

可见光与红外融合目标跟踪技术研究进展综述

方彦策¹, 赵君灵², 黄昭龙², 李旗挺³, 杜立超³, 张宏江³, 宋眉眉²

(1. 中国海洋大学 信息科学与工程学部, 山东 青岛 266100;

2. 天津理工大学 理学院, 天津 300384; 3. 中国运载火箭技术研究院 研究发展部, 北京 100076)

摘要: 目标跟踪是计算机视觉领域中的一个重要的问题, 自从被提出以来逐渐发展出了不同类型的技术方法; 由于可见光与红外光在目标跟踪方面存在互补性, 两者融合的目标跟踪技术在性能和鲁棒性上比传统目标跟踪方法更具有优势; 近年来, 人工智能技术的发展推动了可见光与红外融合目标跟踪技术的快速进步; 面向可见光与红外融合目标跟踪技术的发展历程, 首先针对可见光与红外融合领域经典方法和近年来出现的技术进行梳理和总结, 特别是对近两年的研究进展进行了归纳整理, 具体包括基于 Transformer、基于注意力机制、基于时间序列、自适应融合和基于多模态编解码器的可见光与红外融合目标跟踪方法; 然后介绍了可见光与红外融合数据集及目标跟踪的评价指标; 最后对未来的发展方向作了展望。

关键词: 目标跟踪; 可见光与红外融合 (RGBT); 深度学习; Transformer; 注意力机制

Review of Research Progress on Visible and Infrared Images Fusion Target Tracking

FANG Yance¹, ZHAO Junling², HUANG Zhaolong², LI Qiting³, DU Lichao³,
ZHANG Hongjiang³, SONG Meimei²

(1. College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China;

2. College of Science, Tianjin University of Technology, Tianjin 300384, China;

3. R&D Department, China Academy of Launch Vehicle Technology, Beijing 100076, China)

Abstract: Object tracking is an important problem in the field of computer vision, and different types of technical methods have been gradually developed since it is proposed. Due to the complementarity of visible and infrared light in target tracking, the fusion of the two methods has more advantages than traditional target tracking in performance and robustness. In recent years, the development of artificial intelligence has promoted the rapid progress of the visible and infrared light fusion target tracking. Facing the development history of the visible and infrared light fusion target tracking technology, firstly, the classic method of visible and infrared light fusion and the technologies emerging in recent years are sorted out and concluded, especially the research progress in the past two years is summarized, which are based on Transformer, attention mechanism, time series, self-adaptive fusion and multi-modal decoder; then the visible and infrared light fusion datasets and target tracking evaluation index are introduced; Finally, the future development of this field is prospected.

Keywords: target tracking; RGB-infrared fusion tracking (RGBT); deep learning; transformer; attention

0 引言

目标跟踪是计算机视觉领域中的一个重要的问题, 在自动驾驶^[1]、无人机^[2]、机器人^[3]等领域有广泛应用。在目标跟踪的早期阶段, 卡尔曼滤波^[4], 光流法^[5]等方法已被陆续用于目标跟踪。然而, 缓慢的计算速度和较低的精度限制了目标跟踪水平进一步的发展。同时, 传统的可见光目标跟踪容易受到天气、光线、视频图像质量等各种干扰。近年来, 在通过可见光视频进行目标跟踪的基础上, 将红外视频图像引入目标跟踪的做法, 有效地提高了传统目标跟踪的效果, 这一方法被称为可见光与红外融合目标跟踪

方法 (RGBT, RGB-infrared fusion tracking)。Yilmaz 等人^[6]在 2006 年对当时最先进的目标跟踪方法进行了总结, 但当时的方法仍然存在精度等方面的不足。近年来, 人工智能技术的持续发展有力地促进了目标跟踪技术的进步, 突破了传统可见光与红外融合目标跟踪的发展瓶颈, 为可见光与红外融合的目标跟踪技术不断提供新的思路、框架和工具, 并逐渐在速度和准确性之间取得平衡, 成为进一步研究的重点。Zhang 等人在 2020 年^[7]从稀疏表示、相关滤波、深度学习等方面梳理了可见光和红外融合的主要方法。本文面向可见光与红外融合目标跟踪技术的发展历程,

收稿日期: 2022-06-29; 修回日期: 2022-07-18。

基金项目: 中国航天科技集团公司钱学森青年创新基金资助项目。

作者简介: 方彦策 (2001-), 男, 浙江杭州人, 大学本科生, 主要从事计算机视觉方向的研究。

通讯作者: 李旗挺 (1989-), 男, 浙江杭州人, 博士, 高级工程师, 主要从事人工智能与量子计算方向的研究。

引用格式: 方彦策, 赵君灵, 黄昭龙, 等. 可见光与红外融合目标跟踪技术研究进展综述[J]. 计算机测量与控制, 2022, 30(10): 7-16.

特别针对近两年来出现的新方法和新技术,对可见光与红外融合目标跟踪技术的研究现状进行总结。

本文首先介绍可见光与红外融合跟踪的相关工作,然后从传统方法和基于深度学习的方法两方面进行阐述,接着介绍可见光与红外数据集和相关评估指标,最后提出了对该领域未来理论研究和应用领域发展方向的展望。

1 基本概念

1.1 可见光和红外图像

可见光指的是波长在 390~780 nm 之间的光线,可见光图像具有明显的颜色和纹理信息。红外光指的是波长为 760 nm~1 000 μm (真空中)的辐射电磁波,红外图像反映的是目标和背景向外辐射能量的差异,具有较大的作用距离和较强的抗干扰能力。可见光图像容易受光照影响,但可以提供丰富的图像细节;红外图像不容易受光照影响,虽然能大致描绘物体的形状和位置,但缺乏纹理等细节信息^[8]。由此可见,如果将可见光和红外图像进行融合,可以实现较好的互补作用。

1.2 可见光和红外图像融合跟踪技术方法

图像融合是指用特定的算法将多幅图像合成一幅图像,原图像集合的大部分特点都能从合成后的图像中体现(图1)。将可见光图像和红外图像进行融合,既可以减少外界光照等环境因素的影响,也可以获得尽可能丰富的图像信息。通常来说,可见光和红外图像融合的方法有:像素级融合、特征级融合和决策级融合^[9]。

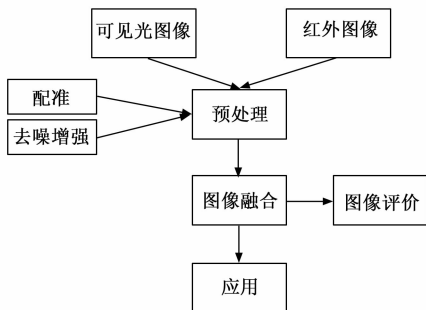


图1 可见光与红外图像融合过程

1.2.1 像素级融合

像素级图像融合是指对可见光和红外图像进行匹配后,在两种图像的像素之上计算出新的像素值的方法,新的像素值将融合两种图像的信息。像素级融合实现难度比较低,但是计算量较大,在实时性方面有所欠缺。

1.2.2 特征级融合

特征级融合的过程是对可见光和红外图像分别进行特征提取之后,进行基于特征(如边缘、形状、轮廓等)的数据融合,并用融合后的特征做出判断。

1.2.3 决策级融合

决策级融合是先对于可见光和红外图像分别进行单独的处理,得到各自判断和识别的初步结果,然后将这些初步结果按照一定的规则和权重进行调整,最终获得最优的

结果的过程。决策级融合在鲁棒性、实时性、开放性等方面都具有优势。

1.3 目标跟踪

目标跟踪是一种在给定视频序列初始帧的目标大小和位置的情况下,利用特定方法获得在后续帧中目标的大小和位置的技术方法。目标跟踪在经过了早期经典方法的发展之后,又出现了基于相关滤波、深度学习等理论的方法。

早期经典方法主要根据目标的特征进行跟踪。例如,光流法是^[5]通过相邻帧之间像素的位置变化来判断目标的运动状态,但这种方法要求限制像素的位移距离,所以光流法有很大的局限性。此外,粒子滤波等经典算法也被用于进行目标跟踪,但还不能很好地解决目标跟踪中的各种问题,也无法处理各种复杂的现实情况。

在基于相关滤波的目标跟踪方法方面,最早利用相关滤波器进行目标跟踪的方法是“最小输出平方和误差”算法(MOSSE, minimum output sum of squared error)^[10],在实现了较高速度的同时提升了目标跟踪的效果。

在基于深度学习的目标跟踪方面,“视觉几何”研究组(VGG, visual geometry group)提出了VGG-19网络^[11]并应用于目标跟踪,该网络利用部分卷积层作为特征提取层,得到的特征经由相关滤波器进行处理,实现了很好的跟踪性能。此后,该领域还出现了对冲深度跟踪(HDT, hedged deep tracking)^[12]、视觉跟踪连续卷积算子(C-COT, continuous convolution operators for visual tracking)^[13]等基于深度学习的方法,在目标跟踪的实时性和性能上都有所提高。

2 可见光和红外融合目标跟踪

可见光与红外融合目标跟踪的早期方法有卡尔曼滤波、粒子滤波和均值漂移等,经过较长一段时间的研究之后,发展出了基于相关滤波的方法、基于图的方法和基于稀疏表示的方法。近年来,神经网络表现出了强大的学习能力,突破了传统目标跟踪方法存在的精度瓶颈,成为了可见光与红外融合目标跟踪的一种主要方法。有鉴于此,本文将可见光与红外融合的目标跟踪方法分为经典方法和基于深度学习的方法。

2.1 经典方法

2.1.1 早期的经典方法

早期的可见光与红外融合目标跟踪方法主要基于三种不同种类的算法,分别是卡尔曼滤波算法、粒子滤波算法和均值漂移算法。

卡尔曼滤波是早期目标跟踪方法中常用的算法。卡尔曼滤波算法由R. E. Kalman在1960年提出^[4],卡尔曼滤波是一种高效率的递归滤波器,它能够从一系列不完全和包含噪声的测量中,估计动态系统的状态。基于卡尔曼滤波的方法,论文^[14]提出了一种运动物体检测和跟踪的系统,该系统把红外视频和可见光视频稳定地集成在水平集框架中,将三维结构张量扩展为通量张量,实现无特征分解的快速、鲁棒的运动检测。除此之外,一种压缩时空卡尔曼

滤波融合跟踪算法^[15]将卡尔曼滤波扩展到多传感器融合跟踪的情形, 其中图像来源包括可见光图像和红外图像。

根据已知的公开文献, 粒子滤波算法在 1993 年由 Gordon 等人提出^[16], 它通过非参数化的蒙特卡罗模拟方法来实现递推贝叶斯滤波, 适用于任何能用状态空间模型描述的非线性系统, 其精度可以逼近最优估计。粒子滤波方法在 1998 年^[17]首次被引入到目标跟踪领域, 该方法将静态非高斯问题的统计因子抽样算法和目标运动的随机模型进行了融合。之后, 有研究人员利用一种融合颜色提示和结构相似性度量的粒子滤波器^[18], 去探讨可见光和红外摄像机的视频像素级融合是否会存在精度上的问题, 实验表明该滤波器在红外的视频中跟踪性能较好, 而在可见光视频中由于光照等环境因素干扰较大导致跟踪性能较差, 融合后不能保证取得更好的效果, 甚至可能还会导致性能下降。对此, 论文^[18]认为基于多分辨率的融合方法, 可以有效解决普通融合方法带来的性能下降问题。另外, 传统的融合前跟踪策略还存在多个单一传感器之间相互影响的情况, 因此出现了一种改进的粒子滤波算法^[19], 它使用带有空间信息的颜色直方图来表示目标模型, 并给予每个粒子的颜色特征权值, 同时在融合可见光和红外序列的跟踪结果的规则上进行了改进, 最后根据目标融合跟踪结果更新模板, 在有效性、鲁棒性和实时性上均有提升。

均值漂移算法最早由 K. Fukunaga 等人提出^[20], 它是一种沿着密度上升方向寻找簇聚点的方法, 通过不断地重复计算距离均值来移动中心点, 实现对目标的跟踪。均值漂移在复杂的背景变化之下, 会导致鲁棒性和跟踪效果的下降。针对此问题, 一种基于区域目标检测和模糊区域规则的 FRDIF 融合方法^[21]采用相似度加权算法, 解决了复杂背景变化带来的影响。除此之外, 为了解决其他多模态跟踪方法遇到的数据存储量和处理量指数增长的情况, 有学者提出了一个可以融合多个空间图跟踪器的框架^[22], 通过有效地组合特征以进行抗干扰的跟踪, 提升了跟踪效果。

2.1.2 基于相关滤波的方法

相关滤波在目标跟踪中是一种重要的方法。它通过设计特定的滤波模板, 与目标所在的候选区域做相应的运算, 从输出的最大相应位置中可以得到目标的近似位置。相关滤波由于它自身在效率和正确性上的优势, 在目标跟踪领域得到了很大的关注。根据公开的文献资料, 第一个被用于可见光与红外融合目标跟踪的相关滤波器是基于软一致性的相关滤波器^[23], 软一致性是指在稀疏不一致的情况下, 使可见光和红外光的光谱保持一致, 从而实现可见光和红外数据更有效地融合, 并利用加权融合机制来计算检测阶段的最终响应图。

在目标跟踪阶段, 相似物体或背景噪声的存在会导致算法的准确率降低, 基于相关滤波的一种大边缘目标跟踪方法和一种多模态目标检测技术的提出解决了相关滤波在此方面的不足^[24]。考虑到在不同模式中的特征也具有一定的相似性, 因此, 基于这个原则可以构建一个相关滤波

器^[25], 它利用低秩约束联合学习不同模态, 在继承相关滤波的优点的同时, 鲁棒性也有一定程度的增强。除此之外, 还有多种基于相关滤波的方法, 例如基于相关滤波器和直方图的融合跟踪方法^[26]、基于相关滤波器的可见光跟踪^[27]和基于马尔可夫链蒙特卡罗的红外跟踪^[28]。

2.1.3 基于图的方法

公开文献显示, 图相关的模型是在 2017 年被提出的^[28], 通过使用可见光和红外数据来学习对象的表示, 该模型被称为加权稀疏表示正则化图。为了抑制可见光与红外跟踪的背景效应, 研究人员^[29]基于跨模态流形排序算法, 通过将软交叉模态一致性整合到排名模型中, 并用最优查询学习方法来处理查询的标签噪声, 将排序结果融合到基于块的对象特征中以解决背景效应。在此基础上, 文章^[30]又通过学习局部和全局多图描述符对先前的图模型进行了改进, 首先用多个图表示对象, 并以一组多模态图像块为节点以防止失真和部分遮挡; 然后随着时间的推移, 使用空间平滑度以及低秩表示动态地学习联合图; 接着将多图信息与对应的图节点权重相结合, 形成鲁棒的对象描述符, 最后采用结构化支持向量机进行跟踪。在之前的研究基础上, 一种两阶段模态图正则化流形排序算法被提出并用于学习可见光与红外融合跟踪对象的鲁棒表示^[31], 该算法利用结构化向量机对目标位置进行预测, 从而实现对目标的表示和跟踪。

稀疏表示是一种使用字典中元素的线性组合来表示样本的方法。据公开资料显示, 稀疏表示在 2009 年首次被引入可见光与红外融合目标跟踪任务^[32], 该方法将跟踪任务建模为稀疏近似问题, 并通过正则化的最小二乘方法来解决。在添加非负性约束和更新动态模板之后, 实验表明基于稀疏表示的方法有很大的潜力^[32]。随后, 另一种基于稀疏表示的数据融合方法也被证明对于可见光与红外融合目标跟踪任务^[33]是有效的, 它来自不同目标候选源的图像块链接到一个一维向量, 然后将其稀疏表示到目标模板空间。对具有异构信息源的多个真实视频的实验表明, 这个方法比当时同类目标跟踪算法具有更强的鲁棒性。在稀疏表示的基础上, 联合稀疏表示提供了一种比较自然的方法来融合多种模态的信息^[34], 使用联合稀疏表示设计的似然函数能有效地判断样本之间的相似性。

除了直接利用稀疏表示进行可见光和红外融合的目标跟踪外, 稀疏表示还可以与贝叶斯框架等多种方法相结合, 从而获得不同的稀疏表示模型, 达到比单一稀疏表示更好的效果。文献^[35]介绍了一种在贝叶斯过滤框架中的方法, 通过拉普拉斯稀疏表示引入生成多模态特征模型, 进行实时在线的灰度可见光与红外目标跟踪。该模型充分利用局部块之间的相似性来细化其稀疏代码, 从而可以无缝融合不同的源数据以进行目标跟踪, 在鲁棒性和有效性上都得到了实验验证。另一种贝叶斯过滤框架中的跨模态稀疏表示的融合方法^[36], 是在模型中引入模态权重以实现自适应融合, 使用重建残差和系数来定义运动模型生成的每

个候选样本的似然概率，最后通过寻找具有最大似然概率的候选样本来定位目标。另外，模型^[37]基于可见光和红外目标跟踪的模式相关感知，通过低秩正则化表征不同模式之间的相关性，并结合稀疏正则化的表示，使其能够进行有效模式融合并处理较大的外观变化。

在贝叶斯过滤框架中，还有一种基于联合稀疏表示的自适应融合方案^[38]，能够自适应地结合来自灰度和红外视频的信息，可在比较复杂的场景中进行目标跟踪，也适用于在线跟踪任务。但是上述方法对于有挑战性的场景仍然不能很好地进行目标跟踪，因此在贝叶斯过滤框架下又有了一种基于多任务拉普拉斯稀疏表示的灰度（可见光）—红外目标跟踪方法^[39]，它基于给定的边界框提取出一组重叠的局部块，寻找灰度和红外模态的多任务联合稀疏表示，并将这两种模态的表示系数连接成一个向量来表示边界框的特性，该方法在比较有挑战性的跟踪任务中具有一定的有效性。

2.2 基于深度学习的方法

2.2.1 基于 Transformer 的方法

Transformer^[40]是一个利用注意力机制来提高模型训练速度的结构，由 Google 在 2017 年提出，最早被设计用于自然语言处理。Transformer 使用注意力结构代替长短时神经网络，同时也跳出了编解码器与卷积神经网络（CNN, convolutional neural network）或循环神经网络（RNN, recurrent neural network）结合的固定模型框架。当前，Transformer 在深度学习的各个领域都表现出了其出色的性能，其中也包括可见光与红外融合目标跟踪。

根据已公开的文献，第一个将 Transformer 引入可见光与红外融合目标跟踪领域的是一种称为跨模式协作上下文表示（CMC2R, cross-modal collaborative contextual representation）的双流混合结构^[41]，通过编码器块转换层融合不同分辨率下的局部特征和全局表征，以及空间和通道的自我注意机制，实现两种模态的信息融合，最终获得上下文信息（图 2），该网络在目标跟踪任务中表现出了较好的性能。

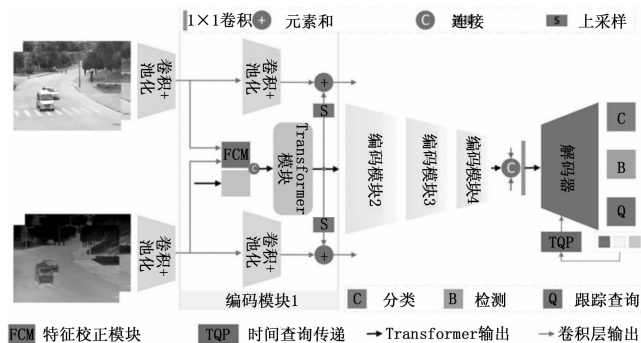


图 2 基于 Transformer 的跨模式协作上下文表示双流混合结构

2.2.2 基于注意力的方法

深度学习中的注意力机制，来源于对人类视觉的注意

力机制的研究。注意力机制能够像人眼一样，在输入图像之后，评估图像中不同区域的重要性，并为其分配不同的权重。分层双传感器交互网络（HDINet, hierarchical dual-sensor interaction network)^[42]较早将注意力机制引入可见光与红外融合目标跟踪领域。该网络的核心是特征交互模块和数据编码模块两个模块，前者基于注意力机制提取和补充双传感器的主导信息，而后者负责将原始数据编码为第一个特征交互模块的初始输入，同时后者的工作质量对整个网络的性能有关键影响。为了高效地融合双模态信息，有学者提出了基于模态感知注意网络和竞争学习（MaC-Net, modal-aware attention network and competitive learning）的可见光和红外融合目标跟踪算法^[43]。该算法由特征提取网络、模态感知注意力网络和分类网络组成，其中特征提取网络利用双流网络结构，从每个模态图像中提取特征；模式感知注意力网络通过整合原始数据建立一个描述不同特征层重要性的注意力模型，来引导特征的融合，以增强不同模式间的信息交互；分类网络由分别作用于可见光数据、热红外数据和融合信息数据的三个平行的二元分类器组成，辅以一个面向多模态的损失函数。在竞争性学习的训练策略引导下，整个网络将向双模态的最佳融合方向进行微调。为进一步充分利用可见光图像和热红外图像在跟踪中的互补优势，跨模态注意网络（CANet, cross-modal attention network)^[44]给出了可见光与红外融合目标跟踪的两个并行网络中的跨模型操作（CM, cross-model operation），在两种模态之间，设计了一种基于注意力机制校正的特征信息的加法运算，在获得更丰富的模态特征信息基础上有效减少了计算冗余。另外，利用一个并行的、分层的交互网络来实现两种学习方式的特征互补，体现了深度学习的互补优势。然而，基于注意力机制的方法可能会导致在较长的时间跨度上，单帧学习的特征权重无法使分类器专注于鲁棒的特征^[45]，为解决此问题，研究人员提出了一种包含全局和局部注意力的双视觉注意力引导的跟踪算法。此外，针对处理硬样本（即难以学习的样本）分类的一种用于可见光与红外跟踪的多模态学习框架（M²L, multi-modal multi-margin metric learning)^[46]将所有样本分为四个部分，利用其相互之间的关系来提高特征嵌入的稳健性，其中的注意力机制融合模块可以实现基于质量感知的数据整合。

2.2.3 基于时间序列的方法

在深度学习中，时间信息是输入神经网络的视频流中不可或缺的因素。目标跟踪任务中，基于前后帧存在的时间上的因果关系，通常可以将时间因素作为线索。因此，在神经网络中引入长短时记忆（LSTM, long short-term memory）等基于时间序列的方法是可见光和红外融合目标跟踪的重要方法之一。

基于历史跟踪结果的自适应融合算法^[47]首先将基于时间序列的深度学习方法引入可见光与红外目标跟踪领域。该方法综合了前向和后向的跟踪结果来评估跟踪精度，在

给定初始目标边界的情况下, 采用卷积神经网络进行特征提取, 分别对目标进行可见光和红外目标跟踪, 同时在这两种模式中实现反向跟踪, 并且计算每对之间的差异。这一基于历史跟踪结果的融合算法, 自适应地汇集了可见光和红外信息, 解决了可见光和红外融合跟踪过程中的一些难题。ChiNet 将长短时记忆跟可见光与红外目标跟踪相结合^[48], 利用航天器交会序列的时间信息来估计航天器的相对姿态, 根据长短时记忆单元在数据序列建模中的性能, 来处理卷积神经网络主干提取的特征, 联合多模态可见光与红外图像的输入, 将平均位置误差缩小近 80%。除此之外, 一种考虑时间信息的自适应可见光与红外目标跟踪器^[49]综合分析了空间和时间因素, 在传统空间信息网络的基础上, 增添了包含时间信息的网络, 构建了一个用于跨模态交互的自适应融合子网络, 从更多的维度中获取有效的信息。

2.2.4 自适应融合的方法

随着深度学习的迅速发展, 原有的传统图像融合方法逐步倾向于与深度学习融合, 大量融合模块的研究和设计, 对于综合可见光图像和红外图像的信息起了重要的作用。

针对不同序列的图像对的个体特征和共同特征, 动态融合网络 (DFNet, dynamic fusion network)^[50] 的双流结构中的每一层可以依据两个非共享卷积核来提取个体特征, 通过共享卷积核为每一层提取共同特征, 并对非共享卷积核和共享卷积核进行自适应加权和求和, 从而可以动态计算个体特征和共同特征在面对模态可靠性变化时的贡献。与 DFNet 相似, 三串流自适应融合网络 (TAFNet, three-stream adaptive fusion network)^[51] 借助成对的可见光和红外图像进行人群计数。TAFNet 分为一个主流和两个辅助流, 主流输入由一对可见光和红外图像结合构成, 两个辅助流分别利用可见光图像和红外图像提取特定模态的特征。与前者所不同的是, 增强背景感知相关滤波方法^[52] 则采用了先融合后跟踪的策略, 该方法将红外图像转换为单通道图像, 利用灰度信息确定目标和整体环境之间的像素差异程度, 通过对可见光和红外图像的自适应加权决策实现目标跟踪。基于响应图评估算法^[53] 改进了高斯回归中的自适应融合权重, 运用分层卷积神经网络分别提取了可见光和红外图像中的深度特征, 每个层的可见光和红外信息互不干扰。为了更好地开发和利用多尺度信息, 多分支自适应融合网络^[54] 从多个分支中聚合多尺度信息, 由多尺度适配器以并行的方式提取特征, 并由多分支融合模块自适应地聚合来自多个分支以及上一层的特征, 从而减轻来自低质量图像和视频中的噪声的影响。为了增强不同模态的特征表示并充分挖掘模态之间的互补性, 三叉戟融合网络 (TFNET, trident fusion network)^[55] 通过递归策略来聚合所有卷积层的特征, 利用聚合特征和模态特定特征进行分类和回归, 实现了更加鲁棒的目标跟踪。

2.2.5 基于多模态编解码器的方法

多模态编解码器可以将多种模态的信息输入 (如可见

光、红外) 转化成特定长度的向量, 再将向量转化成特定形式并进行输出。论文^[56] 中提出了一种多交互双解码器, 旨在解决可见光和红外融合目标跟踪方面存在的两个关键问题: 一是如何实现不同模态之间的有效互补, 防止噪声干扰; 二是如何抑制显著性偏差, 即如何聚焦可见光和红外目标的共同特征, 避免被单一模态主导跟踪。该方法利用多交互块来模拟双模态、多级特征和全局上下文之间的交互, 从而融合了不同模态之间有效的互补特征, 恢复出更多的空间细节, 实现对目标对象的定位并抑制背景噪声。

3 数据集与评价指标

3.1 可见光与红外数据集

数据集在可见光与红外融合目标跟踪的训练和测试过程中是不可或缺的。可见光与红外融合数据集主要有 OTCBVS、LITIV、GTOT、RGBT210、RGBT234、VOT-2016 和 LasHeR 等。这些数据集在各自的历史发展阶段中, 都有力推动了可见光和红外融合目标跟踪技术的研究。

3.1.1 OTCBVS 数据集

OTCBVS 数据集^[57] 是一个公开的基准数据集, 可以用于测试和评估可见光与红外领域的算法。这项工作是由 Riad I. Hammoud 在 2004 年发起的。它由 14 个子数据集组合而成, 分为 7 个红外数据集、1 个可见光数据集、6 个可见光-红外数据集, 包含行人、面部、动作、武器、车辆、船舶等目标 (图 3)。



图 3 OTCBVS 数据集图例

3.1.2 LITIV 数据集

LITIV 数据集^[58] 是利用可见光和红外摄像机以每秒 30 帧的速度, 在不同的场景和不同的时间进行拍摄而得到的图像数据集, 图像分辨率为 320×240 像素。LITIV 数据集中共包括 9 个视频序列。

3.1.3 GTOT 数据集

灰度红外目标跟踪 (GTOT, grayscale-thermal object tracking) 数据集^[38] 包含 50 个不同场景下的视频, 包括道路、水池、实验室等区域, 共约 15 800 帧。其中, 标注的被跟踪目标分为 4 类, 包括车辆、人、天鹅等。

3.1.4 RGBT210、RGBT234 数据集

RGBT210 数据集^[29] 是由一个热红外成像仪 (DLS-

H37DM-A) 和一个 CCD 相机 (SONY EXView HAD CC) 拍摄得到的, 其中包含 210 个视频集, 共约 210 000 帧, 每个视频集最多包含 8 000 帧。

RGBT234 数据集^[59]包含 234 个视频集, 共约 233 800 帧, 每个视频集包括该视频的可见光和红外视频序列。相较于 RGBT210 数据集, RGBT234 数据集拓展了场景的多样性, 增加了在炎热天气下捕获的视频 (图 4)。基于 RGBT234 数据集, 衍生出了 2019 年举办的 Visual Object Tracking 挑战赛数据集 VOT19-RGBT, 该数据集包含 234 个序列, 并且所有序列都根据 VOT 序列聚类协议在 11 维全局属性空间中聚类。



图 4 RGBT234 和 RGBT210 数据集图例

3.1.5 VOT-2016 数据集

由于上述 GTOT、RGBT210 等数据集存在着一定的局限性, 比如数据集中的视频主要由同一种设备采集, 成像特性和图像分辨率基本相同, 这不利于保证目标跟踪算法在不同环境中的应用效果。为了完善数据集、增强数据多样性, “视觉目标跟踪”团队 (VOT, visual object tracking) 使用 10 种不同类型的传感器、从 9 种不同类型的数据来源中收集图像数据, 构建了 VOT-2016 数据集^[60]。该数据集的平均序列长度为 740 帧, 分辨率范围从 305×225 像素到 1920×480 像素不等, 其中的数据还包含了由温度改变而带来的红外特征变化。

3.1.6 LasHeR 数据集

针对大规模数据集短缺、成像平台单一、场景和类别

数量有限、复杂场景数据缺乏的问题, 研究人员又构建了 LasHeR 数据集^[61]。该数据集由 1 224 个可见光和红外视频对组成, 总数据量超过 730 000 帧。LasHeR 数据集收集了广泛的对象类别, 从不同的拍摄点和不同的场景, 进行了跨日夜、跨天气、跨季节的数据采集, 不仅对每一帧进行了空间对齐, 而且还使用边界框进行了手动注释。

3.2 评价指标

可见光和红外融合的目标跟踪性能评估常用的指标有 5 种^[62-63], 即精确率、成功率、准确性、鲁棒性和预期平均重叠 (见表 1)。

表 1 跟踪性能评价指标

评价指标	含义
精确率	其输出位置在给定的真实阈值距离内的帧的比例。
成功率	目标重叠区域大于阈值的成功帧数的比例。
准确性	衡量预测的边界框与真实边界框的重叠程度, $S = \frac{ Area(B_p \cap B_g) }{ Area(B_p \cup B_g) }$ 其中: B_p 和 B_g 分别表示预测边界框和真值。
鲁棒性	衡量跟踪器在跟踪过程中丢失目标 (跟踪失败) 的次数。
预期平均重叠	在一个短时图像序列上的非重置重叠的期望值。

3.3 基于深度学习的不同跟踪方法的性能

可见光与红外融合目标跟踪在引入了深度学习技术之后, 跟踪的效果实现了较大提升 (表 2)。但是, 相比于单一的可见光目标跟踪, 可见光与红外图像的融合处理会导致识别速度的降低。提高实时性将是可见光与红外融合目标跟踪领域需要研究的一个问题。

4 未来发展方向

近年来, 可见光和红外融合目标跟踪领域的发展日新月异, 但仍存在一些问题值得研究, 一方面亟待扩充符合要求的可见光和红外图像, 另一方面需要建立可见光与红外融合目标跟踪方法体系, 从而进一步推动该领域的发展。

表 2 部分可见光与红外融合跟踪器的性能表现

Methods	Year	VOT-2016			GTOT-50		RGBT234		实时性
		预期平均重叠	准确性	鲁棒性	精确率 (ALL)	成功率 (ALL)	精确率 (ALL)	成功率 (ALL)	
C-COT	2016	0.331	1.98	1.95	—	—	—	—	Y
MDNet	2016	0.257	1.72	2.80	81.2	63.3	71	49	N
SiamFC	2016	0.277	1.30	3.17	65.5	56.5	—	—	Y
ACFN ^[64]	2017	—	—	—	70.2	57.6	—	—	Y
CFNet ^[65]	2017	—	—	—	66.1	55.3	55.1	38.0	Y
DAT ^[66]	2018	0.320	1.47	2.10	77.1	61.8	73.1	50.3	Y
SiamDW ^[67]	2019	0.240	0.53	0.46	68.0	56.5	—	—	Y
LTDA	2019	—	—	—	84.3	67.3	78.7	54.5	N
HDINet	2021	—	—	—	88.8	71.8	78.3	55.9	N
CANet	2021	—	—	—	85.8	70.2	76.8	55.3	—
CMC2R	2022	—	—	—	—	—	73.2	52.6	—
MaCNet	2020	—	—	—	88.0	71.4	79.0	55.4	Y

4.1 算法模型提升

4.1.1 发展无需对齐多模态目标跟踪方法

由于能同时捕获可见光和红外图像(或视频)的装置比普通相机的成本更高,所以可见光和红外数据集相对较少。此外,可见光和红外融合的目标跟踪对于两种图像的对齐(Align)程度有较高的要求,但是现有可见光和红外数据集中,能够完全互相对齐的可见光图像和红外图像数据量依然不够充足,而构建图像数据对齐的大规模数据集需要耗费较大的代价。因此,有必要发展无需对齐的多模态目标跟踪方法,实现在可见光和红外对齐图像数据不足的情况下提升目标跟踪的效果。在此方向上,近年来的研究工作^[68-69]取得了一定的进展,所以无需对齐的多模态目标跟踪方法在理论上是可行的^[56]。

4.1.2 在跟踪过程中融入目标的运动模型

在可见光和红外融合目标跟踪任务中,经常会出现尺度变化和热交叉(TC, thermal crossover)的情况,在此情形中很多跟踪器无法较好地完成任务。此外,当前可见光和红外融合的目标跟踪算法在面临运动目标图像模糊问题时,局限性依然存在。针对上述问题,在未来的研究中,可以考虑在可见光和红外融合目标跟踪的过程中,融入被跟踪目标的运动模型,以便利用更多、更丰富的空间和时间信息来提升目标跟踪的效果^[47]。

4.1.3 进一步研究丢失目标重检测的方法

目前,许多可见光和红外融合的目标跟踪算法在具有不确定性因素的环境中,缺少丢失目标后重新进行检测的功能,一旦丢失被跟踪的目标,难以进行二次跟踪。例如环境中存在遮挡物,容易致使同一个目标在遮挡前后被算法分别识别成两个不同的目标,这就使得算法在这类情形下的跟踪准确性较低,甚至导致跟踪任务失败。对于此类问题,需要进一步研究丢失目标后进行重新检测和跟踪的方法,使得在有干扰物存在的复杂场景下,可见光和红外融合的目标跟踪算法依然可以有效工作。

4.1.4 加强基于成熟模型的迁移学习研究

在可见光与红外融合图像数据量较少的客观情况下,依赖大数据量的深度学习方法受到了较多的限制,不能很好地达到预期效果。因此,可以在纯可见光目标跟踪预训练模型的基础上,开展迁移学习技术研究,对已有可见光目标跟踪成熟模型进行适应性调整和改进,从而尽可能多地吸收和利用可见光跟踪模型学习到的色彩、纹理等细节特征,来提升可见光和红外融合的目标跟踪算法的效果。

4.1.5 研究无监督或弱监督目标跟踪方法

可见光和红外数据集的人工标注存在工作量大、耗时长的问题,同时保证标注质量需要花费的人力成本很高。在当前此类数据集标注量相对欠缺的情况下,可以考虑研究基于无监督或弱监督的可见光和红外融合目标跟踪方法,并且研究开发样本自动标注算法,提高标注效率。这样能够在一定程度上缓解数据集标注量不足的问题。

4.1.6 研究探索目标跟踪新方法和新框架

近年来,Transformer 和注意力机制的引入,使得可见光

和红外融合的目标跟踪在性能上得以继续提升。由此可以看出,探索新方法和新框架是提升可见光和红外融合目标跟踪算法性能的有效途径,创新性的图像处理思想和机制将不断推动可见光和红外融合目标跟踪领域持续向前发展。

4.2 技术转化应用

随着可见光与红外融合目标跟踪技术的逐步完善,该技术将在未来各种实际场景中产生很大的应用价值。

4.2.1 提升人物检测和跟踪水平

与传统的可见光目标跟踪技术相比,在现实复杂场景(如低照度或者低能见度的环境)中,可见光与红外融合目标跟踪技术具有较为明显的优势,体现出了一定的鲁棒性,例如在夜晚对行人进行检测、跟踪和数量统计。另外,在新冠疫情背景下,该技术也能够应用于检测进入特定场所的人员的口罩佩戴情况甚至是体温状况^[70]。除此之外,可见光与红外融合目标跟踪技术还可被用于人体模型的构建^[71]。虽然,可见光与红外融合目标跟踪技术在人物检测和跟踪上尚未有大规模的应用,但是未来有望在此领域发挥较大的效用。

4.2.2 推动多样化物体检测发展

可见光与红外融合目标跟踪也适用于多样化的物体检测。例如,科研人员基于该技术,从多光谱和红外遥感信息中解读出内蒙古等我国西北干旱地区的土壤盐渍化情况^[72]。另外,针对电力设备的在线监测和故障检测需求^[73],可以依托可见光与红外融合目标跟踪技术,在移动平台上(例如无人机、无人车)建立适应尺度变化、位移变化的电力设备检测系统,提升了对电力设备的故障点判断水平。此外,在消防安全方面,火灾现场的可见光图像易受到环境干扰,但其红外图像则由于明显的热效应而便于进行红外跟踪。因此,根据这个特点,可以构建基于可见光与红外融合目标跟踪的火灾定位方法^[74],对于火情信息检测、火灾控制扑救具有重要意义。

4.2.3 实现全天候复杂环境感知

可见光与红外融合目标跟踪对于提高计算机视觉技术的环境适应能力具有重要的意义。近年来,基于可见光与红外融合目标跟踪的自动驾驶技术受到日益增多的关注,将红外信息引入自动驾驶,可以有效提高自动驾驶车辆在夜晚、浓雾等恶劣条件下的安全行驶能力。另外,该技术还为全时效无人机的自动作业提供了技术保障,支撑无人机实现了夜晚条件和复杂气象条件下的图像采集和目标跟踪的功能^[75]。

5 结束语

近年来,可见光与红外融合目标跟踪技术发展迅速。本文梳理了当前可见光与红外融合目标跟踪的主流方法,将这些方法分为经典方法和基于深度学习的方法,其中经典的方法分为早期方法、基于相关滤波的方法、基于图的方法、基于稀疏表示的方法。针对深度学习方向,可以分为基于 Transformer 的方法、基于注意力机制的方法、基于时间序列的方法、自适应融合的方法、基于多模态编解码

器的方法。此外, 本文还介绍了当前该领域常用的数据集以及常见的评价指标, 并对该领域的未来发展方向进行了讨论和展望。

参考文献:

- [1] LUO Z. Object tracking for automatic driving [C] //2021 2nd International Conference on Computing and Data Science (CDS), IEEE, 2021: 265–269.
- [2] LI Y, FU C, HUANG Z, et al. Key filter-aware real-time UAV object tracking [C] //2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020: 193–199.
- [3] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking [C] // European Conference on Computer Vision (ECCV16). Springer International Publishing, 2016.
- [4] KALMAN R E. A new approach to linear filtering and prediction problems [J]. Transactions of the ASME-Journal of Basic Engineering, 1960, 1: 35–45.
- [5] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision [C] //7th International Joint Conference on Artificial Intelligence (IJCAI–81) International Joint Conferences on Artificial Intelligence, Inc., 1981: 657–662.
- [6] YILMAZ A, JAVED O, SHAH M. Object tracking: a survey [J]. ACM Computing Surveys (CSUR), 2006, 38 (4): 13.
- [7] ZHANG X, YE P, LEUNG H, et al. Object fusion tracking based on visible and infrared images; A comprehensive review [J]. Information Fusion, 2020, 63: 166–187.
- [8] MORRIS N J W, AVIDAN S, MATUSIK W, et al. Statistics of infrared images [C] //2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007: 1–7.
- [9] SCHNELLE S R, CHAN A L. Enhanced target tracking through infrared-visible image fusion [C] //14th International Conference on Information Fusion, IEEE, 2011: 1–8.
- [10] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C] //2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010: 2544–2550.
- [11] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015: 3074–3082.
- [12] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4303–4311.
- [13] DANELLJAN M, ROBINSON A, SHAHBAZ K F, et al. Beyond correlation filters; learning continuous convolution operators for visual tracking [C] //European Conference on Computer Vision. Springer, Cham, 2016: 472–488.
- [14] BUNYAK F, PALANIAPPAN K, NATH S K, et al. Geodesic active contour based fusion of visible and infrared video for persistent object tracking [C] //2007 IEEE Workshop on Applications of Computer Vision (WACV'07), IEEE, 2007: 35–35.
- [15] YUN X, JING Z, XIAO G, et al. A compressive tracking based on time-space Kalman fusion model [J]. Science China Information Sciences, 2016, 59 (1): 1–15.
- [16] GORDON N J, SALMOND D J, SMITH A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation [C] //IEE Proceedings F-radar and Signal Processing, IET, 1993, 140 (2): 107–113.
- [17] ISARD M, BLAKE A. Condensation—conditional density propagation for visual tracking [J]. International Journal of Computer Vision, 1998, 29 (1): 5–28.
- [18] CVEJIC N, NIKOLOV S G, KNOWLES H D, et al. The effect of pixel-level fusion on object tracking in multi-sensor surveillance video [C] //2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1–7.
- [19] XIAO G, YUN X, WU J. A new tracking approach for visible and infrared sequences based on tracking-before-fusion [J]. International Journal of Dynamics and Control, 2016, 4 (1): 40–51.
- [20] FUKUNAGA K, HOSTETLER L. The estimation of the gradient of a density function, with applications in pattern recognition [J]. IEEE Transactions on information theory, 1975, 21 (1): 32–40.
- [21] XIAO G, YUN X, WU J M. A multi-cue mean-shift target tracking approach based on fuzzified region dynamic image fusion [J]. Science China Information Sciences, 2012, 55 (3): 577–589.
- [22] CONAIRE C ó, O'CONNOR N E, SMEATON A. Thermo-visual feature fusion for object tracking using multiple spatio-gram trackers [J]. Machine Vision and Applications, 2008, 19 (5): 483–494.
- [23] WANG Y, LI C, TANG J. Learning soft-consistent correlation filters for RGB-T object tracking [C] //Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Springer, Cham, 2018: 295–306.
- [24] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4021–4029.
- [25] ZHAI S, SHAO P, LIANG X, et al. Fast RGB-T tracking via cross-modal correlation filters [J]. Neurocomputing, 2019, 334: 172–181.
- [26] YUN X, SUN Y, YANG X, et al. Discriminative fusion correlation learning for visible and infrared tracking [J]. Mathematical Problems in Engineering, 2019: 1–11.
- [27] LUO C, SUN B, YANG K, et al. Thermal infrared and visible sequences fusion tracking based on a hybrid tracking framework with adaptive weighting scheme [J]. Infrared Physics &

- Technology, 2019, 99: 265–276.
- [28] REN K, ZHANG X, HAN Y, et al. Robust night target tracking via infrared and visible video fusion [C] //Applications of Digital Image Processing XLI. SPIE, 2018, 10752: 43–52.
- [29] LI C, ZHAO N, LU Y, et al. Weighted sparse representation regularized graph learning for RGB-T object tracking [C] //Proceedings of the 25th ACM international conference on Multimedia, 2017: 1856–1864.
- [30] LI C, ZHU C, HUANG Y, et al. Cross-modal ranking with soft consistency and noisy labels for robust RGB-T tracking [C] //Proceedings of the European Conference on Computer Vision (ECCV), 2018: 808–823.
- [31] LI C, ZHU C, ZHANG J, et al. Learning local-global multi-graph descriptors for RGB-T object tracking [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29 (10): 2913–2926.
- [32] LI C, ZHU C, ZHENG S, et al. Two-stage modality-graphs regularized manifold ranking for rgb-t tracking [J]. Signal Processing: Image Communication, 2018, 68: 207–217.
- [33] MEI X, LING H. Robust visual tracking using ℓ_1 minimization [C] //2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009: 1436–1443.
- [34] WU Y, BLASCH E, CHEN G, et al. Multiple source data fusion via sparse representation for robust visual tracking [C] //14th International Conference on Information Fusion, IEEE, 2011: 1–8.
- [35] LIU H P, SUN F C. Fusion tracking in color and infrared images using joint sparse representation [J]. Science China Information Sciences, 2012, 55 (3): 590–599.
- [36] LI C, HU S, GAO S, et al. Real-time grayscale-thermal tracking via laplacian sparse representation [C] //International Conference on Multimedia Modeling. Springer, Cham, 2016: 54–65.
- [37] LI L, LI C, TU Z Z, et al. A fusion approach to grayscale-thermal tracking with cross-modal sparse representation [C] //Chinese Conference on Image and Graphics Technologies. Springer, Singapore, 2018: 494–505.
- [38] LAN X, YE M, ZHANG S, et al. Modality-correlation-aware sparse representation for RGB-infrared object tracking [J]. Pattern Recognition Letters, 2020, 130: 12–20.
- [39] LI C, CHENG H, HU S, et al. Learning collaborative sparse representation for grayscale-thermal tracking [J]. IEEE Transactions on Image Processing, 2016, 25 (12): 5743–5756.
- [40] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 6000–6010.
- [41] LIU X, LUO Y, YAN K, et al. CMC2R: Cross - modal collaborative contextual representation for RGBT tracking [J]. IET Image Processing, 2022, 16 (5): 1500–1510.
- [42] MEI J, ZHOU D, CAO J, et al. HDINet: hierarchical dual-sensor interaction network for RGBT tracking [J]. IEEE Sensors Journal, 2021, 21 (15): 16915–16926.
- [43] ZHANG H, ZHANG L, ZHUO L, et al. Object tracking in RGB-T videos using modal-aware attention network and competitive learning [J]. Sensors, 2020, 20 (2): 393.
- [44] YANG Y, LIANG H, YANG Y, et al. Cross-modal attention network for RGB-T tracking [C] //2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC), IEEE, 2021: 341–346.
- [45] YANG R, ZHU Y, WANG X, et al. Learning target-oriented dual attention for robust RGB-T tracking [J]. arXiv preprint arXiv: 1908.04441, 2019, <https://doi.org/10.48550/arXiv.1908.04441>.
- [46] TU Z, LIN C, ZHAO W, et al. M5L: Multi-modal multi-margin metric learning for RGBT tracking [J]. IEEE Transactions on Image Processing, 2021, 31: 85–98.
- [47] WANG Y, WEI X, TANG X, et al. Adaptive fusion CNN features for RGBT object tracking [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23 (7): 7831–7840.
- [48] RONDAO D, AOUF N, RICHARDSON M A. ChiNet: deep recurrent convolutional learning for multimodal spacecraft pose estimation [J]. arXiv Preprint arXiv: 2108.10282, 2021, <https://doi.org/10.48550/arXiv.2108.10282>.
- [49] TANG Z, XU T, WU X J. Temporal aggregation for adaptive RGBT tracking [J]. arXiv preprint arXiv: 2201.08949, 2022, <https://doi.org/10.48550/arXiv.2201.08949>.
- [50] PENG J, ZHAO H, HU Z. Dynamic fusion network for RGBT tracking [J]. arXiv preprint arXiv: 2109.07662, 2021, <https://doi.org/10.48550/arXiv.2109.07662>.
- [51] TANG H, WANG Y, CHAU L P. TAFNet: A three-stream adaptive fusion network for RGB-T crowd counting [J]. arXiv preprint arXiv: 2202.08517, 2022, <https://doi.org/10.48550/arXiv.2202.08517>.
- [52] ZENG X, LUO Z, XIONG X. Reinforced background aware correlation filter with adaptive weighting strategy for RGB-T tracking [C] //2021 4th International Conference on Pattern Recognition and Artificial Intelligence, IEEE, 2021: 8–12.
- [53] WANG Y, WEI X, TANG X, et al. Response map evaluation for RGBT tracking [J]. Neural Computing and Applications, 2022: 1–13.
- [54] LI Y, LAI H, WANG L, et al. Multibranch adaptive fusion network for RGBT tracking [J]. IEEE Sensors Journal, 2022, 22 (7): 7084–7093.
- [55] ZHU Y, LI C, TANG J, et al. RGBT tracking by trident fusion network [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32 (2): 579–592.
- [56] TU Z, LI Z, LI C, et al. Multi-interactive dual-decoder for RGB-thermal salient object detection [J]. IEEE Transactions on Image Processing, 2021, 30: 5678–5691.
- [57] DAVIS J W, SHARMA V. Background-subtraction using contour-based fusion of thermal and visible imagery [J]. Comput-

- er Vision and Image Understanding, 2007, 106 (2/3): 162 - 182.
- [58] TORABI A, MASSÉ G, BILODEAU G A. An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications [J]. Computer Vision and Image Understanding, 2012, 116 (2): 210 - 221.
- [59] LI C, LIANG X, LU Y, et al. RGB-T object tracking: Benchmark and baseline [J]. Pattern Recognition, 2019, 96: 106977.
- [60] LEBEDA K, HADFIELD S J, BOWDEN R. The thermal infrared visual object tracking VOT-TIR2016 challenge result [C] //Proceedings, European Conference on Computer Vision (ECCV) Workshops, 2016: 1 - 26.
- [61] LI C, XUE W, JIA Y, et al. Lasher: a large-scale high-diversity benchmark for RGBT tracking [J]. IEEE Transactions on Image Processing, 2021, 31: 392 - 404.
- [62] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411 - 2418.
- [63] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking vot2017 challenge results [C] //Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 1949 - 1972.
- [64] CHOI J, JIN CHANG H, YUN S, et al. Attentional correlation filter network for adaptive visual tracking [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4807 - 4816.
- [65] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4807 - 4816.
- (上接第 6 页)
- [42] 胡海峰, 宋征宇, 孙海峰. 新型中型运载火箭控制系统架构及发展 [J]. 导弹与航天运载技术, 2019 (2): 49 - 53.
- [43] TAO F, QI Q. Make more digital twins [J]. Nature, 2019: 490 - 491.
- [44] 陶 飞, 刘蔚然, 张 萌, 等. 数字孪生五维模型及十大领域应用 [J]. 计算机集成制造系统, 2019, 25 (1): 1 - 18.
- [45] ZHANG X, JIANG D, HAN T, et al. Rotating machinery fault diagnosis for imbalanced data based on fast clustering algorithm and support vector machine [J]. Journal of Sensors, 2017 (12): 1 - 15.
- [46] BAPTISTA M, et al. Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling [J]. Computers & Industrial Engineering, 2018, 115: 41 - 53.
- [47] SCHLEICH B, ANWER N, MATHIEU L, et al. Shaping the digital twin for design and production engineering [J]. CIRP Annals, 2017, 66 (1): 141 - 144.
- [48] 刘秀罗, 王 佳, 尚国强, 等. 航天发射场数字孪生技术初探 [J]. 西北工业大学学报, 2019, 37: 114 - 119.
- [49] 陶 飞, 程 颖, 程江峰, 等. 数字孪生车间信息物理融合理论与技术 [J]. 计算机集成制造系统, 2017, 23 (8): 1603 - 1611.
- [50] HA N, XU K, REN G, et al. Machine learning-enabled smart sensor systems [J]. Advanced Intelligent Systems, 2020, 2 (9): 2000063.
- [51] 陶 飞, 张 萌, 程江峰, 等. 数字孪生车间——一种未来车间运行新模式 [J]. 计算机集成制造系统, 2017, 23 (1): 1 - 9.
- [52] TAO F, SUI F, LIU A, et al. Digital twin-driven product design framework [J]. International Journal of Production Research, 2019, 57 (12): 3935 - 3953.
- [53] JORDAN M I, MITCHELL T M. Machine learning: trends, perspectives, and prospects [J]. Science, 2015, 349 (6245): 255 - 260.
- [54] XIA M, SHAO H, WILLIAMS D, et al. Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning [J]. Reliability Engineering & System Safety, 2021, 215: 107938.
- [55] ZHANG W, HAN G, WANG J, et al. A BP neural network prediction model based on dynamic cuckoo search optimization algorithm for industrial equipment fault prediction [J]. IEEE Access, 2019, 7: 11736 - 11746.
- [66] PU S, SONG Y, MA C, et al. Deep attentive tracking via reciprocal learning [J]. Advances in Neural Information Processing Systems, 2018, 31: 1931 - 1941.
- [67] ZHANG Z, PENG H. Deeper and wider Siamese networks for real-time visual tracking [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4591 - 4600.
- [68] FAN D P, LIN Z, JI G P, et al. Taking a deeper look at co-saliency object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2919 - 2929.
- [69] JIN W D, XU J, CHENG M M, et al. Inet: Intra-saliency correlation network for co-saliency detection [J]. Advances in Neural Information Processing Systems, 2020, 33: 18749 - 18759.
- [70] 白梅娟, 查祖福水, 王杨洋, 等. 可见光与红外图像联合口罩佩戴检测研究 [J/OL]. 控制工程: 1 - 9 [2022-07-17]. DOI:10.14107/j.cnki.kzgc.20220073.
- [71] LIPTON A J, FUJIYOSHI H, PATIL R S. Moving target classification and tracking from real-time video [C] //Proceedings Fourth IEEE Workshop on Applications of Computer Vision, WACV'98 (Cat. No. 98EX201). IEEE, 1998: 8 - 14.
- [72] 崔文轩. 基于无人机热红外遥感的土壤盐渍化监测模型研究 [D]. 杨凌: 西北农林科技大学, 2021.
- [73] 国 朕. 基于红外与可见光图像融合算法及其在电力设备检测中的应用 [D]. 扬州: 扬州大学, 2021.
- [74] 段琳锋, 侯新国. 基于红外与可见光图像火灾定位研究 [J]. 舰船电子工程, 2022, 42 (3): 183 - 187.
- [75] 陈海华, 杨 磊, 陈凤翔, 等. 全时效巡检无人机的研究与应用 [J]. 科技创新与应用, 2021, 11 (20): 164 - 166.