DOI:10.13878/j.cnki.jnuist.2019.06.006

丁正彤[1] 徐磊[1] 张研[1] 李飘扬[1] 李阳阳[1] 罗斌[1] 涂铮铮[1]

RGB-T 目标跟踪综述

摘要

RGB-T 目标跟踪是基于 RGB 目标跟踪问题发展而来的.为了提高复杂环境下的目标跟踪性能,学者们提出结合可见光和热红外的信息来克服单一成像受限的问题.本文首先介绍了 RGB-T 目标跟踪的研究背景,并指出该任务所面临的挑战,然后归纳并介绍了目前已有的 RGB-T 目标跟踪的几类方法,包括传统方法和深度学习方法.最后,本文对现有的 RGB-T 数据集、评价指标进行了分析和对比,并指出 RGB-T 跟踪中值得研究的方面.

关键词

可见光-热红外;多模态;目标跟踪

中图分类号 TP391

文献标志码 A

收稿日期 2019-10-18

资助项目 国家自然科学基金(61602006)

作者简介

丁正彤,男,研究方向为智能视频分析. dingzhengtong@ foxmail.com

涂铮铮(通信作者),女,博士,副教授,硕士生导师,主要研究方向为模式识别与图像处理.zhengzhengahu@ 163.com

1 安徽大学 计算机科学与技术学院,合肥,230601

0 引言

视觉目标跟踪,旨在从连续视频帧中估计出目标在每一帧中的位置和尺度信息,是计算机视觉中的一个热点问题,在视频监控、自动驾驶和机器人感知等方面有着广泛的应用.尽管目标跟踪取得了许多重要突破[1-8],但现阶段的目标跟踪仍然面临许多挑战性问题,尤其是在各种复杂的环境条件下(如低光照、雨天、烟雾等),可见光图像的成像质量受到显著影响,使得跟踪目标物体是非常困难的.

热红外成像主要的优势体现在:它可以捕捉到目标所发出的热辐射,对光照变化不敏感,可以实现在零光照条件下跟踪目标;它还具有很强的穿透烟雾的能力,使得 RGB-T(RGB-Thermal,可见光-热红外)目标跟踪比传统目标跟踪具有更强的潜在应用价值.

因此,结合可见光和热红外信息可以有效地提高目标跟踪性能,较好地实现全天时全天候的目标跟踪.图 1 是低光照(左)和强光照(右)情况下的可见光图像,目标在其中并不明显,但在图 2 相对应的热红外图像中目标轮廓清晰.图 3 呈现的两个热红外图像产生了热交叉现象,目标和背景极难区分,但目标在图 4 相对应的可见光图像中较为明显[9].可见,可见光和热红外信息相互补充,有助于复杂环境条件下的视觉跟踪.

最近几年,随着热红外传感器的普及,RGB-T 跟踪在计算机视觉领域引起了越来越多的关注.由于 RBG-T 目标跟踪相对于单模态目标跟踪起步较晚,至今鲜见关于 RGB-T 目标跟踪进展的文献综述.本文将对该领域前后发展进行一个较为全面的综述.首先介绍 RGB-T 目标跟踪面临的挑战,然后回顾传统的 RGB-T 目标跟踪算法,主要包括基于人工设计特征和传统的机器学习技术的 RGB-T 目标跟踪算法,再介绍近几年出现的基于深度学习的 RGB-T 目标跟踪方法,最后对已有的多个 RGB-T 数据集、评价指标进行分析和对比.

1 RGB-T 目标跟踪的挑战

一些早期的研究[10-11]表明将可见光和热红外数据结合起来,可以有效地提高跟踪性能.相对于传统的单模态目标跟踪,借助红外信息构建的多模态目标跟踪,其跟踪效果得到进一步提升,但在面对更复杂场景的情况下,RGB-T 目标跟踪不仅遇到传统的目标跟踪所面临的挑战,而且也遇到新的挑战.

图 1　可见光图像(在低光照和强光照情况下难以追踪目标)[9]

Fig. 1　Visible images,where the visible spectrum is disturbed by low or high illumination[9]

图 2　热红外图像(在低光照和强光照情况下容易追踪目标)[9]

Fig. 2　Thermal infrared images can overcome the disturbance by low or high illumination[9]

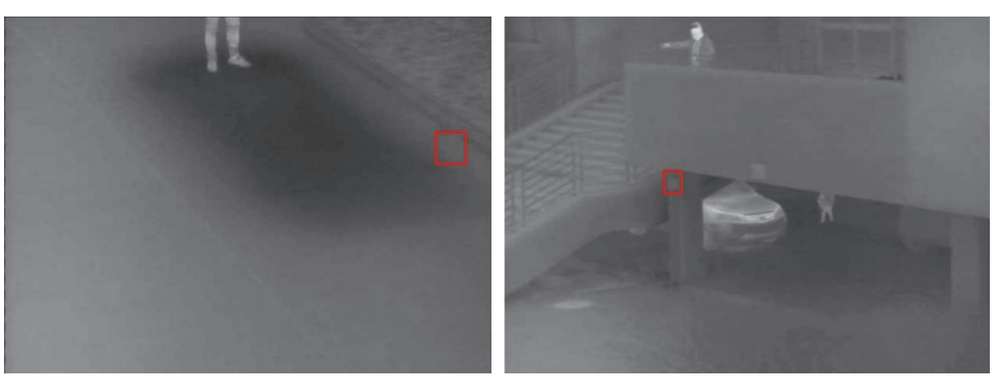

图 3　热红外图像(在热交叉情况下难以追踪目标)[9]

Fig. 3　Thermal infrared images,where thermal spectrum is disturbed by thermal crossover[9]

1.1　传统的目标跟踪的挑战举例

1)目标的形变与快速运动:当跟踪的目标发生较大的形变或尺度变化剧烈时,跟踪框不能及时适应变化,将会引入过多的背景信息污染模型,从而导致跟踪目标的丢失.另外对于快速运动的物体,由于相邻两帧的目标移动跨度较大,超出了候选区域,也将导致跟踪失败.

2)遮挡:目标被遮挡可以分为部分遮挡和完全遮挡.如果目标是部分被遮挡,在遇到遮挡物的第一帧,边界框会将遮挡物的信息包含在内,导致后续跟踪过程中的目标被错误判别;如果是完全遮挡,边界框找不到目标,会直接导致跟踪失败.如图 5 所示,红色框内跟踪目标人物在第 2 张图中部分被树叶遮挡,导致跟踪框内可能包含树叶信息,而在第 3 张图

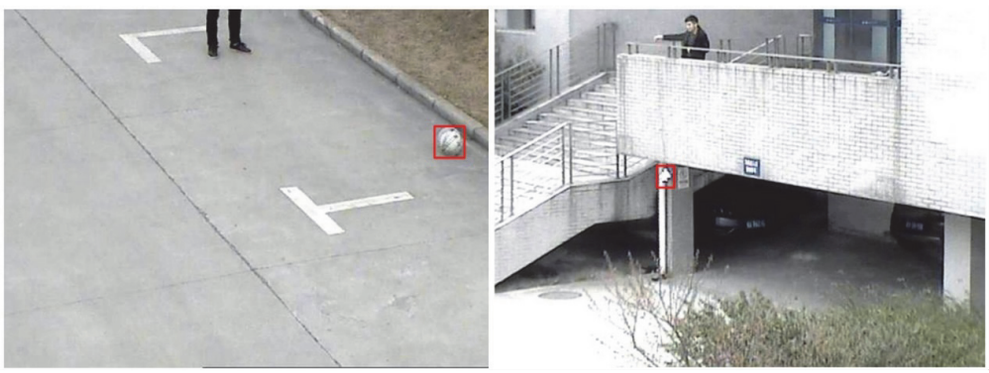

图 4　可见光图像(在热交叉情况下容易追踪目标)[9]

Fig. 4　Visible images are not influenced by thermal crossover[9]

中则是完全被树叶遮挡,候选框内不能找到目标特征,导致目标丢失.

1.2　RGB-T 目标跟踪的新挑战

1)RGB-T 融合:如何将 RGB 和热红外两个模态进行有效的融合是 RGB-T 目标跟踪面临的挑战之一.如果两个模态中的一个模态成像不佳,则直接融合两个模态将会引入噪声,进而影响跟踪性能,所以两个模态的融合策略直接影响 RGB-T 跟踪性能.

2)特征表示:与传统目标跟踪相比,RGB-T 目标跟踪的目标特征由 RGB 与 T 特征共同描述,更鲁棒的 RGB-T 特征表示必然可以提升跟踪的性能,这一点也得到了越来越多的关注.

3)成像受限:在零光照、光线强烈变化、雾霾等情况下,可见光谱成像受限;当目标与周边背景物体的温度差异较小时,则会有热交叉现象发生,热红外成像受限.

2　传统方法

RGB-T 目标跟踪的传统方法多为在线目标跟踪,旨在利用简单有效的人工设计视觉特征,结合浅层外观模型,利用匹配或分类算法进行快速有效的目标跟踪.由于稀疏表示在抑制噪声、减少误差方面表现较好,故稀疏表示理论也被用于解决 RGB-T 目标跟踪问题[12-16],并且取得了较好的效果.但稀疏表示模型计算复杂度较高,难以实时处理.随着相关滤波理论在单模态跟踪中取得了较为鲁棒的效果[17-19],Zhai 等[20]在 RGB-T 目标跟踪中引入交叉模态相关滤波器,更有效地进行可见光和红外模态的融合.为了改善 RGB-T 目标跟踪中的模型漂移现象,研究者在 RGB-T 目标跟踪中引入图的理论[21-23],自适应地使用可见光和红外图像信息来学习模态权重.下面将从以下 3 个方面详细对 RGB-T 跟踪的传统方法进行阐述.

2.1　基于稀疏表示的 RGB-T 目标跟踪

近年来,使用稀疏表示的目标跟踪[11,24-25]能够较好地抑制噪声和减少误差.受此启发,Wu 等[12]将 RGB 和 T 信息结合起来,稀疏地表示目标模板空间中的每个样本;Liu 等[13]使用 RGB 和 T 信息计算出联合稀疏表示系数的最小操作来融合跟踪结果.在这些方法中,RGB 模态和 T 模态贡献相同,故在处理

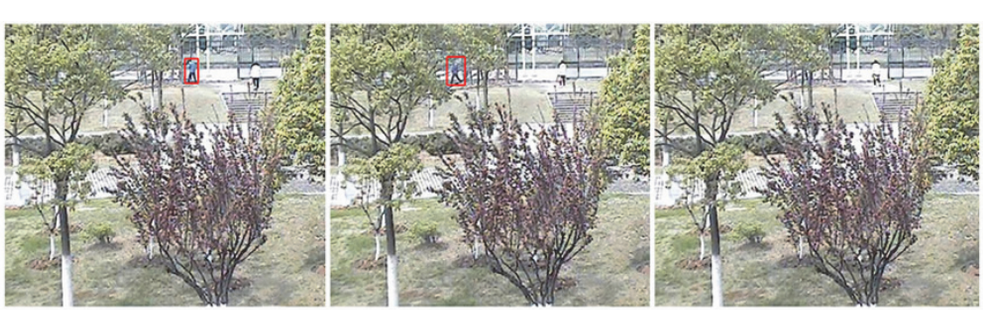

图 5　目标被遮挡时,跟踪失败[9]

Fig. 5　Tracking failed when the target is occluded[9]

干扰或者故障时可能会限制跟踪性能.

为了改善上述情况,Li 等[14]引入反映其跟踪预测可靠性的模态权值,为每种模态引入模态权值来表示该模态的可靠性,实现不同模态的自适应融合.具体来说,在贝叶斯滤波技术的框架下,一种基于协同稀疏表示的自适应融合方法被提出.这种方法在每个模态中引入权值描述模态的可靠性,找到一种自适应的目标跟踪的协作稀疏表示方法,可以自适应地融合可见光信息和红外信息,进而实现全天候地对目标进行鲁棒跟踪,当目标在一种模态中处于不稳定或者故障时,通过赋予不可靠模态信息低权值,利用可靠的模态进行信息补充用于跟踪[10],通过在线方式联合优化稀疏码、不同模态下的权值和最大似然判别法[26]对稀疏码进行有效的优化,并利用封闭形式解法进行求解,能够避免在目标跟踪中产生的模型漂移.该方法可增强跟踪鲁棒性,并防止之前视频帧的可见光和红外信息的累积产生的外观污染问题的发生.

此外,由于每个模态中目标外观的较大变化或背景干扰会给采集的样本带来一些噪声,进而影响分类器的学习,并且视觉和运动特征在不同模态中差异较大.Lan 等[16]针对 RGB-T 目标跟踪,提出了一种新颖的判别学习模型,可以消除由于较大变化产生的异常样本,并学习来自不同模态的具有判别一致性的特征,而且可以协作完成模态可靠性度量及目标与背景的分离,取得了较好的效果.

2.2 基于相关滤波的 RGB-T 目标跟踪

大多数现有单模态方法采用基于贝叶斯滤波框架的稀疏表示去跟踪目标,这些跟踪器在加入红外信息这个模态后可能会受到如下限制:一是对可见光和红外信息的联合稀疏约束使得模态一致性太强而无法实现有效融合;二是为了达到有效跟踪的目的,贝叶斯滤波算法需对大量待选样本进行采样.因此稀疏表示模型的优化的计算复杂度高,耗费时间.Zhai 等[20]利用低秩约束,提出交叉模态相关滤波器以获得可见光和热红外两个模态之间的相互依赖性,实现多种模态的协同融合,使所学习的滤波器可以包含来自不同数据源的有用信息,从而获得鲁棒的跟踪结果.并对交叉模态相关滤波器利用交替方向乘法器(ADMM)优化算法[27]求解,从而实现了超实时的跟踪效果.在利用交叉模态相关滤波器进行跟踪时,最快达到 227 帧/s 的速度.

在一些特殊情况下,RGB 图像和热红外图像之间的融合可能是无效的.如果简单地采用协同稀疏表示的方法在贝叶斯滤波框架下解决,也存在着耗时的问题,无法达到目标跟踪要求的实时跟踪的目的.为了解决以上问题,Wang 等[28]提出了基于相关滤波器的多光谱方法来进行有效的目标跟踪.该方法考虑到了不同光谱信息的协同性和异质性,通过在相关滤波器中加入软一致性来部署多光谱间信息,以实现更有效的融合.同时采用快速傅里叶变化(FFT)来大大减少计算的时间,改进后的方案在进行目标跟踪时以超 50 帧/s 的运行速度展现出良好的跟踪效果.

2.3 基于图的 RGB-T 目标跟踪

由于目标跟踪需要对每一帧图像进行识别,每帧中目标的包围框都可能不同,这些框一般由 RGB 与热红外特征共同描述.由于背景信息的存在,可能导致模型漂移现象的出现.为了解决该问题,Li 等[21]提出了一种加权稀疏表示正则化图,以自适应地使用 RGB 和红外数据来学习权重.其图像块作为图节点,并以块特征矩阵作为输入进行联合的稀疏表示[13,15].为了处理各个源的偶然扰动和故障,给每种模态分配权重以表示可靠性,使得跟踪器可以自适应地融合不同模态的数据,并学习得到更有意义的图亲和矩阵.值得注意的是,通过设计的高效的 ADMM(交替方向乘器)算法[27]来联合优化模态权重、稀疏表示和图(包括结构、边缘权重和节点权重).

由于初始化过程中不正确的图像块权重会影响目标跟踪的正确性,Li 等[22]随后提出了一种新的两阶段模态图正则化流形排序算法,以学习一种更鲁棒的 RGB-T 跟踪对象表示方法.首先给定对象边界框,将其划分为一组不重叠的图片,这些图片用 RGB 和热红外特征共同描述.然后,给每个块分配一个权重,抑制表示中的背景信息,并将这些权重融合,以得到鲁棒的目标表示.该方法以一种联合的方式建立块权重和模态权重的模型,并对它们进行有效的优化.为了提高块权重的鲁棒性,采用了两阶段排序策略.第一阶段,根据初始种子计算块权重;第二阶段,以第一阶段的计算结果为基础进行权重计算.最后,应用结构化支持向量机对目标位置进行预测.

由于使用低秩和稀疏表示来学习具有全局性的动态图[29]没有考虑局部信息,可能会限制性能,例如相邻节点往往较为相似.为解决此问题,研究者提出了一种新颖的通用方法[23]来学习局部和全局多图描述符,以抑制 RGB-T 跟踪的背景信息干扰.该描述符可自动探索具有全局和局部线索的多模态图像

块之间的内在关系,其依赖于一种新颖的图学习算法,主要是用多幅图表示对象,并用一组多模态图像块作为节点,以增强对对象形变和部分遮挡的鲁棒性;将局部相邻信息强加到表示系数中,这使得学习到的图亲和矩阵也可以使用局部空间线索进行编码,并利用学习到的图亲和矩阵计算图节点权重,将多图信息与相应的图节点权重相结合,形成鲁棒的目标描述符,最后采用结构化支持向量机获得最优解作为跟踪结果.

3 基于深度学习的 RGB-T 目标跟踪

近年来,由于神经网络的广泛应用,视觉跟踪取得了新的突破.目前基于深度网络的 RGB-T 跟踪模型大致可以为三类:第一类是以密集特征聚合与剪枝网络(DAPNet)[30]、质量感知特征聚合网络(FANet)[31]和双流卷积神经网络(Two-Stream CNN)[32]为代表的多模态特征融合,利用深度网络自适应地融合可见光和热红外模态的特征,利用模态间的互补优势,获得更加鲁棒的特征,提高跟踪性能;第二类以多适配器卷积网络(MANet)[33]为例,发掘模态共享特征、模态特定特征的潜在价值以及实例感知信息,提高特征融合的质量;第三类是基于注意力机制的 RGB-T 跟踪,例如双重注意力模型(DUALATTENTION)[34].

3.1 基于多模态特征融合的 RGB-T 跟踪

在早期的特征融合研究中,Li 等[32]提出了一种新的卷积神经网络(ConvNet)结构,包括一个通用子网络(Two-Stream CNN)和一个融合子网络(FusionNet).通用子网络用来提取丰富的语义信息以有力地表示目标对象,而融合子网络用来自适应地融合多种模态的信息.具体地说,用 Two-Stream CNN 来提取不同模态的特定特征,其中一个 CNN 用于处理 RGB 流,另一个 CNN 用于处理热红外流.由于多模态特征通常包含一些冗余噪声,这会在一定程度上影响 RGB-T 跟踪的性能.FusionNet 从 Two-Stream CNN 的输出中选择有区分性的特征,以减轻冗余信息的影响,从而在提高精度的同时显著提高效率.

近来,为了有效地融合可见光和热红外信息,提高 RGB-T 跟踪的性能,Zhu 等[30]提出了一种密集特征聚合与剪枝网络(DAPNet).密集特征聚合与剪枝网络(DAPNet)由两个主要模块组成,一个是密集特征聚合,为目标对象提供丰富的 RGB-T 特征表示;另一个是特征剪枝,从聚合的 RGB-T 特征中除去噪声或冗余的特征,选择最具区分性的特征.

在密集的特征聚合模块,将所有层的特征递归地集成到同一个特征空间中,充分地利用了浅层特征和深层特征,获得更鲁棒的特征表示,实现更好的跟踪性能.但是聚合的 RGB-T 特征存在噪声和冗余,这些冗余的特征会干扰目标的定位.也就是说,只有少数通道是有益的,并且其中很大一部分通道在描述某个目标时包含冗余和不相关的信息.为了解决这个问题,Zhu 等[30]提出了一种协同特征剪枝方法来去除噪声和冗余的特征.特征剪枝模块包括两个步骤,即通道评分和通道选择.通过这种特征剪枝方法,在每次训练的迭代中停用一些特征通道,从而得到一个更可靠的卷积特征表示.训练完成后,在线跟踪过程中特征聚合网络的参数将保持不变,而特征剪枝模块将被丢弃.DAPNet 对由于形变、快速移动、背景杂波和每个模态的遮挡而导致的显著外观变化的挑战有较鲁棒的跟踪效果.

Zhu 等[31]提出过一种新的 RGB-T 融合架构——质量感知特征聚合网络(FANet).该网络由两个子网组成:分层特征聚合子网络和多模态信息聚合子网络.分层特征聚合子网以自适应方式集成层次化和多分辨率的深层特征.在分层特征聚合子网中,Zhu 等[31]还提出了一种新的特征聚合方法——密集特征聚合.浅层特征可对目标位置等空间细节进行编码,有助于实现精确的目标定位,而深层特征能更有效地捕获目标的语义特征.在每个模态中,首先将层次化的多分辨率特征聚合到相同分辨率的统一空间中,智能地学习不同层的权重,自适应地融合它们,以突出显示更多具有判别性的特征,并能够抑制噪声,多模态信息聚合子网则使用聚合的特征来预测模态整体权重,根据预测的可靠性程度协同集成所有模态,然后将模态权值与相应的聚合特征结合起来产生一个可靠的目标表示,显著提高了 RGB-T 目标跟踪性能.

3.2 基于多适配器的 RGB-T 跟踪

在多适配器卷积网络(MANet)之前的 RGB-T 追踪工作通常引入模态权重来实现自适应融合或学习不同模态的鲁棒特征表示,从而专注于特定的信息集成.虽然可以有效地利用特定模态的性质,但是它们忽略了模态共享特征的潜在价值以及实例感知信息,而这些对于 RGB-T 跟踪中不同模态的有效融合是很重要的.

Li 等[33]提出了多适配器卷积网络(MANet),用于 RGB-T 跟踪的端到端训练的深度框架,包括模态

共享、模态特定和实例感知的特征学习.MANet 包含三种适配器,包括通用适配器(Generality-Adapter)、模态适配器(Modality-Adapter)、实例适配器(Instance-Adapter).通用适配器用来提取不同模态的共享对象表示,在有效性和网络效率之间进行了良好协调;模态适配器基于通用适配器,可以有效地提取特定模态的特征表示,充分利用 RGB 和热红外模态的互补优势;实例适配器用来对特定对象的外观特性和时间变化进行建模,以解决跟踪过程中实例对象出现变化或环境变化而导致跟踪模型无法跟踪的问题.此外,通用适配器和模态适配器以并行结构方式结合以降低目标跟踪过程的计算复杂度.

3.3 基于注意力机制的 RGB-T 跟踪

视觉注意力在 RGB-T 跟踪中有着巨大的潜力,有助于分类器的学习.与前面的特征表示学习和自适应模态加权融合不同,Yang 等[34]另辟蹊径,提出了双视觉注意力机制(局部注意力和全局注意力)以实现鲁棒的 RGB-T 跟踪.通过利用 RGB 和热红外数据的共同视觉注意来训练深度分类器,从而实现局部注意力.而全局注意力是一个多模态目标驱动的注意力估计网络,它可以为分类器提供全局预测以及从先前跟踪结果中得到局部预测.

局部注意力的训练过程包括前向传播和反向传播两步.在前向传播的步骤中,将成对的 RGB 和热红外图像送入深度检测跟踪网络,并估计相应的分类得分;在反向传播的过程中,取这个分类分数相对于输入成对的 RGB-T 样本的偏导数,从最后一个全连接层朝向第一卷积层进行网络更新.将第一层的偏导数输出作为 RGB 和热红外输入的共同注意图,在训练过程中,通过在损失函数中加入此注意图作为正则化项,使分类器更加关注目标区域.

尽管前面提出的 RGB-T 跟踪器已经可以实现良好的性能,但是它仍然遵循检测跟踪框架下的局部搜索策略.由于前一帧的跟踪结果也许已经失败,局部搜索策略将不能发挥作用.而将目标驱动注意力估计网络和 RGB-T 全局注意力网络结合,可以改善局部搜索策略所不能解决的该问题.高质量的全局候选框可以从注意力区域中提取,并与局部候选框一起输入到分类器,得到有效的分类结果.因此,局部和全局注意力图的互补进一步提高了 RGB-T 目标跟踪器的鲁棒性和准确性.所以该双注意力机制(局部注意力和全局注意力)的思想在未来的目标跟踪中极具潜力.

4 可见光-红外数据集

4.1 数据集

目前的基于深度学习的 RGB-T 目标跟踪,都依赖于大型数据集来训练模型并评估其性能,被设计用于 RGB-T 目标跟踪的视频基准数据集主要有 LITIV[35]、GTOT[14]、RGBT210[21]、RGBT234[9]、VOT2019[36]以及相关测试平台的数据集.

LITIV[35]数据集由热红外和可见光摄像机以 30 帧/s 的速度、不同变焦设置和不同拍摄位置及不同跟踪场景的视频组成.图像大小为 320×240 像素.

GTOT[14]数据集包括 50 个视频对,每个视频对由一个可见光视频和一个热红外视频组成,具有 50 个不同场景,如办公区、公共道路、水池等.每个可见光视频都与一个热红外视频配对.该数据集包含非刚性、移动模糊、小物体、照明条件、热交叉、比例变化、遮挡时长与面积等挑战.

RGBT210[21]数据集包含大量高精度视频帧(总帧数约 210 000 帧).不同模态之间对齐更加准确,不需要预处理和后处理.该数据集包括对无遮挡、部分遮挡和严重遮挡的注释,可用于不同算法的遮挡敏感性评估.

RGBT234[9]数据集是基于 RGBT210 数据集扩展的大规模 RGBT 跟踪数据集.它包含总共 234 对高对齐的 RGB 和热红外视频序列,具有大约 200 000 帧,最长的视频序列达到 8 000 帧.但是此数据集中目标对象的外观随着时间的推移而显著变化,这是由遮挡、运动模糊、相机移动和照明挑战引起的,对于评估不同的跟踪器具有足够的挑战性.

VOT 是当下比较流行的跟踪算法的测试平台,包括数据集、评价标准与评价系统,且每一年都会更新.目前 VOT2019[36]已经发布,可用于 RGB-T 目标跟踪.VOT-RGBT2019 包含 60 个视频序列以及 6 个挑战,包括相机移动、光照变化、目标尺寸变化、目标动作变化、非退化 6 个属性.

表 1 列出了 RGB-T 目标跟踪领域的主要视频基准数据集(LITIV[35]、GTOT[14]、RGBT210[21]和 RGBT234[9]).

4.2 评价标准

为了评估性能,本节重点介绍 6 种广泛使用的跟踪效果评估指标:精确率(PR)、成功率(SR)、准确度(Accuracy)、鲁棒性(Robustness)、PR 曲线(PR curves)和 F 值(F-measure).

表 1　用于 RGB-T 目标跟踪的基准数据集示例

Table 1　Examples of benchmark datasets used for RGB-T object tracking

数据集	视频序列	总帧数	每个序列最大帧数	遮挡注释	相机移动	发表年份
LITIV[35]	9	6 300	1 200	无	无	2012
GTOT[14]	50	15 800	700	有	无	2016
RGBT210[21]	210	210 000	8 000	有	有	2017
RGBT234[9]	234	233 800	8 000	有	有	2018

1)精确率(PR).精确率(PR)是输出位置在给定的真值阈值距离内的帧的百分比.在某些场景下,也可以使用最大精确率(MPR)作为评价指标.

2)成功率(SR).成功率(SR)是输出边界框与真值边界框之间的重叠率大于阈值的帧的百分比.通过改变阈值,可以获得 SR 图.

3)准确度(Accuracy).准确度为对于给定的测试数据集,分类器正确分类的样本数与总样本数之比.

4)鲁棒性(Robustness).鲁棒性用来度量模型受数据扰动、噪声以及离群点的影响程度.

5)PR 曲线(PR curves).即以召回率(Recall)为横坐标,精确率为纵坐标绘制而成的曲线,通过调节分类阈值,可以得到不同的召回率和精确率,从而得 PR 曲线.

6)F 值(F-measure).F 值是精确率(PR)和召回率(Recall)的加权调和平均,精确率和召回率没有绝对联系,但在数据集合规模变大时,二者会互相制约,F 值就可以在维持二者权重相同时,综合二者特性,得出分类模型的优劣.

5　结束语

在目标跟踪过程中,外部环境因素很容易对跟踪的效果产生影响,而有效地利用可见光和热红外的互补优势,可以实现全天候的鲁棒的视觉跟踪,因此 RGB-T 目标跟踪近些年成为计算机视觉中的一个新的研究分支.本文从传统方法和深度学习方法两方面对 RGB-T 目标跟踪方面的相关研究进行阐述.传统方法分为基于稀疏表示的、基于相关滤波的、基于图模型的方法,深度学习方法分为基于多模态特征融合的、基于多适配器的、基于注意力机制的深度学习网络.

RGB-T 目标跟踪有着巨大的研究价值,可以考虑探索更深度的模态融合机制,将 RGB 和热红外这两种模态进行更有效的融合,这也是当前研究面临的难题之一,比如设计新型融合结构、进行多模交互学习等.同时,可以对目标对象进行更有效的表示,如提取出目标的掩模轮廓、关键点、概率分布等有效特征.此外 RGB-T 目标跟踪中的分类器也有待增强.这些都是未来值得研究的方向.

参考文献

References

[1] Grabner H,Grabner M,Bischof H.Real-time tracking via on-line boosting[C]//Proceedings of the 2006 British Machine Vision,2006:47-56

[2] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking[C]//European Conference on Computer Vision,2008:234-247

[3] Avidan S. Ensemble tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(2):261-271

[4] Babenko B,Yang M H,Belongie S.Robust object tracking with online multiple instance learning[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,33(8):1619-1632

[5] Kalal Z,Mikolajczyk K,Matas J.Tracking-learning-detection[J]. IEEE Transactions on Software Engineering,2011,34(7):1409-1422

[6] Hare S,Golodetz S,Saffari A,et al.Struck:structured output tracking with kernels[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(10):2096-2109

[7] Li X,Shen C H,Dick A,et al.Learning compact binary codes for visual tracking[C]//IEEE Conference on Computer Vision & Pattern Recognition,2013:2419-2426

[8] Zhang J M,Ma S G,Sclaroff S.MEEM:robust tracking via multiple experts using entropy minimization[C]//European Conference on Computer Vision,2014:188-203

[9] Li C L,Liang X Y,Lu Y J,et al.RGB-T object tracking:benchmark and baseline[J].Pattern Recognition,2018,96:106977

[10] O'Conaire C, O'Connor N E, Smeaton A. Thermo-visual feature fusion for object tracking using multiple spatiogram trackers [J]. Machine Vision and Applications,2008,19(5/6):483-494

[11] O'Conaire C,O'Connor N E,Cooke E,et al.Comparison of fusion methods for thermo-visual surveillance tracking[C]//2006 9th International Conference on Information Fusion,2006:1-7

[12] Wu Y,Blasch E,Chen G S,et al.Multiple source data fusion via sparse representation for robust visual tracking[C]//14th International Conference on Information Fusion,2011:1-8

[13] Liu H P,Sun F C.Fusion tracking in color and infrared images using joint sparse representation[J].Science China:Information Sciences,2012(3):104-113

[14] Li C L,Cheng H,Hu S Y,et al.Learning collaborative sparse representation for grayscale-thermal tracking[J].IEEE Transactions on Image Processing,2016,25(12):

5743 - 5756
[15] Lan X Y, Ma A J, Yuen P C. Multi-cue visual tracking using robust feature-level fusion based on joint sparse representation[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 1194-1201
[16] Lan X Y, Ye M, Zhang S P, et al. Robust collaborative discriminative learning for RGB-infrared tracking[C]// 32nd AAAI Conference on Artificial Intelligence, 2018: 7008-7015
[17] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// The 23rd IEEE Conference on Computer Vision and Pattern Recognition, 2010, DOI: 10.1109/CVPR.2010.5539960
[18] 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
[19] Bai B, Zhong B N, Ouyang G, et al. Kernel correlation filters for visual tracking with adaptive fusion of heterogeneous cues[J]. Neurocomputing, 2018, 286: 109-120
[20] Zhai S Y, Shao P P, Liang X Y, et al. Fast RGB-T tracking via cross-modal correlation filters[J]. Neurocomputing, 2019, 334: 172-181
[21] Li C L, Zhao N, Lu Y J, 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
[22] Li C L, Zhu C L, Zheng S F, et al. Two-stage modality-graphs regularized manifold ranking for RGB-T tracking[J]. Signal Processing: Image Communication, 2018, 68: 207-217
[23] Li C L, Zhu C L, 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
[24] Li C L, Sun X, Wang X, et al. Grayscale-thermal object tracking via multitask Laplacian sparse representation[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 47(4): 673-681
[25] Gade R, Moeslund T B. Thermal cameras and applications: a survey[J]. Machine Vision and Applications, 2014, 25(1): 245-262
[26] Parikh N, Boyd S. Proximal algorithms[J]. Foundations and Trends ® in Optimization, 2014, 1(3): 127-239
[27] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends ® in Machine Learning, 2011, 3(1): 1-122
[28] Wang Y L, Li C L, Tang J. Learning soft-consistent correlation filters for RGB-T object tracking[C]// Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2018: 295-306
[29] Li C L, Lin L, Zuo W M, et al. Learning patch-based dynamic graph for visual tracking[C]// Thirty-First AAAI Conference on Artificial Intelligence, 2017: 4126-4132
[30] Zhu Y B, Li C L, Luo B, et al. Dense feature aggregation and pruning for RGBT tracking[C]// Proceedings of the 27th ACM International Conference on Multimedia, 2019: 465-472
[31] Zhu Y B, Li C L, Lu Y, et al. FANet: quality-aware feature aggregation network for RGB-T tracking[J]. arXiv Preprint, 2018, arXiv: 1811.09855
[32] Li C L, Wu X H, Zhao N, et al. Fusing two-stream convolutional neural networks for RGB-T object tracking[J]. Neurocomputing, 2018, 281: 78-85
[33] Li C L, Lu A D, Zheng A H, et al. Multi-adapter RGBT tracking[J]. arXiv Preprint, 2019, arXiv: 1907.07485
[34] Yang R, Zhu Y B, Wang X, et al. Learning target-oriented dual attention for robust RGB-T tracking[C]// 2019 IEEE International Conference on Image Processing (ICIP), 2019, DOI: 10.1109/ICIP.2019.8803528
[35] 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
[36] Kristan M, Matas J, Leonardis A, et al. The seventh visual object tracking VOT2019 challenge results[C]// Proceedings of the IEEE International Conference on Computer Vision Workshop, 2019

A survey of RGB-T object tracking

DING Zhengtong[1] XU Lei[1] ZHANG Yan[1] LI Piaoyang[1] LI Yangyang[1] LUO Bin[1] TU Zhengzheng[1]
1 School of Computer Science and Technology, Anhui University, Hefei 230601

Abstract RGB-Thermal object tracking has developed due to its strongly complementary benefits of thermal information to visible data. In this paper, we introduce the research background of RGB-T object tracking and the challenges in this task; then summarize and introduce the existing methods of RGB-T object tracking, including traditional methods and deep learning methods. Finally, we analyze and compare the existing RGB-T datasets and evaluation criteria, and point out the aspects worthy of study in RGB-T object tracking.

Key words RGB-Thermal; multimodality; object tracking