院校级科研基金 (2023KY1205)

收稿时间: 2024-03-25; 修改时间: 2024-04-19; 采用时间: 2024-05-06; csa 在线出版时间: 2024-07-24

CNKI 网络首发时间: 2024-07-25

在单目标跟踪<sup>[1]</sup>中,目标模型的设计<sup>[2,3]</sup>、融合<sup>[4,5]</sup>和更新<sup>[6]</sup>是跟踪问题的关键.在不同的跟踪场景下,如何设计模型是首要考虑因素.其次在进行模型融合和更新时,需要重点关注置信度量化与融合策略以及模型更新策略,这是目标跟踪研究的关键所在.

单模态建模容易降低算法的准确性和鲁棒性,且过分依赖于特定数据.多模态方法通过结合图像、激光雷达和深度图像等多种数据<sup>[7-10]</sup>,提高了处理跟踪问题的能力.例如, Gong 等人提出的 DCFE 算法<sup>[11]</sup>利用颜色特征和注意力机制来强化目标区域的特征,提高检测性能.虽可有效提高目标检测能力,但算法的复杂性限制了其计算效率,并且在非特定(如非水下)场景中可能不够有效,导致信息丢失.

模型融合的成功关键在于选择合适的融合方法,这直接决定了目标跟踪的性能.一种方法是集成学习<sup>[12]</sup>,如 Zhao 等人提出的 GADE-Net 算法<sup>[13]</sup>,它通过结合不同模型的参数来提高目标检测的准确度,并减少决策随机性和过拟合问题,但这增加了计算复杂度和对高质量标注数据的需求.另一种方法是使用多尺度神经网络,比如孙备等人提出的 MUC 算法<sup>[14-16]</sup>,该算法引入了多尺度互交叉注意力模块以提升伪装目标检测的精度,并融合了多种数据类型以更准确地定位目标.尽管这些方法在提高检测准确性方面表现出色,但在需要快速反应且计算资源有限的场景下,它们的优势可能不会完全显现.

目标跟踪在实际应用中常受到光照、遮挡和图像质量变化的影响,这些因素都可能导致目标外观的显著变化,因此模型的及时更新变得至关重要.为应对这一挑战,一些研究提出了不同的模型更新策略.例如, Zhang 等人提出的 SSIT 算法<sup>[17]</sup>采用了一种增量学习方法,它基于样本选择来优化训练过程,避免了训练信息丢失问题,但这也使得训练更加复杂. Li 等人提出了一种基于 DCF 框架的动态学习率算法<sup>[18]</sup>,这种方法可以根据目标的运动变化动态调整学习率,以适应目标外观的变化.然而,在遮挡情况下,这种方法可能会导致过拟合,从而影响跟踪性能.

本文着眼单目标跟踪步骤中所出现的问题,在构建颜色模型的前提下,为 HOG 特征引入双线性插值<sup>[19]</sup>以建立高维度形状模型;针对上述多模态模型<sup>[20]</sup>设计深层次高置信度模型融合策略,特别提出量纲层级及其归一化标准,对模型内、模型间的多种置信度,分别实施平均权衡机制和高置信度倾向的自适应权衡机制,

确保融合目标的高置信度;最后根据不同模型的更新需求,设计模型分级非线性学习率曲线及参数,同时设定学习率的波动上下限,以保持模型跟踪模板的稳定性和可靠性;并采用粒子滤波的搜索方式,提高在线跟踪目标的速度和鲁棒性.

## 1 目标多模态模型设计与搜索策略

### 1.1 目标多模态模型设计

本文将颜色特征视为快速识别目标的关键因素,首先将目标  $M_t$  的 RGB 色彩通道转化为 HSV 色彩通道,再计算色彩通道中各颜色的出现概率,以此建立的颜色直方图作为目标的颜色模型,该模型用以下方式表达:  $O = \{C_1, C_2, C_3, \dots, C_m\}$ , 相应公式如下:

$$C_m = \frac{X_m}{\max X_m - \min X_m} \quad (1)$$

其中,  $X_m$  表示颜色直方图中第  $m$  维的值,  $C_m$  是对 HSV 颜色直方图进行归一化.考虑颜色是非固定属性且易受环境影响,本文引入形状特征以提升跟踪效果.首先通过对目标灰度图的仿射变换生成有效样本,以增加包含平移、旋转等有效样本的数量和密度,增强模型的泛化能力,减少过拟合并提高精确率. KCF 算法<sup>[21]</sup>使用的是经典 HOG 特征,其对目标的姿态变化敏感,目标的形变或旋转可能导致 HOG 特征描述子发生显著变化,并且特征描述子会受到背景噪声的影响.针对此问题,本文引入双线性插值来优化 HOG 特征,该方法通过平滑图像重采样减少局部特征变化,可在一定程度上减少上述问题,并使模型能学习到高密度信息,进一步增强泛化能力,最后采用这些改进特征对模型进行多样化和深层次的训练.步骤见图 1.

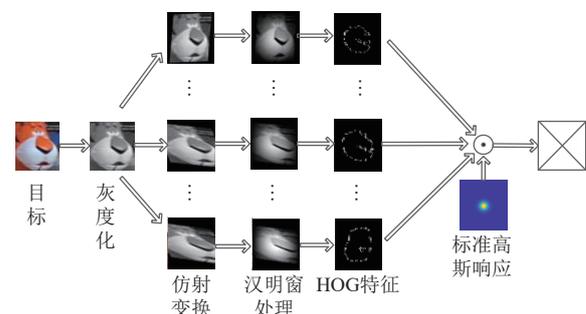


图1 形状模型建模过程

形状模型训练过程如下:依据卷积定理,在频域中变量有如下关系:

$$F(g) = F(f) \odot F^*(h) \quad (2)$$

其中,  $h$  表示滤波器,  $g$  代表相关响应,  $F$  是傅里叶变换,

$F^*(h)$ 描述 $F(h)$ 的复共轭,而 $f$ 为特征候选区.特征增强会生成一些HOG特征样本 $f$ 以及对应的相关响应 $g_i$ ,其中 $i=1,2,\dots,N$ .设傅里叶变换 $F_i=F(f_i),G_i=F(g_i),H=F(h)$ ,则有:

$$H^* = \frac{G_i}{F_i}, \quad i=1,2,\dots,N \quad (3)$$

$H^*$ 可通过求解最优化问题得到:

$$\min_{H^*} \sum_{i=1}^N |F_i \odot H^* - G_i|^2 \quad (4)$$

由此得到最优解 $H^*$ 的数学表达式:

$$H^* = \frac{\sum_{i=1}^N G_i \odot F_i^*}{\sum_{i=1}^N F_i \odot F_i^*} \quad (5)$$

其中, $H^*$ 是 $H$ 的复共轭,后续随着视频推移,滤波器模型会更新,具体见第3.3节.

## 1.2 基于粒子滤波的目标搜索策略

在目标跟踪中,当前帧的目标大致位于前一帧附近,为了降低运行开销,增加算法速度,使用粒子滤波方法<sup>[20]</sup>预测当前帧的目标位置特别合适.这种方式可以限制目标在当前帧的搜索范围,并且搜索过程中粒子只需移动一次,提升了搜索效率.粒子 $S_i$ 的结构形式定义为 $S_i = \{P, R, V, a\}$ ,其中 $P$ 为粒子代表的矩形区域左上角的坐标、 $R$ 为该区域的长和宽、 $V$ 为粒子在 $x$ 和 $y$ 方向上的运动速度、 $a$ 为区域在 $x$ 和 $y$ 方向上尺寸的变化速度.

如图2所示,一旦获得第一帧的目标,就会在目标的位置周围分配一组若干的粒子框,作为初始运动模型.考虑到跟踪速度和搜索范围,本文设置了50个粒子框.接着这些粒子框会根据式(6)进行转移,从而预测目标可能出现在当前帧中的所在位置:

$$\begin{cases} M_t^i(x) = 2 \times M_{t-1}^i(x) - M_{t-2}^i(x) + R_x \\ M_t^i(y) = 2 \times M_{t-1}^i(y) - M_{t-2}^i(y) + R_y \\ M_t^i(s) = 2 \times M_{t-1}^i(s) - M_{t-2}^i(s) + R_s \end{cases} \quad (6)$$

其中, $M_t^i(x)$ 、 $M_t^i(y)$ 、 $M_t^i(s)$ 是第 $t$ 帧的候选框子图像 $M_t^i$ 的左上角 $x$ 坐标、 $y$ 坐标和宽高缩放比例系数. $R_x$ 、 $R_y$ 和 $R_s$ 分别为均值为0,标准差为1、0.5、0.001的随机数.转移完成后,每个粒子在后续的相似度融合和匹配计算后,需要重新采样,以确保粒子始终聚焦在目标周围,确定下一帧目标搜索范围.



图2 运动粒子模型示意图

## 2 多模态模型信息的准确量化

### 2.1 形状模型信息量化表达的局限性

在第1.2节中描述的目标搜索过程中,候选粒子框与目标的多模型需要进行相似度计算.然而,由于颜色模型和形状模型是两种不同类型的跟踪器,候选粒子框在形状模型计算时存在一些容易被忽视的问题.例如,文献[20]中在使用计算形状模型相似度信息时,会以跟踪响应中心位置的值作为两者之间相似度.但这样可能会导致两个问题.

(1) 在目标经历形变、旋转或部分遮挡时,单纯依赖中心响应值来衡量相似度往往不能准确反映实际情况,这一限制可能会降低跟踪的准确度.此外,跟踪响应的中心值会受到局部噪声、背景变动或光照条件改变的干扰,进而影响到相似度评估的准确性.这种情况下,算法可能会误判候选目标框的有效性,尤其在目标与背景相近或背景较为复杂的环境中,错配和跟踪偏差的风险会明显增加.

(2) 如图3(a)所示,形状模型的响应图代表了候选区域在特定空间下的全面表达,包含了丰富的全局信息分布.最高响应值仅占整个响应图的小部分,导致大量有效信息未被充分利用.若仅依赖图3(b)所示的 $g_{\max}$ 作为评估标准,会忽略关键的全局信息,尤其在 $g_{\max}$ 未能全面代表整体信息的场合下更为显著,进而造成目标模型与实际潜在目标框之间的错配.

### 2.2 模型的全局信息量化

如第2.1节所述,模型信息的表述会直接影响模型的正确融合.因此,本文首先将候选框颜色模型与目标颜色模型的巴氏距离 $dt_i^{-1}$ 用作颜色相似度.

DSST算法<sup>[21]</sup>中使用单一模型时,无法全面捕获目标的多维度信息,导致算法在处理复杂背景或光照变

化时的鲁棒性降低. 为解决第 2.1 节中形状信息缺失和 DSST 算法中的问题, 本文设计了一种利用响应图全局信息的多置信度表达, 通过结合 PSR 和 APCE 这两个指标, 分别记为  $gp_i$  和  $ga_i$ , 以此全面评估候选框与形状模型的相似度. 指标计算公式如下:

$$gp_i = \frac{g_{\max} - \mu_{s1}}{\sigma_{s1}} \quad (7)$$

$$ga_i = \frac{|g_{\max} - g_{\min}|^2}{\text{mean}\left(\sum_{w,h} (g_{w,h} - g_{\min})^2\right)} \quad (8)$$

式 (7) 中  $g_{\max}$  为响应图的最大值,  $\mu_{s1}$  为旁瓣的均值,  $\sigma_{s1}$  为旁瓣的标准差; 式 (8) 中  $g_{\min}$  为响应图的最小值,  $w, h$  为响应图的宽和高,  $g_{w,h}$  为响应图中每一个值.

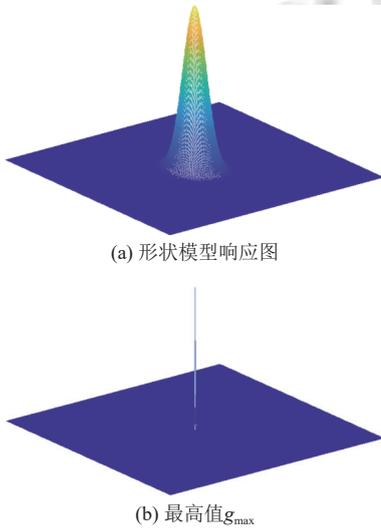


图 3 形状模型两种信息的差异

### 3 多模态模型的融合与更新策略

#### 3.1 模型深层次权衡融合策略设计

如第 2.2 节所述, 候选框具有多个可量化的置信度, 然而在特定场景中确定目标时, 需要对置信度进行评判, 从而权衡它们以获取最终的融合结果.

SAMF 算法<sup>[21]</sup>在如何具体执行模型融合及其权重动态调整的详细描述不够明确, 包括模型选择的标准、权重的计算方法以及自适应策略等具体细节, 为了填补此空白, 本文提出了一种较为简易的深层次权衡融合策略, 能在个别置信度不佳时, 通过不同置信度的相互调整, 确保整体的准确性, 从而提高跟踪的精度与稳定性.

针对以上需求, 本文首先设定一个从 0 到 1 的指标范围作为融合框架, 然后进行形状模型两种置信度的浅层融合. 之后以高置信度为融合标准, 将形状置信度  $gp_i$  和颜色置信度  $ga_i$  进行不同模型间的深层加权融合以获得最终置信度, 这样就会保证模型总是在一个框架下进行融合, 总体融合策略如图 4 所示.

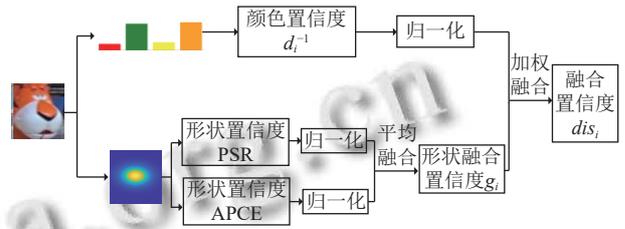


图 4 模型深层次权衡融合策略

两种置信度的相似度极限分别记为  $gp_{\max}$  和  $ga_{\max}$ , 将使用其归一化后的算术平均值作为融合后的置信度. 融合公式如下:

$$g_i = \frac{1}{2} \left( \frac{ga_i}{ga_{\max}} + \frac{gp_i}{gp_{\max}} \right) \quad (9)$$

在不同模型间的融合之前, 同样也需要计算归一化的颜色置信度:

$$d_i^{-1} = \frac{dt_i^{-1}}{d_{c\min}^{-1}} \quad (10)$$

式 (10) 中,  $dt_i^{-1}$  为颜色直方图的置信度,  $d_{c\min}^{-1}$  为颜色置信度的相似度极限,  $d_i^{-1}$  为归一化后的置信度.

针对不同模型间的置信度融合, 本文设计了一种高置信度权衡机制, 根据颜色模型和形状模型的波动程度做动态调整, 力求获取高置信度的候选框. 融合后的值  $dis_i$  可反映当前候选框的最终可信度, 计算公式如下:

$$dis_i = (1 - \tau)d_i^{-1} + \tau g_i, \quad i = 1, 2, \dots, M \quad (11)$$

其中,  $d_i^{-1}$  和  $g_i$  分别为粒子  $S_i$  目标颜色和形状模型的置信度,  $\tau$  为融合权重参数.

#### 3.2 模型间高置信度融合参数设计

由第 3.1 节中式 (11) 可知, 模型间融合的核心点在于融合权重参数  $\tau$  的确定, 以此综合两者的结果并以置信度  $dis_i$  的形式体现在前述粒子滤波候选目标的搜索过程中. 本文通过分析  $d_i^{-1}$  和  $g_i$  的数值反映出当前跟踪情况好坏, 以此决定融合参数  $\tau$  的值.

为解决 SAMF 算法<sup>[21]</sup>因参数调整繁琐而导致初始化设置复杂、调试难度增加的问题, 本文提出了一种

偏向高置信度的融合参数选择策略. 该策略主要基于一个简化的准则: 注意力仅聚焦于高置信度, 在融合过程中为其分配更高的权重, 策略如下所示.

(1)  $d_i^{-1} > \varepsilon_1$  且  $g_i > \varepsilon_2$ , 两个模型都工作良好, 将根据两者之间的置信度差异分配权重  $\tau$ .

(2)  $d_i^{-1} > \varepsilon_1$  且  $g_i \leq \varepsilon_2$ , 形状模型出现波动, 颜色模型状态良好, 将采用颜色模型的结果,  $\tau$  值为 0.

(3)  $d_i^{-1} \leq \varepsilon_1$  且  $g_i > \varepsilon_2$ , 颜色模型出现波动, 形状模型状态良好, 将采用形状模型的结果,  $\tau$  值为 1. 对于这 3 种情况, 实验中  $\tau$  的取值公式如下:

$$\tau = \begin{cases} 0.5 + 0.5 \left( \frac{g_i - d_i^{-1}}{\min\{g_i, d_i^{-1}\}} \right), & d_i^{-1} > \varepsilon_1 \text{ 且 } g_i > \varepsilon_2 \\ 0, & d_i^{-1} > \varepsilon_1 \text{ 且 } g_i \leq \varepsilon_2 \\ 1, & d_i^{-1} \leq \varepsilon_1 \text{ 且 } g_i > \varepsilon_2 \end{cases} \quad (12)$$

(4)  $d_i^{-1} \leq \varepsilon_1$  且  $g_i \leq \varepsilon_2$ , 两个模型同时出现波动, 目标丢失或者目标被遮挡, 融合  $\tau$  的值由两者之间最大值决定. 对于最后一种情况, 实验中  $\tau$  的取值公式如下:

$$\begin{cases} 0, & d_i^{-1} > g_i \\ 1, & d_i^{-1} \leq g_i \end{cases} \quad (13)$$

经过实验, 本文将  $\varepsilon_1$  设置为 0.16;  $\varepsilon_2$  设置为 0.12.

### 3.3 模型非线性分级平衡更新策略设计

在得到当前帧目标的位置后, 需要分别对颜色和形状模型进行更新, 这个过程当中, KCF 算法<sup>[21]</sup>使用固定式的学习率更新模型, 固定学习率意味着模型更新的速度和强度在整个跟踪过程中保持不变, 这可能不适用于所有场景, 特别是在目标外观发生快速或显著变化的情况下. 并且如果目标在连续几帧内遭受遮挡或显著形变, 可能导致模型过度拟合于最近的错误样本, 从而导致跟踪偏差. 为解决此问题, 本文设计一种非线性分级平衡更新策略, 确定非线性分级学习率曲线函数, 作为模型在学习和优化过程中的“步长”控制器. 在设计曲线时, 在不同阶段考虑了模型更新的学习率强度, 并限制其波动范围以保持模型的稳定性和避免过拟合. 记:

$$\begin{cases} d_{\min}^{-1} = \max_i \{d_i^{-1}\}, \\ g_{\max} = \max_i \{g_i\}, \end{cases} \quad i = 1, 2, \dots, M \quad (14)$$

式 (14) 中通过当前帧的最优颜色置信度  $d_{\min}^{-1}$  和最优形状置信度  $g_{\max}$  可判断颜色模型  $H^*$  和形状模型  $C$  的波动情况, 以得到两者更新率  $\eta$  和  $\lambda$ .

(1) 颜色模型更新策略: 当前帧的最优颜色模型记

为  $O^* = \{C_1^*, C_2^*, C_3^*, \dots, C_m^*\}$ , 颜色模型更新表达式为:

$$C_u(t+1) = (1-\lambda)C_u(t) + \lambda C_u^*(t), \quad u = 1, \dots, m \quad (15)$$

其中,  $\lambda$  为更新系数.

$$\lambda = \begin{cases} 0.08(d_i^{-1} - 0.16)^3 + 0.05, & d_{\min}^{-1} > \varepsilon_1 \\ 9.76(d_i^{-1} - 0.16)^3 + 0.05, & d_{\min}^{-1} \leq \varepsilon_1 \end{cases} \quad (16)$$

式 (15) 中  $C_1(I)$  表示颜色模型的直方图中第  $I$  维颜色的概率,  $C_i^*$  表示当前目标  $M_i^*$  所对应的颜色直方图.

当  $d_{\min}^{-1} > \varepsilon_1$  时, 颜色模型良好, 此时模型非线性加速更新, 加速度逐步增大; 当  $d_{\min}^{-1} \leq \varepsilon_1$  时, 颜色模型出现波动, 此时模型非线性加速更新, 加速度逐步减小.

(2) 形状模型更新策略: 当前帧的目标  $M_i^*$  经过预处理后图像  $F_i^*$  之后, 按照式 (17) 进行形状模型  $H^*$  的更新:

$$\begin{cases} A_t = (1-\eta) \times A_{t-1} + \eta \times g_t \odot F_t^* \\ B_t = (1-\eta) \times B_{t-1} + \eta \times F_t \odot F_t^* \\ H^* = A_t / B_t \end{cases} \quad (17)$$

与上述颜色模型更新的分析类似, 形状模型更新系数  $\eta$  的取值范围如下:

$$\eta = \begin{cases} 0.07(g_i - 0.12)^3 + 0.05, & g_{\max} > \varepsilon_2 \\ 23.15(g_i - 0.12)^3 + 0.05, & g_{\max} \leq \varepsilon_2 \end{cases} \quad (18)$$

当  $g_{\max} > \varepsilon_2$  时, 颜色模型良好, 此时模型非线性加速更新, 加速度逐步增大; 当  $g_{\max} \leq \varepsilon_2$  时, 颜色模型出现波动, 此时模型非线性加速更新, 加速度逐步减小.

## 4 算法流程

本文提出的多模态深层次高置信度融合跟踪算法, 通过结合颜色模态和形状模态的多置信度, 利用深层次高置信度融合策略, 实现了高精度的目标跟踪. 该算法具体如算法 1 所示.

算法 1. 多模态深层次高置信度融合跟踪算法

输入: 图像序列  $\{M_t\}$ , 初始位置  $P_1$ , 目标的宽和高  $(W, H)$ .

输出: 每一帧图像的目标位置  $\{P_t\}$ .

step 1. 在第 1 帧对给定目标分别建立颜色和形状模型.

根据式 (1) 建立目标的颜色模型  $O$ ; 根据式 (5) 建立目标的相关滤波器模型  $H^*$ .

step 2. 随机初始化粒子群的状态:

$$S_i = \{P, R, V, a\}, \quad i = 1, \dots, M$$

在第 1 帧目标位置及其附近放置粒子群. 粒子的具体含义参见第 1.2 节.

step 3. 对候选目标进行粒子滤波搜索获得最优解.

- a) 粒子根据式 (6) 进行转移, 根据式 (7) 和式 (8) 分别计算每个粒子  $S_i$  与目标的颜色置信度  $d_i^{-1}$  形状响应图置信度  $g_{P_i}$  和  $g_{A_i}$ .
  - b) 根据式 (9) 融合形状响应置信度  $g_{P_i}$  和  $g_{A_i}$ ; 根据式 (12) 和式 (13) 选取  $\tau$  值, 再根据式 (11) 进一步融合归一化后的颜色置信度  $d_i^{-1}$  和形状置信度  $g_i$ , 得到粒子  $S_i$  最终置信度  $dis_i$ .
  - c) 选择当前帧  $\{M_t\}$  中最高置信度的粒子框  $S_i^*$ , 其坐标作为当前帧的目标位置  $\{P_t\}$ , 并以此位置进行粒子  $S_i$  的重采样.
- step 4. 模型更新.
- a) 根据式 (15) 和式 (16) 更新颜色模型  $O$ .
  - b) 根据式 (17) 和式 (18) 更新形状滤波器模型  $H^*$ .
- step 5. 在下一帧循环 step 3 和 step 4.

图 5 和图 6 分别展示了算法中目标建模流程和算法的具体步骤以及它们在整体跟踪流程中的协同作用.

在跟踪流程的起始阶段, 算法首先对输入视频序列的第 1 帧中的目标进行分析, 分别通过 step 1 和 step 2 建立多模态模型和粒子初始化.

随后算法进入到跟踪搜寻目标的过程, 在每一新帧图像中执行 step 3, 通过这个步骤, 候选框和多模态

模型进行置信度计算与融合, 之后经过排序筛选后得到该帧的目标并进行粒子重采样.

一旦确定了当前帧的目标, 算法即通过 step 4 进行模型更新, 对颜色模型和形状模型进行微调 and 迭代, 确保模型的持续适应性和跟踪的准确性.

最终, 通过 step 5 算法实现对整个视频序列的连贯跟踪. 在这个阶段, step 3 和 step 4 步骤循环执行, 直至视频的最后一帧.

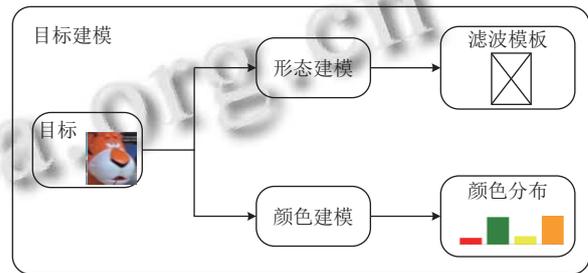


图 5 目标建模示意图

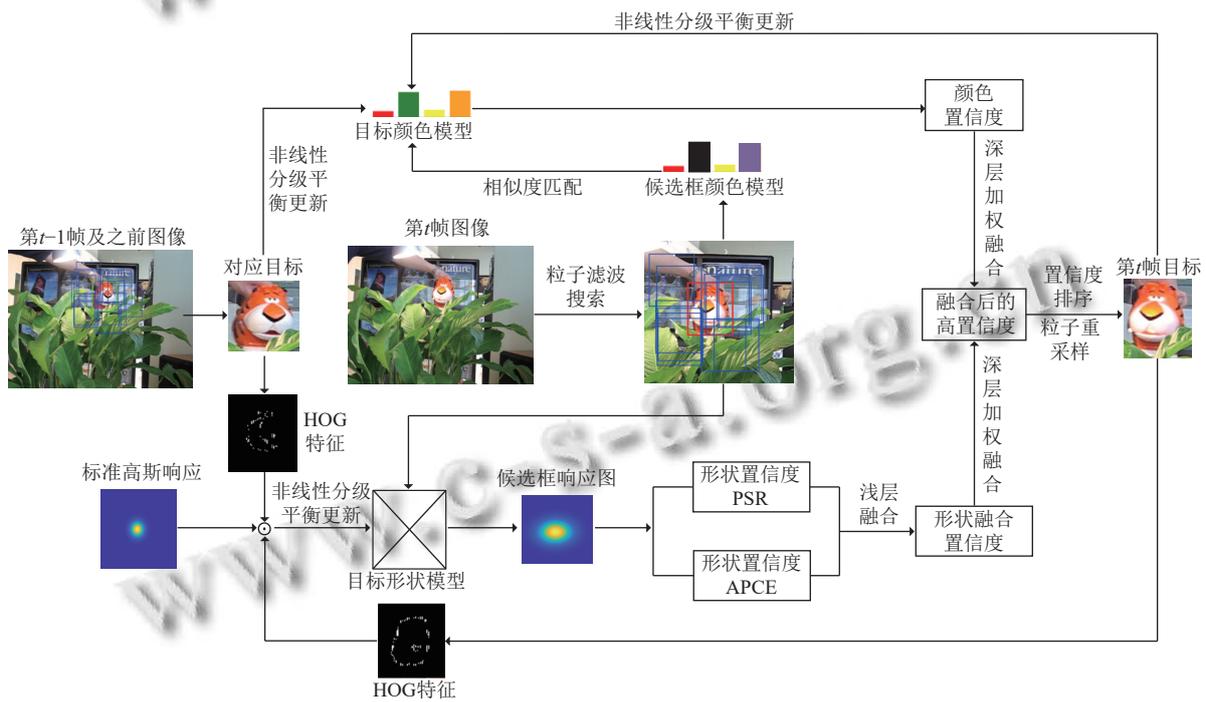


图 6 本文算法跟踪示意图 (视频帧从  $t=2$  开始)

## 5 实验与结果分析

本文所使用的测试数据集 OTB-2015 包含的视频序列有 11 个属性解释, 涉及各种挑战性因素, 是最权威的视觉目标跟踪数据集之一. 实验环境的技术参数为: Intel Core i1135G7 的 CPU, 主频 2.40 GHz, 内存 40 GB, 编程平台 VS2022, 实验评估平台 Matlab R2022a.

为了充分评估本文算法, 实验中选取了一些与本文最相关的顶尖算法<sup>[21]</sup>作为对比, 分别为 SAMF、fDSST、DSST、Staple、KCF、KCF-GaussHog 和 KCF-LinearHog.

本文算法改进了 SAMF 算法的模型融合方式, 以增强目标描述的准确性, 并创新性地引入了一个高置

信度准则,提供了一种比 SAMF 算法更加复杂和精细的策略.本算法融合了 KCF 及其变种算法所依赖的 HOG 特征模型,以及 Staple 算法使用的颜色直方图模型,不同的是本算法通过多模态模型的多置信度融合,超越了前者的单一模型依赖和后者简单模型的融合方法.不同于 DSST 和 fDSST 算法通过尺度滤波器达成的尺度适应,本算法通过模型融合策略提升了对目标尺寸变化的适应能力.此外提出的模型更新机制与上述比较算法中的策略均不相同,体现了明显的创新性.

本文使用 5 种指标评价上述算法的性能:中心误差 ( $CLE$ )、重叠率 ( $OS$ )、精确率 (accuracy rate)、成功率 (success rate) 和平均帧率 (FPS).中心误差和重叠率的计算式为:

$$CLE = \sqrt{(x_T - x_G)^2 + (y_T - y_G)^2} \quad (19)$$

$$OS = \frac{s(R_T \cap R_G)}{s(R_T \cup R_G)} \quad (20)$$

式 (19) 和式 (20) 中,  $(x_T, y_T)$  和  $(x_G, y_G)$  分别表示跟踪结果和真实目标的中心坐标,  $R_T$  表示跟踪算法所获得的目标区域,  $R_G$  表示真实的目标区域,  $s(R_T \cap R_G)$  和  $s(R_T \cup R_G)$  分别表示交集和并集区域的面积;精确率定

义为跟踪结果的  $CLE$  值小于阈值的帧比率;成功率定义为跟踪结果的  $OS$  值大于阈值的帧比率,跟踪效果好则  $CLE$  值越小,  $OS$  值越大.

### 5.1 本文算法整体性能和分析

图 7 展示了本文算法 (以下简称该算法) 与其他相关顶尖算法在 OTB 数据集测试的一次通过性 (OPE) 跟踪精确率和成功率.图 7(a) 中方括号内的数值表示中心误差阈值为 20 时的精确率,而图 7(b) 中方括号内的数值表示重叠率阈值为 0.5 时,对应算法的跟踪成功率.从图 7 中可以观察到,算法在整体 OPE 评估指标中显示出了最好的跟踪精确率 (0.798),尤其是当中心误差阈值大于 20 时,该算法在所有参照算法中的精确率最高;从跟踪成功率角度分析,在重叠率大于 0.3 时,该算法的成功率均显著优于其他参照算法.以上两方面的结果均证明本文方法使目标更容易被跟踪,证明了该算法的优越性.表 1 展示了图 7 中所有算法在数据集上的定量指标.可看出该算法的平均  $CLE$  和  $OS$  指标在所有参照算法中表现最好,平均  $CLE$  为 30.57,比次优的 Staple 算法减少 1.48,平均  $OS$  为 0.609,比次优的 KCF 算法提高 2.3%,且 FPS 为 15.67,满足一般实时性要求.

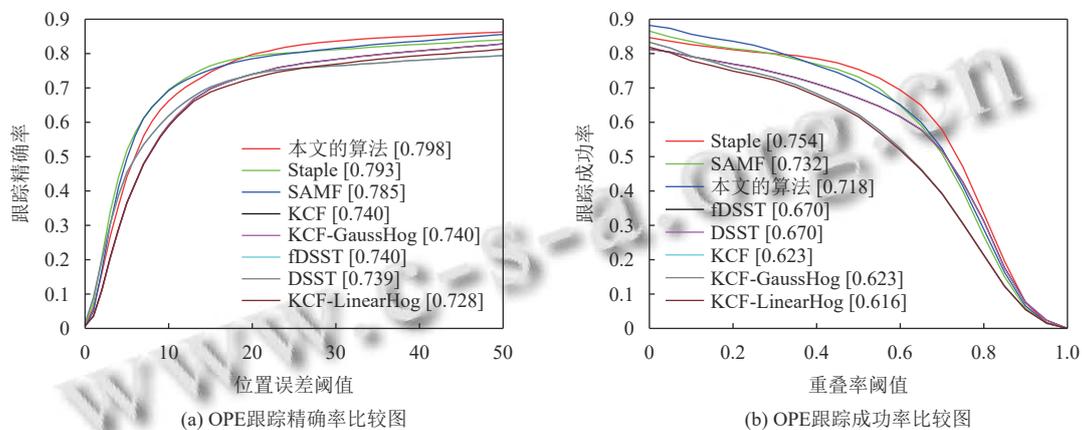


图 7 算法 OPE 跟踪精确率和成功率图

表 1 各个算法的性能对比

算法名称	$CLE$ 平均值	$OS$ 平均值	FPS
本文的算法	<b>30.57</b>	<b>0.609</b>	15.67
SAMF	36.683	0.584	14.63
fDSST	45.113	0.505	45.00
DSST	50.256	0.555	20.73
Staple	32.05	0.406	50.29
KCF	43.88	0.586	58.33
KCF-GaussHog	41.88	0.573	174.10
KCF-LinearHog	40.59	0.571	<b>284.92</b>

### 5.2 不同场景下算法评估

本文选取了光照、平面外旋转、尺度变化、变形以及离开视野等具体场景进行多方位定性定量评估,图 8 给出所参照的所有算法在特定场景的跟踪结果,可定量印证本文算法跟踪效果.

同时本文也列出了所参照算法在特定场景的直观对比,可定性印证本文算法的跟踪效果,如图 9 所示.

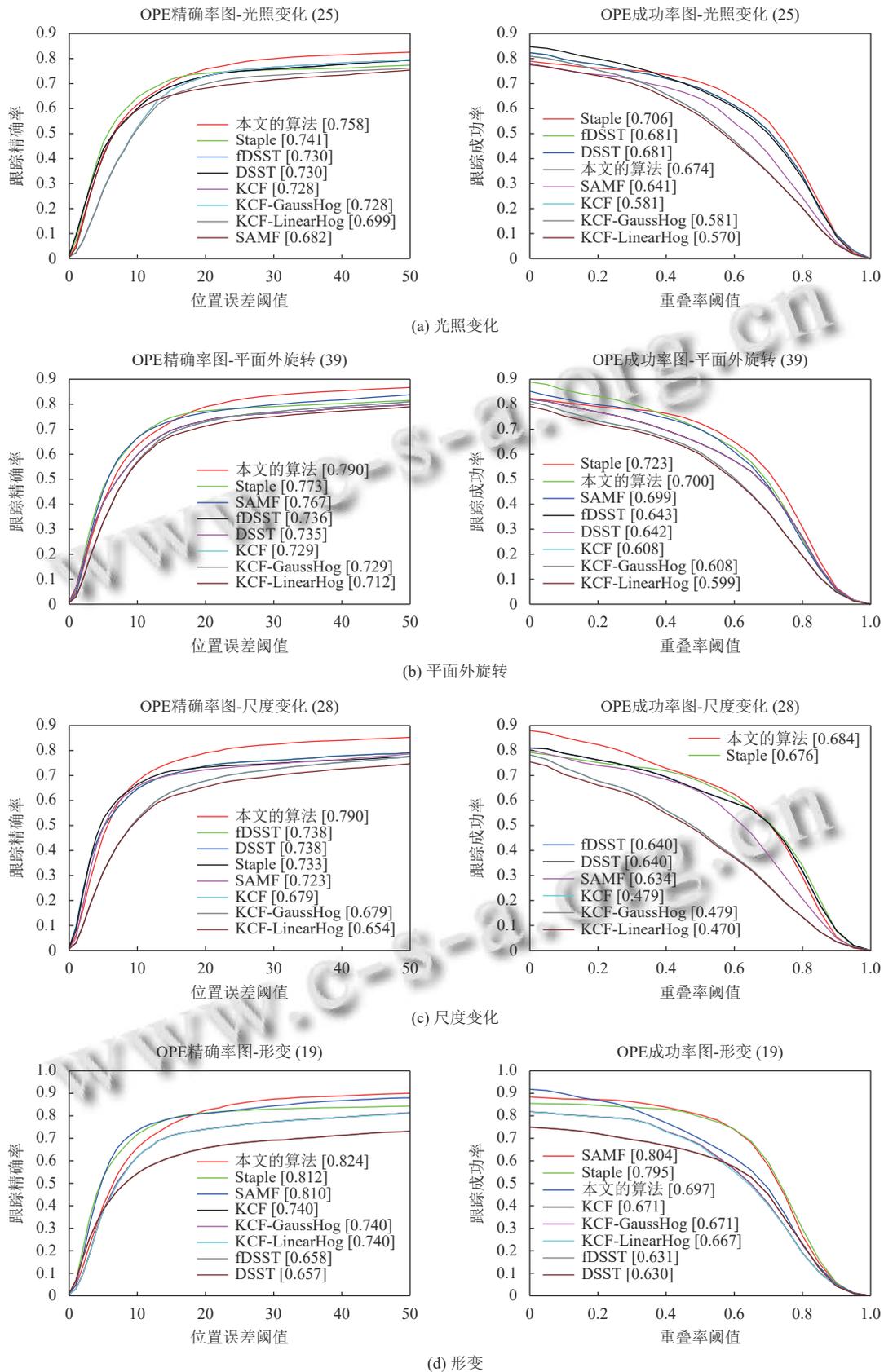


图8 不同场景下 OPE 跟踪精确率和成功率图

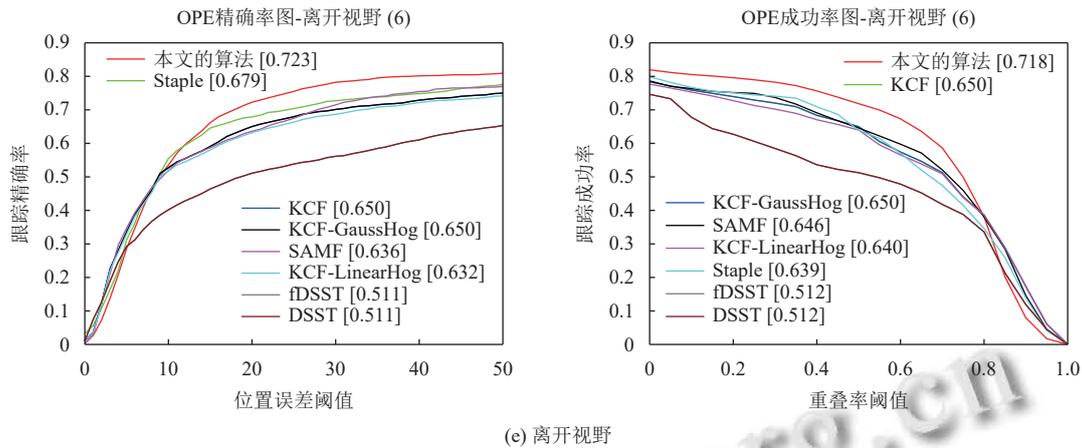


图8 不同场景下 OPE 跟踪精确率和成功率图(续)

如图8和图9所示,该算法在光照变化和平面外旋转场景中的精确率分别为0.758和0.790,均为所有算法中最优,当 $CLE$ 阈值大于20、 $OS$ 阈值大于0.4时,该算法的精确率和成功率显著优于参照算法. *shaking*序列的第10–364帧和*sylvester*序列的第1158、1162和1171帧分别出现了光照变化和平面外旋转的现象,本文算法在这两种场景中皆可优于其他算法.这表明目标的高维度多模态模型能全面、丰富地捕获深度目标特征,扩展了目标表示空间,有效地提升模型泛化能力.在尺度变化和离开视野的场景中,该算法的精确率分别为0.790和0.723,而正确率分别为0.684和0.718,同样在所有算法中表现最优,当中心误差阈值大于10,且重叠率阈值小于0.4和0.8时,该算法的精确率和成功率显著优于参照算法. *couple*序列的第10–134帧,以及*freeman4*序列的第228、278和

282帧中存在尺度变化和离开视野的现象,在这两类场景中,该算法也能成功跟踪目标,且表现优于其他算法,这显示出模型深层次融合的全面性和有效性;同时也验证了本文算法中粒子滤波搜索策略的有效性.在遇到目标形变的场景中,本文算法的精确率可达0.824,为所有算法中最高,当中心误差阈值大于20时,该算法的精确率也优于参照算法. *skiing*序列的第22帧和第35帧,以及*jogging-2*序列的第39–106帧展示了目标形变和被遮挡的情况,该算法在这两种情况下也同样能成功跟踪目标,且表现优于大部分参照算法.这表明,模型的非线性分级平衡更新能有效避免因目标变化和遮挡产生的过拟合,使模型更新、更稳健,保证跟踪的稳定性.以上定性定量的结果和分析均表明,该算法具有较高的鲁棒性、稳定性和有效性.

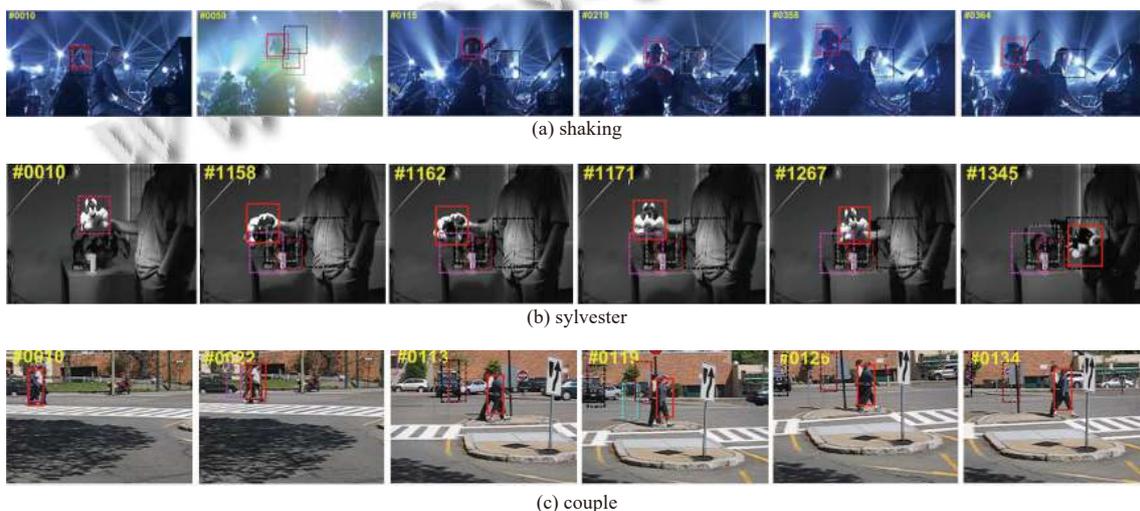


图9 算法在不同场景的直观对比结果(正红色实线框为本文算法)



图9 算法在不同场景的直观对比结果(正红色实线框为本文算法)(续)

## 6 结论

本文提出了一种多模态深层次高置信度融合跟踪算法. 通过构建高维度多模态模型, 提出了量纲层级统一的归一化标准; 此外, 设计了平均权衡机制和高置信度倾向的自适应权衡机制, 实现多模态模型的置信度融合; 同时, 设计了模型分级非线性学习率曲线及参数, 设定了学习率的波动上下限, 完成模型的非线性分级平衡更新; 采用了粒子滤波搜索方式筛选潜在目标, 提高了跟踪速度. 实验结果表明, 在 OTB-2015 数据集上, 该算法取得了较好的效果, 验证了算法的有效性和泛化能力. 但是本文算法仍存在不足: 模型更新的超参数调整耗时且困难, 导致在某些跟踪场景下无法进行有效跟踪, 并且仅能对预选框等比例缩小或放大, 影响正确目标的选取. 后续将着重于解决这些问题, 向更多场景进行高指标泛化.

## 参考文献

- Ni JJ, Wang XT, Gong T, *et al.* An improved adaptive ORB-SLAM method for monocular vision robot under dynamic environments. *International Journal of Machine Learning and Cybernetics*, 2022, 13(12): 3821–3836. [doi: 10.1007/s13042-022-01627-2]
- 齐滨, 田金, 邹男, 等. 量测驱动的自适应似然无源弱目标跟踪. *声学学报*, 2023, 48(5): 959–970. [doi: 10.12395/0371-0025.2022029]
- 徐文, 吴雨桑, 张婷. 多阵列水下多目标跟踪的分布式算法研究. *信号处理*, 2023, 39(10): 1764–1774.
- Xiu CB, Ma YF. Kernel correlation filter tracking strategy based on adaptive fusion response map. *IET Image Processing*, 2022, 16(4): 937–947. [doi: 10.1049/ipr2.12156]
- 陈仁祥, 谢文举, 徐向阳, 等. 基于数据融合和改进 MoCo 的工业机器人抖动原因识别. *仪器仪表学报*, 2023, 44(7): 112–120.
- 张惊雷, 宫文浩, 贾鑫. 基于自引导注意力的双模态校准融合目标检测算法. *模式识别与人工智能*, 2023, 36(9): 793–805.
- Yue YY, Yang Y, Yu YT, *et al.* Improving multi-object tracking by full occlusion handle and adaptive feature fusion. *IET Image Processing*, 2023, 17(12): 3423–3440. [doi: 10.1049/ipr2.12874]
- 段可欣, 闫文君, 刘凯, 等. 融合双谱特征的雷达辐射源个体识别方法. *海军航空大学学报*, 2023, 38(5): 382–390.
- 夏鸿斌, 黄凯, 刘渊. 多特征融合短会话推荐模型. *模式识别与人工智能*, 2023, 36(4): 354–365.
- Anzid H, Le Goic G, Bekkari A, *et al.* A new SURF-based algorithm for robust registration of multimodal images data. *The Visual Computer*, 2023, 39(4): 1667–1681.
- Gong TY, Zhang MM, Zhou Y, *et al.* Underwater image enhancement based on color feature fusion. *Electronics*, 2023, 12(24): 4999. [doi: 10.3390/electronics12244999]
- Nanni L, Faldani G, Brahmam S, *et al.* Improving foraminifera classification using convolutional neural networks with ensemble learning. *Signals*, 2023, 4(3): 524–538. [doi: 10.3390/signals4030028]
- Zhao M, Yang R, Hu M, *et al.* Deep learning-based technique for remote sensing image enhancement using multiscale feature fusion. *Sensors*, 2024, 24(2): 673. [doi: 10.3390/s24020673]
- 孙备, 党昭洋, 吴鹏, 等. 多尺度互交叉注意力改进的单无

- 人机对地伪装目标检测定位方法. 仪器仪表学报, 2023, 44(6): 54–65.
- 15 Bhat G, Danelljan M, van Gool L, *et al.* Learning discriminative model prediction for tracking. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 6181–6190.
- 16 周天奕, 丁卫平, 黄嘉爽, 等. 模糊逻辑引导的多粒度深度神经网络. 模式识别与人工智能, 2023, 36(9): 778–792.
- 17 Zhang YC, Liu YD, Yang GM, *et al.* SSIT: A sample selection-based incremental model training method for image recognition. Neural Computing and Applications, 2022, 34(4): 3117–3134. [doi: [10.1007/s00521-021-06515-4](https://doi.org/10.1007/s00521-021-06515-4)]
- 18 Li D, Li S, Wei Q, *et al.* Dynamic learning rate of template update for visual target tracking. Mathematics, 2023, 11(9): 1988. [doi: [10.3390/math11091988](https://doi.org/10.3390/math11091988)]
- 19 Zhou RG, Wan C. Quantum image scaling based on bilinear interpolation with decimals scaling ratio. International Journal of Theoretical Physics, 2021, 60(6): 2115–2144. [doi: [10.1007/s10773-021-04829-6](https://doi.org/10.1007/s10773-021-04829-6)]
- 20 陈昭炯, 叶东毅, 林德威. 基于背景抑制颜色分布新模型的合成式目标跟踪算法. 自动化学报, 2021, 47(3): 630–640.
- 21 刘艺, 李蒙蒙, 郑奇斌, 等. 视频目标跟踪算法综述. 计算机科学与探索, 2022, 16(7): 1504–1515. [doi: [10.3778/j.issn.1673-9418.2111105](https://doi.org/10.3778/j.issn.1673-9418.2111105)]
- (校对责编: 孙君艳)