

多模型协作的分块目标跟踪*

刘明华¹, 汪传生², 胡强¹, 王传旭¹, 崔雪红¹

¹(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

²(青岛科技大学 机电学院, 山东 青岛 266061)

通讯作者: 汪传生, E-mail: wcsmta@qust.edu.cn



摘要: 为了解决复杂场景下,基于整体表观模型的目标跟踪算法容易丢失目标的问题,提出了一种多模型协作的分块目标跟踪算法.融合基于局部敏感直方图的产生式模型和基于超像素分割的判别式模型构建目标表观模型,提取局部敏感直方图的亮度不变特征来抵抗光照变化的影响;引入目标模型的自适应分块划分策略以解决局部敏感直方图算法缺少有效遮挡处理机制的问题,提高目标的抗遮挡性;通过相对熵和均值聚类度量量子块的局部差异置信度和目标背景置信度,建立双权重约束机制和子块异步更新策略,在粒子滤波框架下,选择置信度高的子块定位目标.实验结果表明,该方法在复杂场景下具有良好的跟踪精度和稳定性.

关键词: 协作模型;局部敏感直方图;粒子滤波;分块目标跟踪;超像素

中图法分类号: TP391

中文引用格式: 刘明华,汪传生,胡强,王传旭,崔雪红.多模型协作的分块目标跟踪.软件学报,2020,31(2):511-530. <http://www.jos.org.cn/1000-9825/5656.htm>

英文引用格式: Liu MH, Wang CS, Hu Q, Wang CX, Cui XH. Part-based object tracking based on multi collaborative model. Ruan Jian Xue Bao/Journal of Software, 2020,31(2):511-530 (in Chinese). <http://www.jos.org.cn/1000-9825/5656.htm>

Part-based Object Tracking Based on Multi Collaborative Model

LIU Ming-Hua¹, WANG Chuan-Sheng², HU Qiang¹, WANG Chuan-Xu¹, CUI Xue-Hong¹

¹(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

²(College of Mechanical and Electrical Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: A part-based tracking approach based on multi collaborative model is proposed that can address the problem of losing object based on the holistic appearance model in complex scenarios. Object appearance model is constructed by fusing the generative model based on local sensitive histogram (LSH) and discriminative model based on superpixel segmentation, by extracting the illumination invariant feature of the LSH resist the influence of the illumination changes on the object model effectively; for the lack of effective occlusion handling mechanism of the LSH algorithm, the part-based adaptive model segmentation method is introduced to improve the performance of resistance occlusion; by through the relative entropy and mean shift cluster method, measuring the differences confidence value and the foreground-background confidence value of the local part, establish the dual weights constraint mechanism and asynchronous update strategy for the part model, the partes with high confidence are selected to locate object in the particle filter framework. Experimental results on challenging sequences confirm that the proposed approach outperforms the related tracking algorithm in complex scenarios.

Key words: collaborative model; local sensitive histogram; particle filter; part-based object tracking; superpixel

* 基金项目: 国家自然科学基金(61472196, 61672305); 山东省重点研发项目(2017GGX10133)

Foundation item: National Natural Science Foundation of China (61472196, 61672305); Key Research and Development Program of Shandong Province (2017GGX10133)

收稿时间: 2017-05-22; 修改时间: 2018-03-12, 2018-05-28; 采用时间: 2018-08-21; jos 在线出版时间: 2019-03-27

CNKI 网络优先出版: 2019-03-27 16:40:15, <http://kns.cnki.net/kcms/detail/11.2560.TP.20190327.1640.002.html>

视觉目标跟踪(visual object tracking,简称 VOT)是机器视觉领域的研究热点之一,在智能视频监控、人机交互和机器人导航定位等领域有着重要的应用.视觉目标跟踪一般包括目标初始化、表观建模、运动估计和目标定位,而表观建模是关键.复杂环境下,由于光照变化、相似目标干扰以及遮挡等因素的影响,导致目标表观模型产生较大变化,使得对运动目标的跟踪变得非常困难.因此,选择有效的视觉特征建立目标模型,以及提高目标模型对复杂背景的可分性和自适应性,是实现鲁棒跟踪的关键^[1].根据目标模型构建方式的不同,跟踪算法一般可分为产生式算法^[2-7]和判别式算法^[8-14]两种.

产生式算法是一种基于模式匹配的处理方法.该方法首先经过若干帧的在线学习构建目标的先验特征模型,然后通过最大似然或最大后验概率估计,在后续帧中搜索与目标模板最匹配的区域,以确定目标位置.产生式算法的模型描述方式主要包括子空间学习^[2,3]、特征模板^[4,5]和稀疏表示^[6,7]这3种.例如,Ross等人^[2]提出了增量视觉跟踪器(incremental visual tracking,简称 IVT)算法,利用低维子空间学习对目标表观建模,通过增量学习子空间在线更新目标模型,并根据候选目标与特征子空间中心距离的大小来定位目标.该方法较好地解决了目标表观变化的问题.Kwon等人^[3]提出了视觉跟踪分解(visual tracking decomposition,简称 VTD)算法.该算法在粒子滤波框架下,将运动模型和观测模型分解成多个子模型来描述较大程度的目标表观变化.He等人^[4]提出一种基于局部敏感直方图(local sensitive histogram,简称 LSH)特征的目标模型描述方法.该方法充分利用每个像素点和周边像素的位置关系,计算其加权直方图特征,并基于该特征实现了一个多区域目标跟踪算法.该算法在光照变化、目标姿态变化等条件下均能稳定地跟踪目标.Jia等人^[6]提出一种基于结构化局部稀疏表观模型的跟踪算法,充分利用局部图像块的稀疏编码和空间结构信息,结合增量子空间学习的自适应模板更新策略,较好地解决了遮挡和相似目标的干扰问题.产生式算法主要着眼于对目标自身的刻画,提取的目标信息较为丰富,具有很好的目标模型描述能力,但忽略了对背景信息的利用,在目标自身产生较大变化或者被遮挡时容易产生跟踪漂移.

判别式算法是一种基于目标检测的处理方法.该方法将目标跟踪视为目标和背景的二分类问题.其基本思想是,构建能够区分目标和背景的判别模型,通过训练和更新在线分类器对候选区域的目标和背景进行准确分类来确定目标位置.判别式算法包括传统判别式算法、基于相关滤波和深度学习的目标跟踪算法.传统判别式算法根据选择的分类器不同,主要分为基于 boosting 分类器^[8,9]、支持向量机分类器^[10,11]、贝叶斯分类器^[12]和随机学习分类器^[13]的目标跟踪算法.例如,Babenko等人^[8]提出一种基于样本多样性的多示例学习(multiple instance learning,简称 MIL)算法,根据跟踪结果,利用正负样本训练分类器,构建目标表观模型并进行在线更新,以降低目标模板中包含背景信息的风险.Kalal等人^[9]提出一种基于结构约束的跟踪学习检测(tracking-learning-detection,简称 TLD)算法,将目标检测、在线学习和目标跟踪这3个模块有机结合,通过在线学习策略不断更新跟踪模块的显著特征点和检测模块的目标模型及相关参数,保证跟踪的稳定性.Horst等人^[12]提出一种干扰感知目标跟踪方法(distractor-aware tracking,简称 DAT),将预先探测出的干扰区域与正确的目标区域相结合,构建判别式干扰感知模型,通过密集采样获得候选目标,并利用目标概率图的概率值与距离值加权定位目标.传统判别式算法因同时考虑到目标和背景信息,在跟踪过程中表现出较高的性能,但复杂环境下,传统判别式算法需要大量高质量的样本训练分类器来保证跟踪精度,计算复杂度很高,很难满足跟踪实时性的要求.

相关滤波跟踪方法^[15-26]的出现,有效地解决了传统判别式算法样本贫乏和计算效率低的问题.相关滤波跟踪算法利用误差最小化模型构建目标函数,通过训练目标背景分类器,得到一个判别式相关滤波模型,对后续帧所有待检测样本,利用该模型测试分类器的响应输出,通过寻找响应峰值来预测目标位置.例如,Bolme等人^[15]首次将相关滤波思想应用到目标跟踪领域,提出了以最小均方差为目标的相关滤波跟踪算法.该方法通过快速傅里叶变换将时域的卷积运算转化为频域的点积运算,简化了相关操作的计算过程,提高了跟踪速度.Henriques等人^[16]提出一种基于循环矩阵的密集采样策略(circulant structure of tracking with kernels,简称 CSK)来解决训练样本不足的问题.算法具有很高的跟踪速度和较高的跟踪精度,但未考虑目标尺度变化,并且仅使用了灰度特征(raw pixel),在复杂环境下容易丢失目标.针对 CSK 算法存在的问题,Henriques等人^[17]提出了核相关滤波(kernelized correlation filter,简称 KCF)目标跟踪算法,将 CSK 算法扩展成多通道,并引入了梯度方向直方图

(histogram of gradient,简称 HOG)特征,提高了跟踪精度和鲁棒性.但 KCF 算法仍使用固定尺度的跟踪窗,在目标模型更新时也没有考虑目标的遮挡.Danelljan 等人^[18]在相关滤波框架中引入了颜色名属性(color name,简称 CN)特征来扩展 CSK 跟踪器,并通过自适应维度下降法避免了高维颜色属性特征的计算,保证了跟踪的实时性.其他类似研究可参见文献[19–26],这些研究主要解决目标的尺度变化问题和复杂环境下高质量样本的选取机制问题.基于相关滤波的跟踪算法将所有的计算过程都转换到频域空间,通过快速傅里叶变换,极大地提升了运算速度,但多数算法依然选择灰度、颜色和 HOG 等特征构建目标整体表观模型,忽视了局部背景信息的影响,在目标被遮挡、存在相似目标或复杂背景干扰时,训练样本中很容易引入背景像素,导致训练模型的判别性不足,分类能力退化,在分类器的响应输出中呈现多峰现象,产生跟踪漂移.

近年来,基于深度学习的目标跟踪方法在 VOT^[27]竞赛中表现出了巨大的潜力,深度学习由于强大的特征提取和表达能力,得到了研究者的重视.目前,基于深度学习的目标跟踪方法主要有两种研究思路:一是利用深度网络学习到的特征的可迁移性,先在大规模训练样本上离线训练某一特定类型的深度神经网络,使其在具体的在线跟踪时可以对目标特征进行提取,并利用获取的在线数据对该深度网络进行微调,以适应目标表观变化^[28–32];另一种研究思路是通过改变深度神经网络的结构(如控制网络层数和每层神经元的数量),使其能够以完全在线的方式利用卷积神经网络进行目标跟踪^[33,34].基于深度学习的目标跟踪算法面临的主要问题是,(1) 深度特征模型的构建依赖于对大量标注样本的有效学习,而视觉目标跟踪是一个典型的小样本在线选择和学习问题,一方面能够获取的在线训练样本数量非常有限,另一方面获取的训练样本通常是基于对目标整体表观模型的学习,这些因素都会导致学习到的深度模型判别性不足;(2) 构建规模庞大的深度学习网络计算复杂度非常高,很难满足跟踪的实时性要求,虽然部分算法跟踪效果很好,但是距离实际应用依然有很大的距离.

复杂环境下,目标跟踪容易受到光照变化和遮挡而导致跟踪失败,基于分块的目标跟踪算法将目标整体划分为若干空间结构化的子块集合,通过度量不同子块的匹配程度,实现对整个目标的跟踪.由于在子块内部图像模式的不变性较强,算法对目标局部变化和遮挡具有较强的适应性而得到研究者的广泛关注^[5,24,35–43],例如,Adam 等人^[35]将目标模型分割成多个均匀的矩形分块,提取各子块的直方图特征,在候选目标区域进行搜索匹配,通过所有子块加权投票确定目标位置;Li 等人^[36]在粒子滤波框架下,对目标子块模型引入可靠性度量机制,通过有效的模板更新和尺度估计实现目标的精确跟踪;Yang 等人^[37]针对高层和低层信息对目标模型描述的不足,提出一种基于中层信息的超像素目标模型,通过 Mean Shift 聚类构建由前景和背景置信图描述的判别式目标模型,以提高算法在目标遮挡和形变时跟踪的鲁棒性.为了提高分块跟踪的计算效率,有研究者将分块策略和相关滤波框架相结合,例如,Liu 等人^[24]提出一种基于自适应相关滤波的分块目标跟踪方法,该方法将目标分割成固定数量的矩形分块,对每个目标分块分别利用相关滤波分类器计算其滤波响应值,通过对不同分块的预测结果加权融合定位目标.Liu 等人^[38]根据目标模型的高宽比,将目标划分为固定数量的矩形分块,通过独立训练每一个分块,得到联合的相关滤波分类器,在定位目标时,利用学习到各分块之间的空间结构关系来进一步提高跟踪精度.类似研究可参见文献[39–43].上述基于分块的目标跟踪算法通常采用固定数量和大小的均匀矩形分块对目标模型进行分割,未考虑子块的局部差异性和多样性,当目标模型特征分布均匀时,各子块之间容易产生信息冗余,当目标模型特征分布复杂时,又无法适应目标模型的多变性,复杂环境下,跟踪精度不高.

综上所述,产生式模型关注数据产生的过程,易于增量更新;判别式模型关注后验概率估计,提取的信息具有判别性.前者强调目标模型的完整性和不变性,后者强调目标模型的可分性.鉴于两者的互补性,本文分别构建了基于局部敏感直方图的产生式模型和超像素分割的判别式模型,并将二者融合对目标模型进行描述,提出一种多模型协作的自适应分块目标跟踪算法,以提高光照变化及遮挡等复杂条件下跟踪的鲁棒性.

本文的主要创新性研究成果如下:

- (1) 利用超像素模型构建自适应目标子块,通过相对熵和均值聚类度量各子块的局部差异置信值和目标背景置信值,建立双权值联合约束的目标子块置信度量机制,通过联合高置信度目标子块进行跟踪,有效地降低了相似目标和背景的干扰.
- (2) 提取目标子块的局部敏感直方图特征,给出子块相似性度量方法,在粒子滤波框架下,根据相似性度

量结果计算各分块权重,联合权重较大的候选目标块定位目标,避免了全局搜索匹配,保证算法具有一定的执行效率.

- (3) 对不同的目标子块采取异步更新策略,通过度量目标子块的权重变化,及时确定目标模型何时需要更新以及哪些子块需要更新,保证更新后的目标模型几乎不包含背景信息,避免了基于全局特征模型的更新方式容易引入背景像素导致的跟踪漂移,更好地适应目标和场景的多变性.

实验结果表明,本文算法能够很好地适应目标表现变化和遮挡影响,有效地提高复杂环境下跟踪的精确性和鲁棒性.

1 局部敏感直方图特征

1.1 局部敏感直方图

局部敏感直方图(local sensitive histogram,简称LSH)是对图像中某一像素点周围区域的亮度值进行的加权统计,将周围区域某个灰度级的所有像素点的亮度值乘以一个浮点权值后累加,作为相应位置的像素值.设 $H_p(b)$ 表示像素点 p 的局部敏感直方图第 b 个灰度级的值,则计算公式如下.

$$H_p(b) = \sum_{q=1}^W \alpha^{|p-q|} Q(I_q, b), b=1, \dots, B \quad (1)$$

其中,

- $\alpha \in (0,1)$ 是像素位置权重控制参数,使得远离中心像素点的像素权值较低.
- I_q 为像素点 q 的像素值,若 $I_q \in b$,则 $Q(I_q, b)=1$;否则, $Q(I_q, b)=0$.
- W 为目标区域总像素数.
- B 为目标区域灰度级数.

公式(1)的含义是,像素点 p 的局部敏感直方图中,第 b 个灰度级分量的值是所有灰度级为 b 的像素点按照距离 p 的远近乘以位置权重 $\alpha^{|p-q|}$ 累加得到.

局部敏感直方图为每个像素都赋予了相应权值,除考虑到像素的统计特性外,还充分考虑到了像素的空间位置信息,利用该特征描述目标模型具有非常强的鲁棒性.

1.2 光照不变特征

为了克服光照变化的影响,利用局部敏感直方图,可建立光照不变特征如下:

$$\Gamma_p = \sum_{b=1}^B \exp\left(-\frac{(b-b_p)^2}{2\max(\kappa, \kappa I_p)}\right) \cdot H_p(b) \quad (2)$$

其中, b_p 为 p 点的灰度级; κ 为常量,其值取0.1; I_p 为目标区域的平均亮度; B 为目标区域灰度级数; $\exp(\cdot)$ 为 $H_p(b)$ 的权重.实验结果表明,当光照剧烈变化时,不考虑量化误差的情况,公式(2)的计算结果基本不变.因此,选择特征 Γ_p 描述目标,可以有效地避免光照变化对跟踪鲁棒性的影响.图1给出了不同光照条件下的光照不变特征图像.其中,图1(a)和图1(b)分别为正常光照和较暗光照条件下的图像,图1(c)和图1(d)为对应的经过光照不变特征变换之后的图像.对比图1(c)和图1(d)可知,公式(2)给出的特征具有良好的光照不变性.

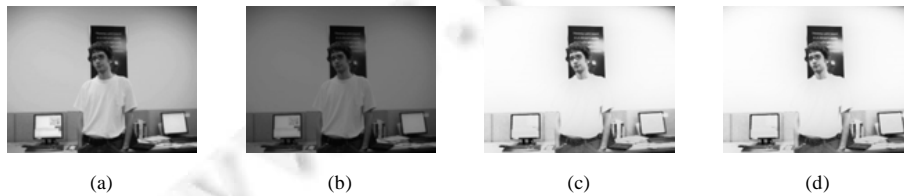


Fig.1 Illumination invariant feature images under different illumination conditions

图1 不同光照条件下的亮度不变特征图像

2 基于多模型协作的自适应分块目标跟踪算法

2.1 基于超像素模型的目标分块

固定大小和数量的目标分块虽然简单,却容易破坏目标区域内部结构及目标与背景的边界,无法适应目标表现的多变性,给后续跟踪带来困难.超像素能够将颜色、亮度、纹理等一些特征属性相似的像素点聚合在一起,作为一个区域整体进行处理,同一超像素内的特征信息比较统一,利用超像素模型能够得到自适应的目标子块,并且得到的目标子块能够很好地保持目标边界和空间结构特征,保证得到的目标子块是具有一定语义的独立分块.相对于固定数量和形状的分块,超像素分块的模式不变性更好,更适合于目标跟踪.鉴于此,本文算法首先利用超像素模型将目标区域自适应划分为大小不同的若干目标子块,然后在特征空间中度量每个目标子块的权重,根据权值的大小,自适应选择适当的目标分块进行跟踪.基于超像素分割产生目标分块的过程如图 2 所示.图 2(a)中红色框为选定目标边界框,图 2(b)为超像素分割的结果,图 2(c)为目标区域对应的超像素分割结果,图 2(d)为通过超像素分割获取的目标分块.由于超像素分割不会破坏图像的边界信息,而目标框并非图像的自然边界,为消除基于超像素分割的目标分块不确定性,本文将与目标框交叉的超像素均判定为背景,此划分策略能够剔除目标框内的所有背景像素,保证学习到的目标表现模型具有很强的判别力.实验结果表明,上述目标分块划分策略在抗遮挡能力和跟踪效率方面都有较好的效果.

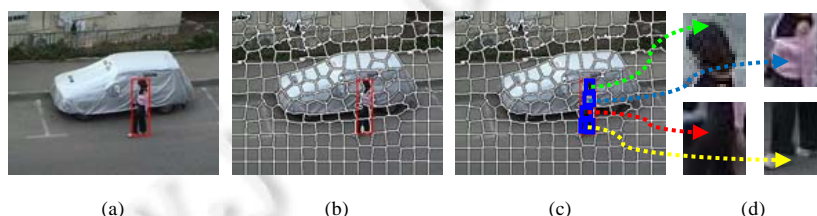


Fig.2 Flow of target fragment based on superpixel model

图 2 基于超像素模型的目标分块流程

2.2 基于分块的目标跟踪

设 t 时刻包含目标的图像帧为 I_t , 目标模板的 N 个子块为 $\{P_t^{(i)} | P_t^{(i)} = s_t(d_y^{(i)}, w_0, h_0)\}_{i=1}^N$, 其中, $d_y^{(i)}$ 为第 i 个子块中心相对目标中心的偏移量, w_0 和 h_0 为子块的初始宽和高, s_t 为 t 时刻目标模板的尺度因子, 初始化 $s_t=1$, 定义 $S_c(P_t^{(i)})$ 表示 t 时刻第 i 个候选目标子块的面积, $S_o(P_{t-1}^{(i)})$ 表示 $t-1$ 时刻第 i 个目标子块的面积, M 是根据子块置信度的度量结果选择的目标子块数量, 则尺度因子为

$$s_t = \frac{\sum_{i=1}^M S_c(P_t^{(i)})}{\sum_{i=1}^M S_o(P_{t-1}^{(i)})} \quad (3)$$

定义每个子块的置信度 $r_t = \{r_t^{(i)}\}_{i=1}^N$, 其中, $r_t^{(i)} \in [0, 1]$, $\sum_{i=1}^N r_t^{(i)} = 1$, 其值越大, 可靠性越高. 利用目标子块估计当前目标状态 X_t , 有:

$$p(X_t | I_t) \propto \sum_{i=1}^N p(X_t | O(P_t^{(i)})) p(O(P_t^{(i)}) | I_t) = \sum_{i=1}^N r_t^{(i)} p(X_t | O(P_t^{(i)})) \quad (4)$$

其中, $O(P_t^{(i)})$ 映射为第 i 个子块的中心.

利用局部敏感直方图建立子块特征模型, 为了避免穷尽搜索, 利用粒子滤波对每个子块构建跟踪器, 根据运动模型 $p(y_t | y_{t-1})$ 和当前观测值预测每个子块的位置:

$$\hat{O}(P_t^{(i)}) = \arg \max_{O(P_t^{(i)})} p(I_t, O(P_t^{(i)}) | \hat{O}(P_{t-1}^{(i)})) \quad (5)$$

通常,目标和其各子块的相对位置在短时间内固定,则由各子块对目标当前位置估计为

$$X_t = X_{t-1} + \sum_{i=1}^N r_i^{(i)} (\hat{O}(P_i^{(i)}) - s_i d_y^{(i)}) \quad (6)$$

2.3 子块置信度计算

(1) 局部差异置信度

利用子块进行目标跟踪时,选择受周围图像影响较小、与周边图像特征差异较大的子块对目标进行定位,结果将更为准确.为了描述目标子块受周边图像干扰程度的大小,定义局部差异性 $d_t(r)$ 为第 t 帧第 r 个目标子块与其周边图像的特征差异程度,并利用 Kullback-Leibler 相对熵来进行度量:

$$d_t(r) = K_L(h_f, h_b) = \sum_i h_f(i) \log \frac{h_f(i)}{h_b(i)} \quad (7)$$

其中, $h_f(\cdot)$ 和 $h_b(\cdot)$ 分别为目标子块直方图和周边区域直方图.对于子块 $P_i^{(i)} = (d_y^{(i)}, w, h)$ 的周边区域采样范围为 $B_i^{(i)} = (d_y^{(i)}, \lambda_s w, \lambda_s h)$, 其中, $d_y^{(i)}$ 为采样中心; λ_s 为采样范围控制参数,本文取 $\lambda_s = 1.5$.

图 3 给出了不同子块的局部差异置信度变化曲线.

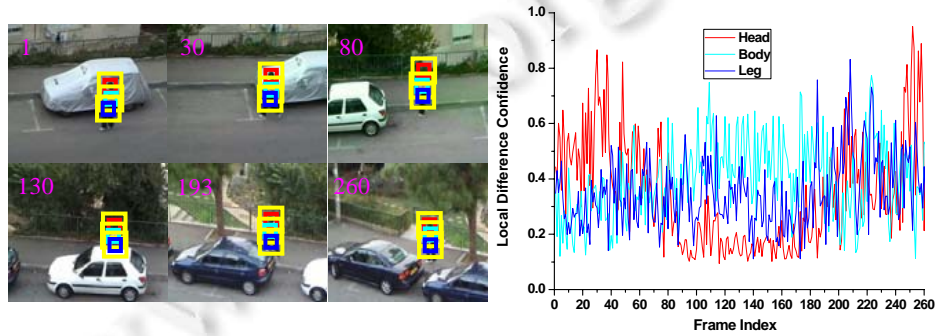


Fig.3 Local difference confidence change curve of different parts

图 3 不同子块的局部差异置信度变化曲线

(2) 目标背景置信度

首先将目标和局部背景区域分割为 m 个超像素,设 $s(t,r)$ 代表第 t 帧中第 r 个超像素, $s(t,r)$ 对应的特征向量为 f_t^r , 这里提取的特征为局部敏感直方图特征, $F = \{f_t^r | t = 1, \dots, N; r = 1, \dots, m\}$ 为前 N 帧中所有超像素特征向量构成的特征池,对集合 F ,在特征空间中利用均值迁移(mean-shift)聚类方法得到 n 个不同的聚类,每个聚类 $clst(i) (i=1, \dots, n)$ 包含聚类中心 $f_c(i)$ 、聚类半径 $r_c(i)$ 及聚类成员 $\{f_t^r | f_t^r \in clst(i)\}$.

每个聚类 $clst(i)$ 都对应训练图像中的不同区域,根据这些区域隶属于目标或背景的大小,可以计算聚类 $clst(i)$ 的目标背景置信度.假设 $s_c^+(i)$ 表示聚类 $clst(i)$ 包含的所有超像素覆盖区域和训练帧中目标重合区域的大小, $s_c^-(i)$ 表示聚类 $clst(i)$ 包含的所有超像素覆盖区域和训练帧中背景重合区域的大小,令 $r_s = s_c^+(i) / s_c^-(i)$, 显然, r_s 越大,表明 $clst(i)$ 的聚类成员属于目标区域的概率越高.给每个聚类赋予一个 $(0,1)$ 之间的目标背景置信度 $p_c(i)$, 表示聚类成员属于目标或背景的概率,其值定义如下:

$$p_c(i) = \frac{r_s}{r_s + 1} \quad (8)$$

对新到来的每一帧,进行超像素分割,利用已经训练好的数据集,在聚类 $clst(i)$ 置信度 $p_c(i)$ 的基础上,通过聚类中心和半径,度量每个超像素的目标背景置信度 $p_c(r)$ 如下:

$$p_c(r) = w(r,i) \times p_c(i) \quad (9)$$

其中,权值 $w(r,i)$ 由第 t 帧第 r 个超像素对应特征向量 f_t^r 与超像素对应聚类中心间的距离计算得到,定义如下:

$$w(r, i) = \exp\left(-\lambda_d \times \frac{\|f_i^r - f_c(i)\|_2}{r_c(i)}\right) \quad (10)$$

其中, λ_d 是归一化参数, 本文取 $\lambda_d=2$. 由公式(10)可知, 超像素 $s(t, r)$ 离聚类中心越远, 其所属聚类的目标背景置信度 $p_c(i)$ 对 $s(t, r)$ 的影响越小. 根据得到的目标背景置信度的大小, 确定目标超像素. 假设与超像素 $s(t, r)$ 对应的目标子块为 $P_t^{(r)}$, 则利用超像素 $s(t, r)$ 的目标背景置信度 $p_c(r)$ 可定义第 t 帧第 r 个目标子块的目标背景置信度 $p_t(r)=p_c(r)$. 图 4 给出了不同子块基于超像素的目标背景置信度变化曲线图.

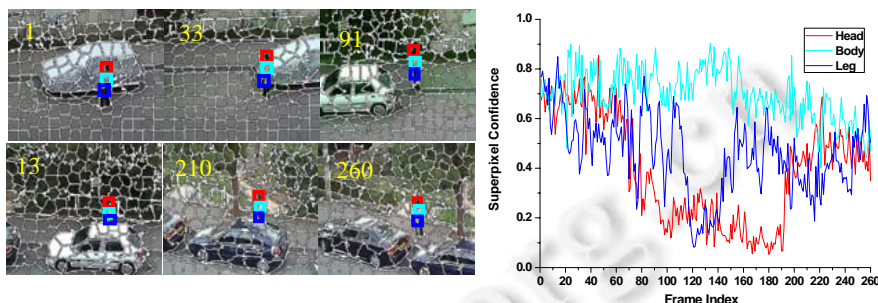


Fig.4 Target-background confidence change curve of different parts

图 4 不同子块的目标背景置信度变化曲线

综合子块的局部差异置信度和目标背景置信度可得:

$$r_t^{(i)} = [w_d, w_s][d_t(i), p_t(i)]^T \quad (11)$$

其中, $r_t^{(i)}$ 为第 t 帧第 i 个目标子块的置信度; w_d 和 w_s 为设定的归一化置信度参数, 控制局部差异置信度和目标背景置信度对当前目标子块置信度的贡献程度, 根据前 k 帧场景的变化动态调整, $w_d \propto \frac{1}{k} \sum_{i=1}^k d_t(i)$, $w_s \propto \frac{1}{k} \sum_{i=1}^k p_t(i)$, 初始化 $w_d=1, w_s=0$. 根据 $r_t^{(i)}$ 的大小, 从目标子块中选择 N 个置信度最高的子块进行目标定位. 为了避免置信度较低子块参与目标定位, 导致跟踪精度降低, 设置可靠性阈值 θ_r , 有:

$$N = \arg \max_i (r_t^{(i)} > \theta_r) \quad (12)$$

θ_r 根据场景和目标模型的变化动态选择, 若 $N=0$, 则根据上一帧的目标位置, 重新进行超像素分割, 训练超像素模型, 重新初始化目标分块.

2.4 基于局部敏感直方图的相似性度量

假设第 t 帧和目标子块 $P_t^{(i)}$ 对应的中心在 (x, y) 处的候选目标子块为 $Q_t^{(i)}(x, y)$, 目标子块和候选目标子块均由 K 个子区域构成, K 值的大小可根据子块大小的不同来动态调整, 本文固定取 $K=100$. p_j 为目标子块中的第 j 个子区域, $q_{(x, y, j)}$ 为对应的候选目标子块的第 j 个子区域, p_j 中的每个像素由公式(1)计算的结果为 $H_p^{(j)}$, $q_{(x, y, j)}$ 中的每个像素由公式(1)计算的结果为 $H_q^{(j)}$, $C_p^{(j)}$ 为 p_j 中所有像素 $H_p^{(j)}$ 的累加值, 即 $C_p^{(j)}(b) = \sum H_p^{(j)}(b)$, $C_q^{(j)}$ 为 $q_{(x, y, j)}$ 中所有像素 $H_q^{(j)}$ 的累加值, 即 $C_q^{(j)}(b) = \sum H_q^{(j)}(b)$. 定义 $d(p_j, q_{(x, y, j)})$ 为子区域 p_j 和 $q_{(x, y, j)}$ 的差值, 其计算方式如下:

$$d(p_j, q_{(x, y, j)}) = \sum_{b=1}^B \|C_p^{(j)}(b) - C_q^{(j)}(b)\| \quad (13)$$

对目标子块和候选目标子块的所有子区域差值进行累加, 定义 $d(P_t^{(i)}, Q_t^{(i)})$ 为第 t 帧第 i 个目标子块和候选目标子块总的相似性度量, 有:

$$d(P_t^{(i)}, Q_t^{(i)}(x, y)) = \sum_{j=1}^K d(p_j, q_{(x, y, j)}) \quad (14)$$

对目标子块 $P_t^{(i)}$ 的每个候选目标子块 $Q_t^{(i)}(x, y)$ 均按上述方法进行计算, 其和最小的位置即为当前帧目标子

块的位置,记为 $\Psi^{(i)}(x,y)$,则有:

$$\Psi^{(i)}(x,y) = \arg \min_{(x,y)} d(P_t^{(i)}, Q_t^{(i)}(x,y)) \quad (15)$$

在粒子滤波框架内,根据目标子块特征模型和观测值 $Y_{1:t}$ 的一致性,利用上述相似性度量方法计算各子块的似然函数:

$$p_{p_t^{(i)}}(Y_t | X_t^{(i)}) = \exp\left(-\frac{d(P_t^{(i)}, I_{p_t^{(i)}}^{(i)}(x,y))^2}{2\sigma^2}\right) \quad (16)$$

其中, $l=1, \dots, M$ 为每个子块的采样粒子数; $I_{p_t^{(i)}}^{(i)}(x,y)$ 为粒子中心在 (x,y) 处,和目标子块 $P_t^{(i)}$ 对应的候选目标子块; $d(\cdot, \cdot)$ 代表基于局部敏感直方图特征的目标子块相似性度量。

获取单个子块的似然函数后,根据不同子块的置信度,计算所有子块的联合似然函数:

$$p(Y_t | X_t^{(i)}) = \exp\left(-\frac{\sum_{i=1}^N r_t^{(i)} d(P_t^{(i)}, I_{p_t^{(i)}}^{(i)}(x,y))^2}{2\sigma^2}\right) \quad (17)$$

选择高斯运动模型 $p(Y_t | Y_{t-1}) = \mathcal{N}(Y_t; Y_{t-1}, A)$, 其中, A 为对角协方差矩阵,其元素值为目标位置和尺度的标准偏差,结合公式(17)可求得公式(6)给出的目标位置:

$$\hat{X}_t = \arg \max_{X_t^{(i)}} p(Y_t | X_t^{(i)}) \quad (18)$$

2.5 遮挡处理和目标模型更新

复杂场景下,跟踪目标很容易出现遮挡情况,如果不加区分的进行目标模型的更新,很容易造成跟踪漂移,因此有必要对遮挡进行检测处理,并据此对目标模型采用不同的更新策略.本文从目标相似性度量和超像素判别两个方面对每个目标子块进行遮挡处理.

(1) 遮挡检测

对每个目标子块 $P_t^{(i)}$,首先通过度量与其对应的超像素目标背景置信度的变化进行判断,定义如下:

$$\frac{\|p_t(i) - \mu_H\|}{\mu_H} > \lambda_0 \quad (19)$$

其中, $p_t(i)$ 为第 t 帧第 i 个目标子块的目标背景置信度, μ_H 为过去 H 帧目标子块的目标背景置信度的均值, λ_0 为基于超像素的遮挡判定阈值.本文初始化 $\lambda_0=0.3$,并根据场景变化动态调整.公式(19)反映了当前目标子块包含目标和背景像素的比例关系和过去 H 帧平均值的变化程度,这种变化程度的大小可作为遮挡严重程度判定的依据,如果公式(19)成立,则表示当前目标子块 $P_t^{(i)}$ 可能发生了遮挡.由于跟踪过程中目标自身形变,光照变化等因素均会对遮挡的判定产生干扰,此时可由公式(16)得到的似然函数值的变化作为子块是否被遮挡进一步判定的依据.由于子块 $P_t^{(i)}$ 的似然函数值取决于该子块和对应候选目标子块的相似性度量,因此,可利用相似性度量值的变化程度作为 $P_t^{(i)}$ 是否被遮挡的判定依据,定义如下:

$$\frac{\|d(p_t^{(i)}) - \eta_K\|}{\eta_K} > \lambda_1 \quad (20)$$

其中, $d(p_t^{(i)})$ 为第 t 帧第 i 个目标子块 $P_t^{(i)}$ 的相似性度量值, η_K 为过去 K 帧该目标子块相似性度量的均值, λ_1 为基于相似性度量的遮挡判定阈值.本文初始化 $\lambda_1=0.3$,并根据场景变化动态调整.图 5 给出了基于子块置信度变化的遮挡检测过程.由图中黑色曲线可知,在 111 帧“Leg 子块”产生严重遮挡时,置信度的平均变化率达到最大值(蓝色圆圈所示),在 153 帧“Leg 子块”从遮挡状态恢复到正常状态,此时置信度的平均变化率达到最小值(绿色圆圈所示),从 197 帧开始,“Leg 子块”又进入遮挡状态,此时置信度的平均变化率再次达到峰值点(紫色圆圈所示);从红色曲线可以看出,“Body 子块”的置信度平均变化率比较平缓,并且没有处于指定阈值之外的区域,这说明“Body 子块”受背景区域的干扰较小,没有遮挡情况发生,计算结果和实际跟踪过程基本一致.可见,通过度量子块置信度的平均变化率能够有效地感知目标是否处于遮挡状态.

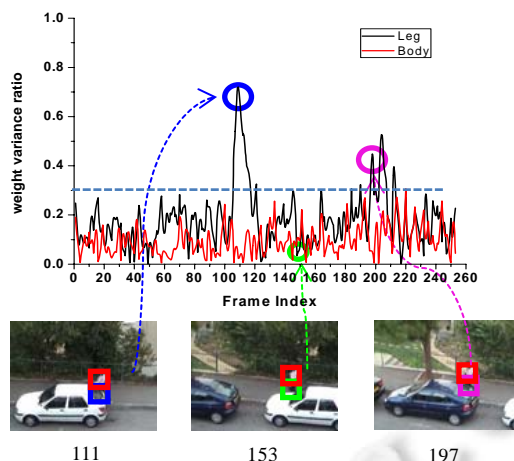


Fig.5 Confidence change rate curve of different parts

图 5 不同子块置信度变化率曲线

(2) 目标模型异步更新

跟踪过程中,由于目标和场景的动态变化以及受光照变化、遮挡等不确定因素的影响,每一帧提取的目标及其背景超像素块在外观、形状以及数量上都可能会发生变化.为了保证跟踪的稳定性,需要对新一帧中经过超像素分割后的目标子块模型进行更新.与基于全局特征模型的整体更新策略很容易引入背景像素不同,本文对不同目标子块采取异步更新的策略.首先,建立超像素目标模型缓冲池 P 来缓存前 N_0 帧(本文固定取 $N_0=10$)的目标子块模型. $P = \{P_t^{(i)} | t=1, \dots, N_0; i=1, \dots, m\}$, 其中, $P_t^{(i)}$ 为 t 时刻的目标子块模型, m 为目标子块的个数.通常情况下,目标模型在跟踪过程中处于两种状态.

- 目标模型的稳定状态:此时的目标模型受自身及环境因素的干扰很小,目标模型整体处于缓慢、渐变的过程中.
- 目标模型的不确定状态:在目标的渐变过程中,目标模型(或其中的一部分)如果突然受到自身形变或是光照变化、遮挡等不确定因素的影响,将会从稳定状态进入到不确定状态,此时目标模型的不同部分可能处于不同状态,有的部分可能依然处于渐变状态,而有的部分可能已被遮挡,处于不可见状态.

利用公式(19)和公式(20)可以对目标状态进行双重感知,能够有效检测出各目标子块的状态变化.例如, t 时刻,如果目标子块 $P_t^{(i)}$ 被遮挡,则其对应的超像素块 $s(t, i)$ 当前所处的位置将会被背景(遮挡物)超像素块所取代,此时,公式(19)和公式(20)的计算结果都有所反映.对处于不确定状态下的目标子块(如遮挡)不进行更新,对稳定状态下的每个目标子块 $P_t^{(i)}$ 按公式(21)进行更新.

$$P_t^{(i)} = (1 - \tau)P_{t-1}^{(i)} + \tau \hat{P}_t^{(i)} \quad (21)$$

其中, $\hat{P}_t^{(i)}$ 为根据各带权子块联合似然函数估计得到的当前目标子块; τ 为更新因子,控制目标子块更新的速率(本文固定取 $\tau=0.01$).这种根据子块置信度的变化对不同目标子块有选择性的异步更新策略,能够有效地防止更新过程中背景信息的引入,避免跟踪模型的漂移.为了保证目标模型从不确定状态重新回到稳定状态时能够恢复跟踪的持续性,更新后的目标子块模型 $P_t^{(i)}$ 将被重新保存到缓冲池 P 中,保证 P 中始终保存着处于稳定状态下的最新目标子块模型(相当于为跟踪序列加了一个大小为 N_0 的滑窗),一旦检测到目标子块进入不确定状态,则立即保存发生突变时目标子块的 t 和 i 值(什么时间、哪个子块),作为该子块模型重新恢复跟踪时的初始化参数,并且在后续跟踪中扩大搜索区域,以便可以重新检测到目标.

2.6 算法流程

本文算法的具体流程如下.

- (1) 初始化: $t=1$.手工选定跟踪目标,对其进行超像素分割,获取目标分块 $\{P_t^{(i)}\}_{i=1}^N$.
- (2) 跟踪过程: $t=2,3,\dots$
 - (a) 子块置信度计算:当 $t \leq 10$ 时(超像素模型训练阶段),对子块 $P_t^{(i)}$,置 $w_d=1, w_s=0$,利用公式(7)度量其局部差异置信度 $d_t(i)$,并另子块置信度 $r_t^{(i)} = d_t(i)$.
 - (b) 目标模型描述:对子块 $P_t^{(i)}$,利用公式(1)计算其局部敏感直方图特征模型.
 - (c) 目标定位:利用公式(16),在粒子滤波框架内计算每个子块 $P_t^{(i)}$ 的观测似然函数,由公式(17)计算所有带权子块的联合似然函数,并根据公式(18)定位目标.
 - (d) 遮挡处理及更新:根据对当前帧目标定位的结果,利用公式(19)、公式(20)进行目标状态判断,对处于稳定状态的目标子块 $P_t^{(i)}$,利用公式(21)进行更新.
 - (e) 超像素聚类:对训练帧($t \leq 10$)利用上一帧定位的目标位置,对目标及其邻域背景区域超像素进行 Mean-shift 聚类,训练超像素模型.
 - (f) 目标背景置信值计算:对新到来的每一帧($t > 0$),使用步骤(e)训练好的超像素模型,利用公式(8)~公式(10)计算子块 $P_t^{(i)}$ 基于超像素分割的目标背景置信值 $p_t(i)$,并利用公式(11)重新计算子块置信度 $r_t^{(i)}$.
 - (g) 计算 $\max(r_t^{(i)}) < \theta$, 是否成立:若成立,则根据目标上一帧的位置进行超像素分割,重置目标,转步骤(a);否则,转步骤(b).

3 实验结果分析

3.1 实验设置

为验证本文算法的优越性,在 2.3GHz CPU、8G 内存的笔记本,VS2015+OPENCV3.1.0 平台下实现了本文算法(为保证对比的公平性,选择的对比算法中对基于 MATLAB 发布的源程序也在该平台下进行了重新实现).利用改进的 SLIC 算法^[44]实现超像素分割,跟踪目标在第 1 帧人工标定,每个子块的初始粒子数设为 100.

选择目标跟踪数据集中常用的 13 个视频序列进行测试^[45],这些序列涵盖了目标跟踪面临的主要挑战,包括遮挡、复杂背景和相似目标干扰、目标快速移动、目标姿态/尺度及光照发生较大变化等.表 1 给出了选择测试序列特性的具体描述.

Table 1 Experimental sequences

表 1 实验序列

| Sequences | Resolution | Length | Object | Description |
|------------|------------|--------|---------|---|
| Basketball | 576×432 | 725 | 34×81 | Occlusion, deformation, clusters |
| Bird2 | 720×400 | 99 | 69×73 | Deformation, clusters, occlusion |
| Bolt1 | 640×360 | 350 | 26×61 | Deformation, clusters, fast motion |
| Lemming | 640×480 | 1 336 | 61×103 | Occlusion, fast motion, scale variation |
| Faceocc1 | 352×288 | 892 | 114×162 | Occlusion |
| Woman | 352×288 | 597 | 21×95 | Occlusion, illumination variation |
| Tiger2 | 640×480 | 365 | 68×78 | Occlusion, clusters, fast motion |
| Car2 | 320×240 | 913 | 64×52 | Fast motion, illumination variation |
| David1 | 320×240 | 770 | 75×95 | Deformation, illumination variation |
| Diving | 400×224 | 231 | 21×129 | Deformation, clusters, fast motion |
| Jogging | 352×288 | 307 | 25×101 | Occlusion, fast motion |
| Skater | 320×240 | 160 | 39×137 | Deformation, fast motion |
| Shaking | 624×352 | 365 | 61×71 | Clusters, illumination variation |

选择目前主流跟踪算法中综合表现出色的 10 种跟踪算法(KCF^[17]、DSST^[19]、Frag^[35]、LSH^[4]、CN^[18]、DAT^[12]、Staple^[21]、SPT^[37]、STC^[46]、BACF^[23])和本文算法进行定性、定量和计算效率这 3 个方面的分析和

对比.程序参数为作者公布源程序的默认参数.表 2 给出了 10 种跟踪算法的相关信息.

Table 2 Algorithms for comparison
表 2 对比算法

| Algorithm | Year | Publication | Model |
|-----------|------|-------------|----------------|
| Frag | 2006 | CVPR | Generative |
| LSH | 2013 | CVPR | Generative |
| CN | 2014 | CVPR | Discriminative |
| DAT | 2015 | CVPR | Discriminative |
| Staple | 2016 | CVPR | Discriminative |
| KCF | 2015 | PAMI | Discriminative |
| DSST | 2016 | PAMI | Discriminative |
| SPT | 2014 | ITIP | Discriminative |
| STC | 2014 | ECCV | Context |
| BACF | 2017 | ICCV | Discriminative |

3.2 定性分析

本文从遮挡、目标快速移动、光照及目标姿态/尺度变化、相似目标及复杂背景干扰这 4 个角度分析本文算法的有效性,并将本文算法与上述方法进行综合分析对比,图 6(a)~图 6(m)给出了不同测试序列相应帧号的跟踪结果对比.



Fig.6 Comparison of some tracking results on different sequences

图 6 不同序列部分跟踪结果对比

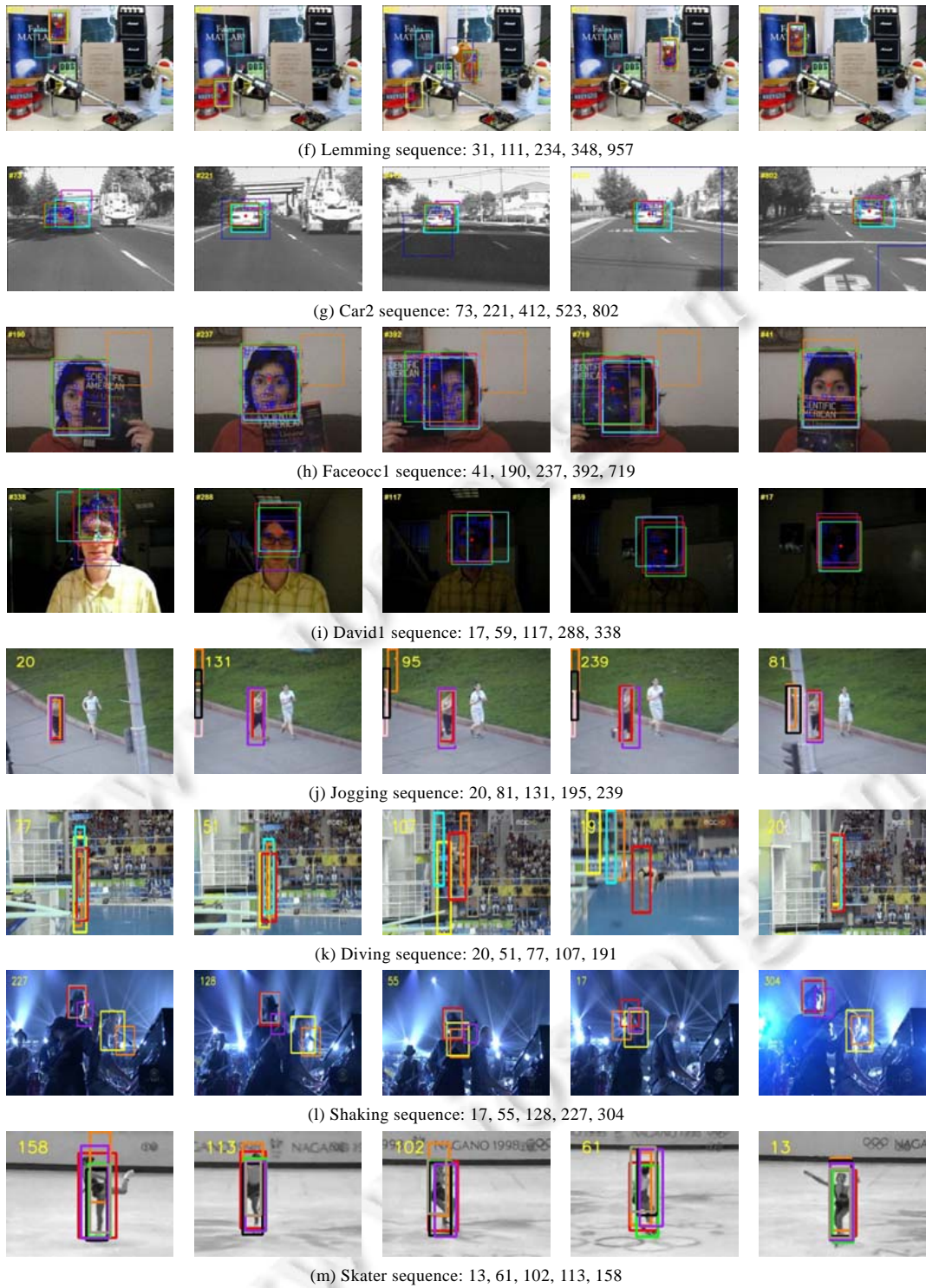


Fig.6 Comparison of some tracking results on different sequences (Continued)

图 6 不同序列部分跟踪结果对比(续)

(1) 遮挡实验

图 6(a)~图 6(j)给出的测试序列均存在不同程度的遮挡,序列(d)展示了行人行走过程中多次被汽车遮挡的跟踪效果.经历第 1 次遮挡时,除 CN 算法丢失目标之外,其他算法都表现出了一定的抗遮挡能力.在经历连续的部分遮挡之后,Staple、DSST 和 BACF 均能稳定跟踪目标,Frag 算法在 217 帧完全丢失目标.除 LSH 之外,其他基于整体特征模型的跟踪算法均出现了较大的跟踪漂移,如 DAT 算法在第 101 帧就几乎丢失目标.KCF 算法由于更新目标模型之前缺少有效的遮挡检测,在经历两次遮挡之后,目标模型出现漂移,并在第 289 帧完全丢失目标.本文算法通过动态选择局部差异性和稳定性最好的目标子块(head 子块+body 子块)定位目标,可有效抵抗连续遮挡的影响,在经历长时间遮挡时,仍能识别出未被遮挡的有效目标子块,准确定位目标.

序列(f)、序列(j)展示了复杂背景下对快速移动目标的跟踪结果,跟踪起始阶段,除 Frag 算法之外,其他算法均能稳定跟踪目标.由于遮挡直接影响基于密集采样策略的相关滤波分类器的性能,在跟踪目标经历了一次完全遮挡后,Staple、DSST 算法均丢失目标;STC 算法只利用第 1 帧样本通过后续迭代更新模型,缺少在线训练过程,经历遮挡后也无法持续稳定跟踪;BACF 算法由于在整帧中采用真实的背景信息作为负样本训练相关滤波模型,相对于在目标区域利用循环移位的密集采样策略,获取的负样本具有更强的判别能力和抗遮挡性,在这两个视频序列的表现相当出色,只在部分帧丢失目标(序列(f):1 028 帧~1 058 帧,序列(j):71 帧~78 帧);本文算法通过度量遮挡前后权重最高的子块(序列(f)的白色眼睛子块、序列(j)的黑色腿部子块)权值的变化程度及时检测到目标状态的变化,判定遮挡发生后对各目标子块通过扩展粒子滤波的采样范围又重新捕捉到目标.除 BACF 之外的其他方法由于不能有效地连续检测和处理遮挡,在遮挡情况下很容易产生跟踪漂移,甚至丢失目标,丢失目标后也基本失去了重新捕捉目标的能力(如 STC 算法在序列(c)83 帧、序列(f)348 帧、序列(h)237 帧,SPT 算法在序列(c)107 帧、序列(f)380 帧,Staple、DSST、CN 算法在序列(f)381 帧均完全丢失目标).

(2) 目标快速移动实验

图 6(a)~图 6(c)、图 6(f)展示了不同序列目标快速移动时的跟踪结果.由于 Frag 算法根据上一帧的目标位置采用较小的固定半径对其邻域进行搜索,因此对目标的快速移动非常敏感,对快速移动目标的跟踪效果较差,在(b)、(c)、(f)这 3 个测试序列中都是最先出现跟踪漂移;LSH 和 SPT 方法对快速运动的物体也有较强的捕捉能力,但容易受到局部出现的相似目标以及遮挡的干扰引起跟踪失败(如 LSH 在序列(a)168 帧丢失目标,SPT 在序列(c)107 帧丢失目标);STC 算法因遮挡及时空上下文模型无法有效地适应目标和背景的快速变化,也很快产生了跟踪漂移(如序列(c)83 帧所示).实验显示,BACF、KCF、Staple、DSST、CN 对目标快速运动均有较强的适应性,而本文方法通过度量子块置信度的变化能够及时感知到目标状态变化,对目标快速移动也有具有较强的鲁棒性.

(3) 光照及目标姿态、尺度变化实验

图 6(a)、图 6(b)、图 6(e)~图 6(g)、图 6(i)、图 6(k)~图 6(m)展示了姿态、尺度和光照变化条件下的目标跟踪结果.

序列(g)展示的是连续光照变化条件下对运动轿车的跟踪,受光照变化的影响,基于颜色特征的 DAT 算法很快丢失目标;SPT 和 Frag 算法也产生了跟踪漂移(如 73 帧所示);STC 算法虽然能够锁定跟踪目标,但是跟踪窗尺度变化不准确,包含了较多的背景像素(如 221 帧所示);其他算法在此序列测试中均表现出很好的鲁棒性.序列(i)展示了亮度缓慢变化过程中对移动人脸的跟踪结果,随着目标的移动和光照的变化,Frag 和 DAT 出现较大的跟踪漂移,其他算法都能稳定地跟踪目标.

序列(a)、序列(b)、序列(e)展示的跟踪过程,目标姿态产生了较大变化,基于整体特征模型的 STC 跟踪算法在目标模板发生显著变化时,容易产生欠更新或过更新导致跟踪漂移(如序列(a)130 帧、序列(b)60 帧所示);由于局部邻域内目标背景的可分性高,基于超像素模型的 SPT 算法和本文算法能够获取到稳定的高置信度目标子块,都能较为准确地跟踪目标;LSH 算法对目标姿态变化也表现出了较强的鲁棒性;基于相关滤波的 BACF、KCF、CN、Staple、DSST 算法在这 3 个序列中表现出了较大的差异,主要是由于除了目标姿态变化较大外,还伴有其他复杂因素的干扰(详见相似目标及复杂背景实验).

序列(f)展示了目标尺度变化的跟踪结果,除 Frag 算法对目标尺度变化较敏感外,其他算法均能在目标尺度变化时较稳定的跟踪目标,直到出现遮挡。

序列(l)是光照发生较强变化、目标和背景可分性较低情况下的跟踪结果。由于可分性较低,目标局部密集采样策略获取的样本质量不高,训练的相关滤波分类器判别能力受到影响,导致 Staple 和 KCF 算法很快丢失目标;BACF 算法通过扩大样本搜索区域,能够提取到更多判别性较高的负样本,训练的相关滤波分类器模型表现出色,始终能够精确定位目标;LSH 和本文方法对光照变化不敏感,都能准确地跟踪目标;由于目标和背景始终没有产生较大的变化,使得 STC 算法能够学习到一个相对稳定的空间上下文模型,也取得了很好的跟踪结果。

序列(k)是所有测试序列中最具挑战性的一个,目标自身产生了巨大形变,运动轨迹不够平滑,并伴有复杂背景的干扰,所有测试算法的跟踪正确率(TA)都很低。本文算法在目标形变过程中,初始阶段利用权值相对较高的腿部子块捕捉目标;随着目标自身的剧烈形变,通过增加分块数量,扩大搜索区域,利用置信度较高的腿部和头部子块定位目标,在测试算法中取得了最好结果。DAT 算法由于引入了基于颜色统计特征的干扰感知模型,对目标的形变不敏感,跟踪效果也好于其他方法。基于相关滤波框架的跟踪算法在此条件下无法获取高质量的训练样本,很快产生漂移,丢失目标。

序列(m)是一个简单的灰度图序列,目标对背景有较好的可分性,虽然目标姿态产生了变化,但测试算法都能稳定跟踪,是所有测试算法整体表现最好的一个序列。这也说明无论采用何种特征模型,可分性都至关重要。

(4) 相似目标及复杂背景实验

图 6(a)、图 6(b)、图 6(e)、图 6(f)、图 6(k)、图 6(l)展示了相似目标干扰及复杂背景条件下的跟踪结果,各序列除了相似目标干扰外,还存在着遮挡和复杂背景的影响,对算法的鲁棒性具有很大的挑战。

序列(a)除了目标姿态发生较大变化外,跟踪目标周围还存在很多相似目标的干扰,但在局部邻域内目标和背景的可分性很好,基于密集采样策略的 CN、KCF 和 Staple 算法均能获得高质量的训练样本,跟踪准确性很高;由于整帧背景和背景的可分性高,BACF 算法的整帧采样策略同样能够获得到高质量的训练样本也表现出了很高的跟踪精度和鲁棒性;本文算法和干扰感知算法 DAT 对相似目标的干扰同样表现出了较好的抗干扰能力(如 183 帧所示);其他算法对相似目标的干扰较为敏感,均出现了丢失跟踪目标或将跟踪窗定位在干扰目标上的现象(如 183 帧、243 帧、349 帧所示)。

序列(e)展示的篮球比赛的跟踪结果类似于序列(a)的情形,本文算法始终能够选择到较高置信度的目标子块和基于干扰感知模型的 DAT 算法都表现出较高的跟踪性能。由于初始阶段目标和局部背景的可分性较高,基于相关滤波的 KCF、CN、DSST 和 Staple 算法都能稳定跟踪目标。随着目标姿态、尺度变化以及相似目标的相互遮挡等因素的干扰,基于模型整体更新的 CN 算法和 KCF 算法逐渐出现了较大的跟踪漂移,虽然仍能跟踪目标,但跟踪窗已经包含了大量的背景像素(如 407 帧、635 帧所示)。Staple 算法由于融合了互补性很强的 HOG 特征和颜色特征,表现出了较强的鲁棒性。DSST 算法在位置滤波器的基础上引入了尺度滤波器,定位目标的同时能够获取精确的目标尺度模型,也具有很高的跟踪精度。由于大量相似目标分布于整帧背景,BACF 算法通过扩大采样范围,利用真实背景信息循环移位产生的负样本中包含了太多相似目标,导致分类器很难判定真实目标位置,很快丢失目标。

序列(b)展示的是对一只往复运动的小鸟的跟踪过程,其中有较大范围的相似背景区域,目标和背景颜色相近,目标姿态产生了较大变化,并出现短暂的遮挡,除 Frag、STC、BACF 和 DSST 算法受上述因素的影响,跟踪窗出现较大偏差外,其他算法都能有效地跟踪目标(如 49 帧、98 帧所示)。

序列(f)展示了复杂背景下对快速移动目标的跟踪,跟踪过程中伴有目标尺度变化和长时间遮挡,除 Frag 算法很快丢失目标外,其他算法均能较稳定地跟踪目标,直到出现长时间遮挡后丢失目标。

由上述分析对比可知,在复杂场景下,特别是在遮挡、光照变化和相似背景干扰情况下,所有对比算法中,本文算法、BACF 以及 Staple 算法综合性能表现出色,在部分测试序列中二者的表现略好于本文算法,但总体看来,本文算法的表现更加稳定,平均跟踪正确率和精确度更高。

3.3 定量分析

本文采用平均中心误差和跟踪正确率对跟踪结果进行定量分析比较,利用 PASCAL VOC 检测准则^[47]给出的跟踪正确率和平均重合率(AOR)作为跟踪成功与否的评价指标,当 AOR 大于指定阈值时,认为跟踪成功,本文选择 $AOR > 0.5$, 设 R 为跟踪结果框, T 为真实目标框, M 为测试序列总帧数, M_0 为成功跟踪帧数, 则 TA 和 AOR 计算如下:

$$AOR = \frac{R \cap T}{R \cup T} \quad (22)$$

$$TA = \frac{M_0}{M} \quad (23)$$

跟踪正确率(TA)描述视频序列中成功跟踪的帧数,该值越大,说明跟踪成功的帧数越多,跟踪结果的准确性越高.平均中心误差反映了跟踪结果与实际目标中心点偏差的均值,该值越小,说明跟踪精确度越高.表 3 给出了本文算法和其他方法跟踪正确率对比,表 4 给出了本文算法和其他方法平均中心误差对比(红色为最优结果,蓝色为次优结果).为了避免较大的误差引入,对不同的测试序列动态调整了中心点误差阈值.

Table 3 Comparison of TA

表 3 跟踪正确率对比

| Sequences | Proposed | Staple | BACF | KCF | DAT | DSST | CN | SPT | LSH | STC | Frag |
|------------|----------|--------|------|------|------|------|------|------|------|------|------|
| Basketball | 0.92 | 0.88 | 0.65 | 0.89 | 0.96 | 0.90 | 0.70 | 0.86 | 0.87 | 0.79 | 0.78 |
| Bird2 | 0.93 | 0.96 | 0.67 | 0.85 | 1.00 | 0.49 | 0.99 | 0.90 | 0.92 | 0.56 | 0.72 |
| Bolt1 | 0.90 | 1.00 | 1.00 | 0.98 | 0.68 | 0.02 | 1.00 | 0.90 | 0.69 | 0.86 | 0.63 |
| Lemming | 0.96 | 0.27 | 0.96 | 0.96 | 0.91 | 0.27 | 0.31 | 0.90 | 0.98 | 0.88 | 0.36 |
| Faceocc1 | 1.00 | 1.00 | 1.00 | 1.00 | 0.10 | 1.00 | 0.90 | 1.00 | 1.00 | 1.00 | 0.87 |
| Woman | 0.97 | 0.94 | 0.99 | 0.48 | 0.21 | 0.94 | 0.25 | 0.80 | 0.83 | 0.78 | 0.71 |
| Tiger2 | 0.90 | 0.89 | 0.90 | 0.96 | 0.62 | 0.92 | 0.64 | 0.31 | 0.65 | 0.54 | 0.05 |
| Car2 | 1.00 | 1.00 | 1.00 | 1.00 | 0.07 | 1.00 | 1.00 | 0.72 | 0.94 | 0.83 | 0.77 |
| David1 | 0.92 | 1.00 | 0.99 | 1.00 | 0.01 | 1.00 | 0.90 | 0.92 | 0.76 | 1.00 | 0.64 |
| Diving | 0.62 | 0.36 | 0.44 | 0.01 | 0.53 | 0.40 | 0.25 | 0.42 | 0.46 | 0.20 | 0.32 |
| Jogging | 0.99 | 0.23 | 0.99 | 0.99 | 0.23 | 0.23 | 0.23 | 0.22 | 0.90 | 0.23 | 0.47 |
| Skater | 0.92 | 0.96 | 0.96 | 0.91 | 0.82 | 0.89 | 0.68 | 0.80 | 0.96 | 0.52 | 0.90 |
| Shaking | 0.96 | 0.04 | 0.98 | 0.31 | 0.23 | 0.95 | 0.69 | 0.64 | 0.91 | 0.98 | 0.62 |
| Average | 0.92 | 0.73 | 0.89 | 0.80 | 0.49 | 0.69 | 0.66 | 0.72 | 0.84 | 0.71 | 0.60 |

Table 4 Comparison of the average center errors

表 4 平均中心误差对比

| Sequences | Proposed | Staple | BACF | KCF | DAT | DSST | CN | SPT | LSH | STC | Frag |
|------------|----------|--------|------|-------|------|-------|------|-------|------|------|-------|
| Basketball | 6.7 | 19.1 | 23.9 | 40.1 | 7.2 | 9.7 | 14.6 | 6.9 | 9.4 | 9.2 | 25.2 |
| Bird2 | 7.7 | 8.2 | 16.7 | 10.8 | 7.2 | 51.9 | 6.0 | 17.4 | 8.3 | 19.4 | 27.6 |
| Bolt1 | 5.1 | 3.7 | 3.7 | 5.5 | 16.9 | 103.0 | 4.1 | 6.3 | 6.5 | 8.6 | 39.2 |
| Lemming | 8.2 | 118.0 | 12.4 | 9.7 | 22.3 | 136.0 | 90.7 | 7.7 | 10.4 | 11.9 | 84.6 |
| Faceocc1 | 6.6 | 19.6 | 9.5 | 5.4 | 98.3 | 13.5 | 12.9 | 6.8 | 5.7 | 76.0 | 7.2 |
| Woman | 4.5 | 5.9 | 5.3 | 27.9 | 58.5 | 10.6 | 89.0 | 11.4 | 6.4 | 15.1 | 19.3 |
| Tiger2 | 6.3 | 10.3 | 10.8 | 13.6 | 13.9 | 11.2 | 18.3 | 24.7 | 8.5 | 9.4 | 45.7 |
| Car2 | 6.8 | 3.5 | 4.6 | 3.6 | 57.2 | 1.4 | 2.7 | 36.4 | 4.1 | 12.4 | 16.8 |
| David1 | 9.4 | 6.7 | 8.3 | 17.5 | 29.8 | 7.1 | 13.5 | 9.2 | 11.6 | 8.7 | 71.3 |
| Diving | 18.6 | 72.7 | 43.5 | 106.0 | 36.7 | 62.8 | 77.5 | 86.3 | 83.4 | 92.0 | 87.2 |
| Jogging | 9.7 | 93.2 | 6.7 | 9.1 | 98.4 | 90.2 | 96.3 | 23.8 | 13.2 | 58.4 | 59.8 |
| Skater | 10.7 | 7.3 | 11.8 | 11.2 | 16.3 | 11.5 | 13.8 | 13.6 | 9.7 | 29.8 | 12.2 |
| Shaking | 8.2 | 125.0 | 8.5 | 17.8 | 86.6 | 8.0 | 15.1 | 102.0 | 9.8 | 8.7 | 109.7 |
| Average | 8.3 | 37.9 | 12.7 | 21.4 | 42.3 | 39.8 | 35.0 | 27.1 | 14.4 | 27.7 | 46.6 |

需要说明的是,本文所提算法的跟踪性能和参数设置密切相关,在不同测试序列中,由于初始跟踪目标的大小、位置等存在较大差异,并且超像素子块存在聚合、分裂或消失等不确定因素,为了在不同的测试序列中均能较好地分离目标和背景,获取到一定数量的稳定目标子块以保证跟踪的精确性,本文对 SLIC 超像素分割算法做了改进和优化,保证每一帧都能获取到大小均匀、不变性好的超像素子块,计算效率也有较大的提高.在超像素模型的训练阶段,对均值聚类算法的核窗宽以及改进的 SLIC 超像素分割算法的压缩因子、超像素个数和迭

代次数分别进行动态调整.初始化核窗宽 $h=0.1, h \in [0.1, 0.3]$, 压缩因子 $cf=10, cf \in [10, 60]$, 超像素个数 $sc=10, sc \in [10, 50]$, 迭代次数 $ic=3, ic \in [3, 15]$, 分别以步长 0.02, 5, 1 和 1 动态增加.较大的核窗宽有助于将目标超像素块从背景中分离出来,在此前提下,较小的核窗宽有助于将目标超像素块分离到不同聚类,保证子块的多样性.

对多数实验序列的测试结果表明,每个目标模型能够获取到 2 个~5 个稳定的目标子块时,跟踪精度与跟踪效率能够达到较好的平衡.因此,本文选择在这一条件下进行对比实验.如果损失一定的计算效率,那么在训练和更新阶段,通过约束优化子块模型获取数量更多、显著性更高、稳定性更好的目标子块进行跟踪,跟踪精度可以进一步提升.表 5 给出了本文算法的详细参数设置.

Table 5 Experimental parameters setting of different sequences

表 5 不同序列实验参数设置

| Sequences | Sample range | sc | cf | ic | h |
|------------|--------------|----|----|----|------|
| Basketball | 51×122 | 15 | 40 | 7 | 0.18 |
| Bird2 | 104×110 | 19 | 30 | 6 | 0.20 |
| Bolt1 | 39×92 | 12 | 60 | 4 | 0.18 |
| Lemming | 92×155 | 29 | 60 | 4 | 0.18 |
| Faceoccl | 171×243 | 41 | 60 | 6 | 0.20 |
| Woman | 32×143 | 12 | 60 | 6 | 0.16 |
| Tiger2 | 102×117 | 17 | 60 | 5 | 0.20 |
| Car2 | 96×78 | 11 | 50 | 3 | 0.16 |
| David1 | 113×143 | 27 | 50 | 4 | 0.20 |
| Diving | 32×194 | 25 | 40 | 12 | 0.22 |
| Jogging | 38×152 | 12 | 50 | 6 | 0.20 |
| Skater | 59×206 | 21 | 60 | 4 | 0.18 |
| Shaking | 92×107 | 18 | 60 | 11 | 0.22 |

图 7 给出了本文算法在不同视频序列中使用的部分训练样本.从图中可以看出,利用超像素子块模型结合有效的子块置信度量机制,本文算法能够获取到高质量的样本参与训练与跟踪过程,提高了跟踪的鲁棒性.

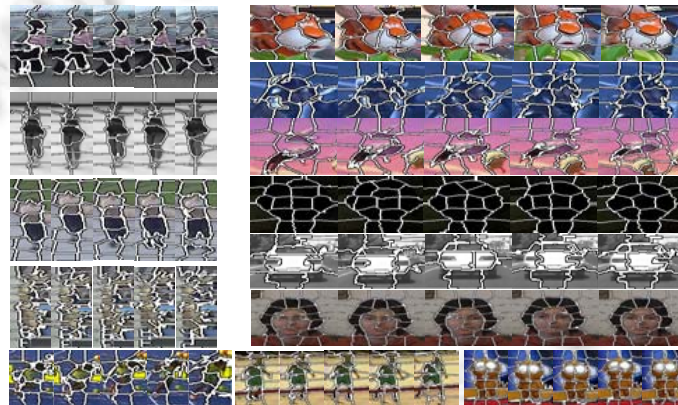


Fig.7 Training samples of different sequences

图 7 不同序列训练样本

3.4 时间复杂度

从算法的时间复杂度考虑,本文算法的耗时与选择训练超像素模型的样本数量、选择目标区域的大小、对目标区域进行超像素分割的数量以及选择的目标子块粒子数量等因素有关.为了提高运算速度,本文进行目标分块时,仅对目标及其周边邻域进行超像素分割,能够有效地减少计算时间;将目标区域超像素分割后得到的稳定目标子块数量控制在 2 个~5 个之间,分块数量过多时,计算复杂度会明显增加,在训练和更新阶段如果未对子块模型进一步筛选和优化的话,跟踪精度并没有明显的提高.

表 6 给出了各算法在不同测试序列中的跟踪效率对比,表中数据为各算法每秒处理的帧数.

Table 6 Comparison of the algorithm efficiency

(frame/s)

表 6 算法效率对比

(帧/秒)

| Sequences | proposed | Staple | BACF | KCF | DAT | DSST | CN | SPT | LSH | STC | Frag |
|------------|----------|--------|-------|-------|-------|-------|-------|------|-------|-------|-------|
| Basketball | 5.97 | 10.59 | 22.33 | 19.39 | 13.69 | 75.90 | 96.8 | 1.76 | 7.02 | 12.68 | 4.16 |
| Bird2 | 4.26 | 11.42 | 18.47 | 19.42 | 31.44 | 42.90 | 44.5 | 1.67 | 6.62 | 14.49 | 3.90 |
| Bolt1 | 6.49 | 12.53 | 19.96 | 20.47 | 14.11 | 88.20 | 141.0 | 1.38 | 10.87 | 10.52 | 4.34 |
| Lemming | 2.72 | 9.41 | 17.54 | 20.01 | 21.26 | 50.80 | 39.7 | 0.77 | 4.57 | 14.96 | 3.17 |
| Faceocc1 | 1.18 | 11.32 | 13.08 | 21.89 | 38.16 | 18.50 | 23.2 | 0.66 | 1.98 | 9.70 | 7.64 |
| Woman | 9.88 | 12.43 | 20.66 | 22.53 | 9.77 | 86.80 | 152.0 | 2.45 | 11.26 | 12.49 | 7.92 |
| Tiger2 | 3.46 | 10.44 | 19.07 | 17.66 | 23.78 | 43.80 | 64.2 | 1.98 | 6.33 | 16.59 | 3.38 |
| Car2 | 7.56 | 12.07 | 23.98 | 22.13 | 27.72 | 72.00 | 194.0 | 1.53 | 11.77 | 28.34 | 10.32 |
| David1 | 2.89 | 6.77 | 17.26 | 18.63 | 29.13 | 39.70 | 31.6 | 0.62 | 5.31 | 22.31 | 9.51 |
| Diving | 5.77 | 11.26 | 21.02 | 25.16 | 4.02 | 87.50 | 88.2 | 1.66 | 7.00 | 28.07 | 7.90 |
| Jogging | 5.23 | 12.24 | 17.01 | 23.79 | 6.31 | 77.90 | 94.7 | 1.53 | 6.94 | 16.31 | 7.37 |
| Skater | 5.97 | 13.99 | 19.52 | 26.63 | 12.51 | 45.20 | 103.0 | 1.23 | 6.17 | 24.23 | 9.16 |
| Shaking | 3.82 | 11.92 | 16.48 | 19.88 | 28.17 | 51.70 | 61.1 | 0.88 | 4.13 | 25.96 | 4.73 |
| Average | 4.62 | 11.26 | 18.95 | 21.35 | 20.01 | 60.07 | 87.23 | 1.39 | 6.92 | 18.20 | 5.88 |

从表 6 可以看出,本文算法在计算效率上,整体高于 SPT 算法,低于基于相关滤波的目标跟踪方法,略低于 Staple 和 LSH 算法,在有些序列中高于 Frag 算法而有些序列中低于 Frag 算法.这是由于本文算法在运行过程中,根据实际选定的目标大小动态调整了超像素分割的压缩因子、超像素数量等相关参数.通过在跟踪过程中加入超像素模型的训练和子块异步更新策略,结合局部敏感直方图良好的光照不变特性,保证目标模型对复杂场景具有很强的适应性,虽然损失了一部分计算效率,但换取了更高的跟踪正确率和精确度.

本文算法对最耗时的两个序列 Faceocc1 和 David1 训练阶段的平均耗时为 4.66s,跟踪阶段,每帧得到跟踪结果平均需要 0.47s.通过优化算法,在保证跟踪精度的同时,基本能够满足实时性要求.

4 结 论

本文针对复杂环境下目标跟踪算法面临的主要问题展开研究,提出了基于多模型协作的自适应分块目标跟踪方法,有效地解决了光照变化及遮挡等因素影响跟踪鲁棒性的问题.该方法融合基于局部敏感直方图的产生式模型和基于超像素分割的判别式模型构建目标表现模型,利用相对熵和均值聚类度量各分块的局部差异置信度和目标背景置信度,建立双权值约束机制,保证高置信度目标子块参与到跟踪过程中以提高跟踪的精确性和鲁棒性.为了避免全局搜索,算法在粒子滤波框架下,联合高置信度分块进行目标定位.此外,本文算法还提供了有效的遮挡检测和处理机制,对目标子块采取异步更新策略,通过度量目标子块置信度的变化,能够及时感知目标模型的当前状态,确定目标模型何时需要更新以及哪些子块模型需要更新,保证更新后的目标模型几乎不包含背景信息,能够很好地适应目标的局部变化和遮挡影响.实验结果表明,本文算法与其他跟踪方法相比更加稳定,在多种复杂场景下(包括遮挡光照及姿态/尺度变化、快速运动、相似目标和复杂背景干扰),具有更高的平均跟踪正确率和精确度.

下一步研究工作的重点是:(1) 将本文算法框架和相关滤波跟踪框架相结合,进一步提高目标跟踪的效率;(2) 将本文算法模型和深度学习网络结合,寻找更有效的目标划分和融合策略,建立简单、高效的目标模型,进一步提高目标跟踪的精度.

References:

- [1] Li X, Hu W, Shen C, *et al.* A survey of appearance models in visual object tracking. *ACM Trans. on Intelligent Systems & Technology*, 2013,4(4):1-42. [doi: 10.1145/2508037.2508039]
- [2] Ross DA, Lim J, Lin RS, Yang MH. Incremental learning for robust visual tracking. *Advances in Neural Information Processing Systems*, 2008,77(1-3):125-141. [doi: 10.1007/s11263-007-0075-7]
- [3] Kwon J, Lee KM. Visual tracking decomposition. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2010. 1269-1276. [doi: 10.1109/CVPR.2010.5539821]

- [4] He S, Yang Q, Lau R W H, *et al.* Visual tracking via locality sensitive histograms. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013. 2427–2434. [doi: 10.1109/CVPR.2013.314]
- [5] Zhong W, Lu H, Yang MH. Robust object tracking via sparsity-based collaborative model. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012. 1838–1845. [doi: 10.1109/CVPR.2012.6247882]
- [6] Jia X, Lu H, Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012. 1822–1829. [doi: 10.1109/CVPR.2012.6247880]
- [7] Wang D, Lu H, Xiao Z, Yang MH. Inverse sparse tracker with a locally weighted distance metric. IEEE Trans. on Image Processing, 2015,24(9):2646–2657. [doi: 10.1109/TIP.2015.2427518]
- [8] Babenko B, Yang MH, Belongie S. Robust object tracking with online multiple instance learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011,33(8):1619–1632. [doi: 10.1109/TPAMI.2010.226]
- [9] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 34(7):1409–1422. [doi: 10.1109/TPAMI.2011.239]
- [10] Bai Y, Tang M. Robust tracking via weakly supervised ranking SVM. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012. 1854–1861. [doi: 10.1109/CVPR.2012.6247884]
- [11] Hare S, Saffari A, Torr PH. Struck: Structured output tracking with kernels. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Piscataway: IEEE, 2011. 263–270. [doi: 10.1109/ICCV.2011.6126251]
- [12] Possegger H, Mauthner T, Bischof H. In defense of color-based model-free tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 2113–2120. [doi: 10.1109/CVPR.2015.7298823]
- [13] Leistner C, Saffari A, Bischof H. MIForests: Multiple-instance learning with randomized trees. In: Proc. of the European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2010. 29–42.
- [14] Xue MG, Zhu H, Yuan GL. Online robust discrimination dictionary learning for visual tracking. Acta Electronica Sinica, 2016, 44(4):838–845 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2016.04.012]
- [15] Bolme DS, Beveridge JR, Draper BA, *et al.* Visual object tracking using adaptive correlation filters. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010. 2544–2550. [doi: 10.1109/CVPR.2010.5539960]
- [16] Henriques JF, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels. In: Proc. of the European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2012. 702–715. [doi: 10.1007/978-3-642-33765-9_50]
- [17] Henriques JF, Caseiro R, Martins P, *et al.* High-speed tracking with kernelized correlation filters. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(3):583–596. [doi: 10.1109/TPAMI.2014.2345390]
- [18] Danelljan M, Shahbaz Khan F, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014. 1090–1097. [doi: 10.1109/CVPR.2014.143]
- [19] Danelljan M, Hager G, Khan FS, *et al.* Discriminative scale space tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2017,39(8):1561–1575. [doi: 10.1109/TPAMI.2016.2609928]
- [20] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration. In: Proc. of the European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2014. 254–265. [doi: 10.1007/978-3-319-16181-5_18]
- [21] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: Complementary learners for real-time tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 1401–1409. [doi: 10.1109/CVPR.2016.156]
- [22] Wang M, Liu Y, Huang Z. Large margin object tracking with circulant feature maps. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017. 4800–4808. [doi: 10.1109/CVPR.2017.510]
- [23] Galoogahi HK, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Piscataway: IEEE, 2017. 1144–1152. [doi: 10.1109/ICCV.2017.129]
- [24] Liu T, Wang G, Yang Q. Real-time part-based visual tracking via adaptive correlation filters. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 4902–4912. [doi: 10.1109/CVPR.2015.7299124]
- [25] Zuo W, Wu X, Lin L, *et al.* Learning support correlation filters for visual tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018. [doi: 10.1109/TPAMI.2018.2829180]
- [26] Ma C, Yang X, Zhang C, *et al.* Long-term correlation tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 5388–5396. [doi: 10.1109/CVPR.2015.7299177]

- [27] Kristan M, Leonardis A, Matas J, *et al.* The visual object tracking VOT2017 challenge results. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Piscataway: IEEE, 2017. 1949–1972.
- [28] Wang N, Yeung DY. Learning a deep compact image representation for visual tracking. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. Curran Associates Inc., 2013. 809–817.
- [29] Wang LJ, Ouyang WL, Wang XG, Lu HC. Visual tracking with fully convolutional networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Piscataway: IEEE, 2015. 3119–3127. [doi: 10.1109/ICCV.2015.357]
- [30] Ma C, Huang JB, Yang XK, Yang MH. Hierarchical convolutional features for visual tracking. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Piscataway: IEEE, 2015. 3074–3082. [doi: 10.1109/ICCV.2015.352]
- [31] Wang L, Ouyang W, Wang X, *et al.* STCT: Sequentially training convolutional networks for visual tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 1373–1381. [doi: 10.1109/CVPR.2016.153]
- [32] Qi Y, Zhang S, Qin L, *et al.* Hedged deep tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 4303–4311. [doi: 10.1109/CVPR.2016.466]
- [33] Li HX, Li Y, Porikli F. DeepTrack: Learning discriminative feature representations online for robust visual tracking. IEEE Trans. on Image Processing, 2016,25(4):1834–1848. [doi: 10.1109/TIP.2015.2510583]
- [34] Li HX, Li Y, Porikli F. Robust online visual tracking with a single convolutional neural network. In: Proc. of the 12th Asian Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2014. 194–209.
- [35] Adam A, Rivlin E, Shimshoni I. Robust fragments based tracking using the integral histogram. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2006. 798–805. [doi: 10.1109/CVPR.2006.256]
- [36] Li Y, Zhu J, Hoi SCH. Reliable patch trackers: Robust visual tracking by exploiting reliable patches. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 353–361. [doi: 10.1109/CVPR.2015.7298632]
- [37] Yang F, Lu HC, Yang MH. Robust superpixel tracking. IEEE Trans. on Image Processing, 2014,23(4):1639–1651. [doi: 10.1109/TIP.2014.2300823]
- [38] Liu S, Zhang T, Cao X, *et al.* Structural correlation filter for robust visual tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 4312–4320. [doi: 10.1109/CVPR.2016.467]
- [39] Kwon J, Lee KM. Highly nonrigid object tracking via patch-based dynamic appearance modeling. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013,35(10):2427–2441. [doi: 10.1109/TPAMI.2013.32]
- [40] Bouachir W, Bilodeau GA. Collaborative part-based tracking using salient local predictors. Computer Vision & Image Understanding, 2015,137(C):88–101. [doi: 10.1016/j.cviu.2015.03.010]
- [41] Hu ZH, Yuan XT, Li J, *et al.* Robust fragments-based tracking with multi-feature joint kernel sparse representation. Journal of Computer Research and Development, 2015,52(7):1692–1704 (in Chinese with English abstract) [doi: 10.7544/issn1000-1239.2015.20140152]
- [42] Hou ZQ, Dai B, Hu D, *et al.* Online visual tracking based on local patch learning. Journal of Electronics & Information Technology, 2016,38(7):1616–1623 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2015.01.012]
- [43] Wang MH, Liang Y, Liu FM, *et al.* Object tracking based on component-level appearance model. Ruan Jian Xue Bao/ Journal of Software, 2015,26(10):2733–2747 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4737.htm> [doi: 10.13328/j.cnki.jos.004737]
- [44] Achanta R, Shaji A, Smith K, Lucchi A, Fua P. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(11):2274–2282. [doi: 10.1109/TPAMI.2012.120]
- [45] Wu Y, Lim J, Yang MH. Object tracking benchmark. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(9): 1834–1848. [doi: 10.1109/TPAMI.2014.2388226]
- [46] Zhang K, Zhang L, Liu Q, *et al.* Fast visual tracking via dense spatio-temporal context learning. In: Proc. of the European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2014. 127–141. [doi: 10.1007/978-3-319-10602-1_9]
- [47] Everingham M, Gool LV, Williams CKI, *et al.* The pascal visual object classes (VOC) challenge. Int'l Journal of Computer Vision, 2010,88(2):303–338. [doi: 10.1007/s11263-009-0275-4]

附中文参考文献:

- [14] 薛模根,朱虹,袁广林.在线鲁棒判别式字典学习视觉跟踪.电子学报,2016,44(4):838-845. [doi: 10.3969/j.issn.0372-2112.2016.04.012]
- [41] 胡昭华,袁晓彤,李俊,等.基于目标分块多特征核稀疏表示的视觉跟踪.计算机研究与发展,2015,52(7):1692-1704. [doi: 10.7544/j.issn1000-1239.2015.20140152]
- [42] 侯志强,戴铂,胡丹,等.基于感知深度神经网络的视觉跟踪.电子与信息学报,2016,38(7):1616-1623. [doi: 10.3969/j.issn.0372-2112.2015.01.012]
- [43] 王美华,梁云,刘福明,等.部件级表现模型的目标跟踪方法.软件学报,2015,26(10):2733-2747. <http://www.jos.org.cn/1000-9825/4737.htm> [doi: 10.13328/j.cnki.jos.004737]



刘明华(1980-),男,山东阳谷人,博士,讲师,主要研究领域为计算机视觉,目标识别与跟踪,智能信息系统.



王传旭(1968-),男,博士,教授,主要研究领域为计算机图像,视频处理.



汪传生(1960-),男,博士,教授,博士生导师,主要研究领域为机器人技术,机电控制理论,先进制造技术.



崔雪红(1978-),女,博士,讲师,主要研究领域为计算机视觉,目标识别.



胡强(1980-),男,博士,副教授,CCF 专业会员,主要研究领域为 Petri 网理论,服务计算,软件形式化分析方法.