一种时空特征融合的鲁棒视觉跟踪算法

杨志龙[1]，侯志强[1]，余旺盛[2]，蒲　磊[3]，张成煜[1]，马素刚[1]
（1.西安邮电大学计算机学院，西安，710121；2.空军工程大学信息与导航学院，西安，710077；
3.火箭军工程大学作战保障学院，西安，710025）

摘要　针对视觉跟踪在复杂背景下因外观特征表征不足等原因造成的目标丢失问题，结合深度光流网络估计的运动特征，文中提出了一种基于时序信息和空间信息自适应融合的视觉跟踪算法。该算法在相关滤波跟踪框架基础上，引入递归全对场变换(RAFT)深度网络估计光流以获取目标的时序信息，提取目标的 CN 特征和 HOG 特征获取空间信息，然后融合目标时序信息和空间信息，以增强对目标时空特征的表征能力；其次，建立了一种跟踪结果质量判别机制，实时调整时序信息在融合过程中的权重，有效提升了算法在复杂动态环境下的泛化能力。为评估算法的有效性，在 OTB100 和 VOT2019 两个数据集上进行了测试，实验结果表明，与主流视觉跟踪算法相比，所提算法的跟踪性能获得了显著提升，尤其在运动模糊、快速运动等属性的视频中，具有明显优势。

关键词　视觉跟踪；时空特征；特征融合；相关滤波
DOI　10.3969/j.issn.2097-1915.2022.06.008
中图分类号　TP391.41　　**文献标志码**　A　　**文章编号**　2097-1915(2022)06-0057-07

A Robust Visual Tracking Algorithm Based on Temporal and Spatial Feature Fusion

YANG Zhilong[1], HOU Zhiqiang[1], YU Wangsheng[2], PU Lei[3], ZHANG Chengyu[1], MA Sugang[1]
(1. College of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 2. Information and Navigation School, Air Force Engineering University, Xi'an 710077, China; 3. Operational Support School, Rocket Force University of Engineering, Xi'an 710025, China)

Abstract　Aimed at the problems that visual tracking is loss caused by insufficient appearance feature representation under complex backgrounds in combination with the motion features estimated by deep optical flow network, a visual tracking algorithm is proposed based on adaptive fusion of temporal information and spatial information. On the basis of correlation filtering tracking framework, the Recurrent All-Pairs Field Transforms (RAFT) deep network is utilized for estimating the optical flow to obtain the timing information of the target, and extract the CN feature and HOG feature to obtain spatial information, and fuse the target timing information and spatial information to enhance the characterization ability of the target's temporal and spatial characteristics, and then a mechanism for discriminating the reliability of tracking results

收稿日期：2021-10-13
基金项目：国家自然科学基金(62072370)
作者简介：杨志龙(1996—)，男，甘肃会宁人，硕士生，研究方向为目标跟踪。E-mail:2323128817@qq.com

引用格式：杨志龙，侯志强，余旺盛，等. 一种时空特征融合的鲁棒视觉跟踪算法[J]. 空军工程大学学报，2022，23(6)：57-63. YANG Zhilong, HOU Zhiqiang, YU Wangsheng, et al. A Robust Visual Tracking Algorithm Based on Temporal and Spatial Feature Fusion[J]. Journal of Air Force Engineering University, 2022, 23(6): 57-63.

is established, and the weight of time sequence information in the fusion process is adjusted in real time, effectively improving the generalization ability of the algorithm in a complex dynamic environment. In order to evaluate the effectiveness of the algorithm in this paper, the tests are carried out on two data sets, OTB100 and VOT2019 respectively. The experimental results show that compared with the mainstream visual tracking algorithms in recent years, the tracking performance is improved by the algorithm, especially in motion blur, fast motion and other attributes of the video. And this algorithm has obvious advantages.

Key words visual tracking; temporal and spatial characteristics; feature fusion; correlation filtering

视觉目标跟踪是计算机视觉领域研究的热点之一,在视频理解、人机交互、无人机、自动驾驶等方面具有广泛的应用[1]。近年来,视觉跟踪领域涌现出诸多优异的成果,但仍然面临着不小的挑战[2],在跟踪过程中,当目标发生形变、遮挡、旋转、运动模糊以及处于复杂场景等情况下,往往会造成目标丢失。

近几年,相关滤波跟踪算法[3]以其良好的跟踪性能吸引了众多研究人员的关注。相关滤波类跟踪算法通常采用手工特征[4-6]对目标外观进行表征,基于深度学习的 CNN 特征[7]也被广泛应用,并取得了良好的跟踪效果。文献[8]在分析了深度特征和手工特征各自的优势后,提出不同特征应区别对待,调整不同特征的权重系数进行融合,发挥不同特征的优势,进一步提升了跟踪器的性能。在基于相关滤波的跟踪算法中,ECO 算法[9]表现突出,该算法使用了 CN 特征、HOG 特征和 CNN 特征的组合,获得了很好的跟踪性能。

但上述这些特征都属于表征目标表观特征的静态信息,在运动目标的跟踪中,如果引入表征目标运动特征的动态信息或者时序信息,应该能够进一步提升算法的跟踪性能。文献[10]提出递归全对场变换深度网络(recurrent all-pairs field transforms, RAFT),通过递归单元迭代更新的方式提取光流获取运动特征,该网络具有良好的泛化能力,并在推理时间和计算速度等方面具有较高的效率。运动特征被广泛应用于动作识别[11]和人体行为识别[12]中,文献[13]在视觉跟踪算法中运用了深度运动特征,并取得了较好效果,但现有的跟踪算法大多数没有利用目标的运动特征,如何通过对目标的运动特征和空间表观特征的融合来提高算法的跟踪性能,是一个值得研究的工作。

此外,在跟踪过程中,对跟踪结果质量的判定可以有效减少模型累积错误导致的跟踪失败。常用的质量判别指标是根据目标响应图的变化来判断跟踪结果是否可靠,文献[14]提出了峰值旁瓣比(peak-to-sidelobe ratio, PSR),该方法通过响应图峰值尖锐程度来判定跟踪结果是否可靠。文献[15]在 LMCF 算法中提出平均相关峰能量(average peak to correlation energy, APCE)通过对响应图振荡的幅度来判断跟踪结果是否可靠。因此,可以依据相关的判别准则对跟踪结果的质量进行判定,以保证获得较好的跟踪结果。

综上所述,本文将在相关滤波跟踪框架的基础上,通过引入 RAFT 深度网络来估计光流,获取表征目标运动特征的时序信息,提取目标的 CN 特征和 HOG 特征,获取表征目标表观特征的空间信息,然后融合这两种信息以增强对目标时空特征的表征能力;其次,建立一种基于相似度的跟踪结果质量判别机制,根据跟踪结果实时调整时序信息在融合过程中的权重,以期在跟踪过程中更好地发挥不同特征的优势。在 OTB100[16]和 VOT2019[17]数据集上对所提出的算法进行了测试,实验结果表明,基于目标时序信息和空间信息自适应融合的跟踪算法有效提升了视觉跟踪算法的精度与成功率。

1 相关工作

为提升视觉跟踪算法的跟踪性能,本文选择了性能优异的 ECO 算法作为基准算法,将目标的运动特征引入到 ECO 算法中。ECO 算法分析了影响跟踪算法的 3 个重要因素:模型冗余、时间复杂度和模型更新策略。针对模型冗余提出了因式分解卷积,针对时间复杂度提出了样本生成空间模型,在模型更新中主要减少了模型更新次数。接下来介绍 ECO 算法中的因式分解卷积过程和样本生成空间模型。

1.1 因式分解的卷积操作

文献[17]因 C-COT 算法中巨额的计算量以及存在的部分冗余特征向量等问题,提出了基于卷积的因式分解操作。不同于 C-COT 算法中为 D 个特征学习独立滤波器,ECO 算法根据贡献量大小从 D 个滤波器中筛选出 f 个滤波器,ECO 算法根据贡献量大小从 D 个滤波器 $f^1, f^2, \cdots, f^D$ 中筛选了 C 个滤波器 $f^1, f^2, \cdots, f^c$,其中 $C<D$,对于特征层 d 的

滤波器，由 f^c 个学习到的系数集合 $p_{d,c}$ 构建线性组合 $\sum_c^c p_{d,c} \times f^c$。系数集合表示为 $D \times C$ 的矩阵 $\boldsymbol{P}=(p_{d,c})$。因式卷积的分解分为 2 个过程，首先目标特征 $J(x)$ 与矩阵 $\boldsymbol{P}_{D\times C}$ 相乘，然后特征图与滤波器进行卷积运算。其中训练样本定义为 x，因式分解的卷积公式表示如下：

$$S_{\boldsymbol{P}f}(x)=\boldsymbol{P}fJ(x)=\sum_{c,d} p_{d,c} f^c J d\{x^d\}=f\boldsymbol{P}^{\mathrm{T}} J\{x\} \tag{1}$$

式中：$\boldsymbol{P}^{\mathrm{T}}$ 为线性降维算子。傅里叶域中滤波器求解目标函数表示为：

$$\boldsymbol{E}(f,p)=\| \hat{\boldsymbol{Z}}^{\mathrm{T}}\boldsymbol{P}\hat{f}-\hat{y} \|_{i2}^2+\sum_{c=1}^{c} \| \hat{\boldsymbol{W}}\hat{f} \|_{i2}^2+\boldsymbol{\lambda} \| \boldsymbol{P} \|_{\boldsymbol{F}}^2 \tag{2}$$

式中：$\hat{\boldsymbol{Z}}$ 为插值特征图经过傅里叶变换后的表示，式(2)最后一项增加了 $\boldsymbol{P}$ 的 Frobenius 范数作为正则化，由控制权重系数。

1.2　样本生成空间模型

该模型通过样本分类来简化训练集。通过概率生成模型减小冗余样本集进而使样本更紧凑。将样本进行编组，类似的样本划为一组，每一个组都代表一种特定的场景。即组与组之间差异性大，而组内比较相似程度高，该方法极大丰富了训练样本集的多样性。

根据样本特征 x 和对应的期望输出分数 y 的联合概率分布 $p(x,y)$，可将式(2)进一步完善为：

$$E(f)=E\{\| S_f\{x\}-y \|_{L2}^2\}+\sum_d^{D2} \| wf^d \|_{L^2}^2 \tag{3}$$

式中：E 为联合概率分布 $p(x,y)$ 的期望。样本 x 的关联输出 y 的形状是预先定义好的，在这里为一个高斯函数。假设为目标位于图像区域中心点，因此所有 y 都一致。将样本分布简化为 $P(x,y)=p(x)\delta y_0(y)$ 只需计算 $p(x)$ 即可：

$$p(x)=\sum_{i=1}^{L} \pi_i N(x;\mu_i;I) \tag{4}$$

式中：L 表示样本组的个数，$i\in[1,L]$；样本组由 $N(x;\mu;I)$ 表示。将原训练样本替换为高斯模型均值 μ_1，使用高斯主成分先验权值 π_1 代替原更新系数。引入上述模型后滤波器求解目标函数最终可表示为：

$$E(f)=\sum_{i=1}^{L}\pi_i \| S_f\{\mu_i\}-y_0 \|_{L2}^2+\sum_{d=1}^{D} \| wf^d \|_{L2}^2 \tag{5}$$

2　本文算法

本文首先引入递归全对场变换深度网络(RAFT)估计光流，以提取时序信息，然后，利用目标的运动特征和表观特征的互补性，融合目标时序信息和空间信息，以增强目标时空特征的表征能力，同时，建立了一种基于相似度的跟踪结果质量判别机制，根据跟踪结果实时调整时序信息在融合过程中的权重，在特征融合过程中发挥不同特征的优势，以获得更好的跟踪性能。

2.1　RAFT 网络时序信息提取

RAFT 是一种性能优异的光流深层网络结构，如图 1 所示。RAFT 主要由三部分构成：特征编码器、4D 相关联层和基于门控循环单元[17](gate recurrent unit，GRU)的更新运算器。

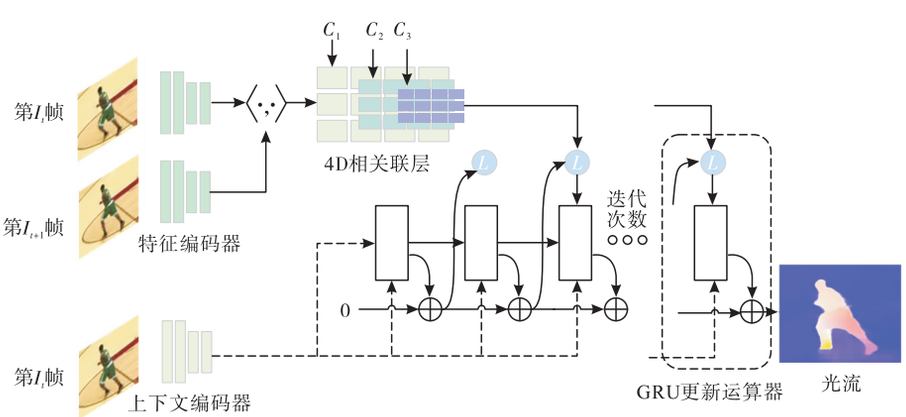


图 1　RAFT 网络结构

定义视频的总序列为 $t=\{1,2,\cdots,i\}$，其中 i 为视频总帧数，第 t 时刻图像为 I_t，选取 t、$t+1$ 时刻连续 2 帧 RGB 图像输入网络。首先通过特征编码器提取 I_t、I_{t+1} 的特征，对 I_{t+1} 的特征图进行平均池化，池化核大小分别为 $\{1,2,4,8\}$；其次将池化后的特征向量分别与 I_t 的特征向量进行内积运算，得到一个 4D 的多尺度相关联层 $c_k(k=1,2,3)$，图 1 中 C_k 的维度大小为 $H\times W\times H/2^{k-2}$；最后将相关联层特征向量和上下文编码器特征向量输入 GRU 更新运算器，GRU 更新运算器的输出为最终光流。RAFT 网络估计了一个密集位移场 (f^1,f^2)，该场将 I_{t+1} 中的每个像素 (u,v) 映射到对应的坐标

$(u',v')=(u+f^1(u),v+f^2(v))$，以描述每个像素在下一帧中的运动光流。其中光流场可以通过GRU更新运算器的迭代来更新。

在测试中，当直接利用RAFT网络计算的原始光流进行特征融合跟踪时，目标非常容易丢失，通过分析得知，ECO算法分别对CN特征和HOG特征进行了归一化处理，而原始光流没有进行归一化处理，使得3种特征融合时得到了错误的融合结果，造成跟踪失败。为此，我们对原始光流采用式(6)进行了归一化处理。

$$\boldsymbol{f}'=\frac{x_{ij}-\min(x_{ij})}{\max(x_{ij})-\min(x_{ij})} \tag{6}$$

式中：x_{ij} 为矩阵中对应的像素值；$\boldsymbol{f}'$ 为归一化处理后的光流特征向量，通过对原始光流的归一化处理，使特征融合结果能够很好地用于目标跟踪，显著提升了算法的跟踪性能。

RAFT网络中，光流提取速度约为10 FPS，不同的迭代次数对跟踪速度和成功率有重要的影响，在OTB100数据集实验结果如表1所示。从表中可以看出，迭代次数为4时跟踪效果最好，而当迭代次数超过4时，不仅影响跟踪器的实时性，同时会降低跟踪器的性能。因此，为平衡跟踪器性能，本文将RAFT网络迭代次数设置为4，并在后续实验中采用该值。

表1 迭代次数对精度、成功率和速度的影响

迭代次数	$i=2$	$i=4$	$i=6$	$i=8$	$i=10$
精度	0.846	0.873	0.861	0.860	0.856
成功率	0.635	0.654	0.649	0.643	0.642
FPS	6.910	6.830	6.810	6.790	6.680

2.2 跟踪结果质量判定

目标跟踪过程中，目标外观和背景随时间的推移不断发生变化，在目标遮挡或者外观出现剧烈变化时容易造成模型污染，从而导致跟踪失败，因此对跟踪结果的质量判定非常重要。这不仅会影响模型更新的策略，也会影响不同类别特征在融合过程中所占的比重。

APCE是根据响应图震荡程度来反映跟踪结果的指标，该指标定义如式(7)所示，当响应图峰值起伏越小则该值越大，说明跟踪结果可靠，反之当出现跟踪失败时该值会急剧下降。

$$APCE=\frac{|R_{\max}-R_{\min}|}{\mathrm{mean}(\sum_{ij}(R_{ij}-R_{\min})} \tag{7}$$

当APCE突然减小时，可能发生目标被遮挡，或者目标丢失的情况，但当目标在快速运动、运动模糊或周围有干扰物等情况下，该值也会突然减小，而大多这种情况下跟踪结果却是可靠的，所以这时该值会对跟踪结果的质量造成误判。

对此，我们采用目标模板和候选模板之间的相似度来判断跟踪结果的可靠程度，该指标定义如下：

$$\delta=\frac{A_{\mathrm{obj}}B_{\mathrm{cdt}}}{\|A_{\mathrm{obj}}\|\cdot\|B_{\mathrm{cdt}}\|} \tag{8}$$

式中：A_{obj} 为目标模板，选取图像的初始模板为目标模板；B_{cdt} 为候选模板，每一帧的跟踪结果为候选模板；$\|\cdot\|$ 为计算矩阵二范式范围为[0,1]。最后根据调整运动特征在融合过程中的权重。

2.3 特征融合

当目标处于复杂动态环境下，仅用空间特征表征目标的特征，其表征能力十分有限。针对不同特征之间的互补性，本文采用时序特征和空间特征的自适应融合方式对目标建模。采用的3种不同特征分别为：HOG、CN和光流。首先从训练样本 X 中提取不同类型的特征 f_j，提取目标的CN特征和HOG特征表征目标空间信息，将 I_t、I_{t+1} 时刻的相邻帧输入RAFT网络中提取光流，利用目标的运动特征和表观特征的互补性，融合目标时序信息和空间信息，增强目标时空信息的表征能力。

光流可以提取目标的运动信息，在复杂动态背景下相对于空间特征更有利于定位，但光流对周边相似运动干扰物十分敏感，这种情况下，如果光流设置权值过大会造成模型漂移。针对不同场景下光流在特征融合过程中体现的重要程度不同，本文采用式(8)对跟踪结果质量的判定，来调整光流在跟踪过程中所占的比重，对光流 f_{flow} 乘以系数 w。定义 w 为光流的权值，取值为：

$$w=\begin{cases}1+\delta,\delta<\alpha\\1-\delta,\delta\geqslant\beta\end{cases} \tag{9}$$

式中：δ 为目标模板和候选模板之间的相似度，α 取值为0.3，β 取值为0.7，当 $\delta<0.3$ 时，跟踪过程中可能发生遮挡、形变等情况，此时增加光流权值，可以更好地提取目标的运动信息，更有利于对目标进行定位；$\delta\geqslant0.7$ 时，为缓解目标周边相似运动干扰物造成的模型漂移，则减小光流的权值。在0.3～0.7时，光流权重为1。

相关滤波器分别与3种特征的特征图进行相关计算得到响应图，为保证响应图的质量，根据响应值大小对响应图通道进行筛选，取3个特征图响应值最高的前 k 个通道进行add融合，最后根据融合后的响应图 Z 估计新的一帧中的目标。特征响应值最高的前 k 个通道的计算过程如式(10)所示：

$$Z=\{Z_P\,|\,p=\arg\max(f^k_{\mathrm{cn}},f^k_{\mathrm{hog}},f^k_{\mathrm{flow}})\} \tag{10}$$

式中：P 为响应值；f_{cn}、f_{hog}、f_{flow} 分别为CN特征、

HOG 特征和光流特征对应的响应图。

本文在 OTB100 数据集测试了 $k(k=1,2,3)$ 对成功率和精度的影响，结果如表 2 所示，实验表明本文算法在 $k=2$ 时，在融合过程中更好地发挥了不同特征优势，得到较好的成功率和精度。

表 2　参数对成功率和精度的影响

通道数	$k=1$	$k=2$	$k=3$
精度	0.865	0.873	0.859
成功率	0.647	0.654	0.641

2.4　算法流程

图 2 为本文算法流程图。首先提取目标的 CN 特征和 HOG 特征，利用 RAFT 网络提取连续两帧图像的光流，然后分别训练 HOG 特征、CN 特征和光流的相关滤波器，通过相关滤波计算得到各自的响应图；根据跟踪结果的可靠程度来计算光流的权重，将光流乘以权重对光流特征进行调整，图 2 中绿色的线条模块为光流权重调整模块；最后，对 3 种特征响应图筛选后进行特征融合，根据融合后的响应图计算目标最终的位置。

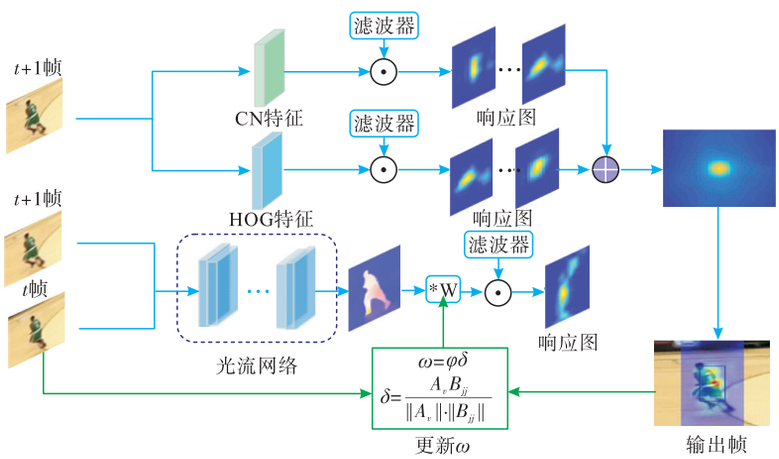


图 2　算法流程图

3　实验

为验证本文算法的有效性，在 Ubuntu 操作系统下，使用 Python 语言编程实现本文算法，在 Intel i5-8400 CPU@2.80 GHz 处理器上进行测试，并利用 GPU(NVIDIA GTX 1080Ti)进行加速。实验部分参数设置为：目标尺度搜索个数 $K=5$，学习率为 $P=0.009$，尺度缩放参数初始化为 $R=4$。为验证算法的实际效果，本文选择 OTB100[18]、VOT2019[19] 这 2 个经典数据平台进行评估。选择 ECO-HC[9] 等经典算法[20-26] 作为参照算法。

3.1　OTB100 实验

3.1.1　消融实验

表 3 为本文在增加不同模块下在 OTB100 数据上的实验结果的比较。实验性能评估指标采用 OTB100 数据集的评价指标，分别为精度和成功率，其中精度衡量算法对目标中心点位置的估计能力。成功率衡量算法对目标尺度大小的估计能力。从表中可以看出，在单独 CN 特征和 HOG 特征下的跟踪精度分别为 0.796 和 0.804，成功率分别为 0.618 和 0.610。在对 3 种特征融合后精度达到 0.873 和 0.642，证明了跟踪过程中运动特征和空间特征互补融合的重要性。与基准算法相比，融合光流特征后精度提升 1.2%，成功率提升 0.8%；在对光流权重根据式(9)进行更新后，算法精度提升 2.4%，成功率提升 1.6%。

表 3　本文算法在 OTB100 上的消融实验结果

CN	HOG	Flow	权重更新	精度	成功率
√				0.796	0.618
	√			0.804	0.610
√	√			0.849	0.638
√	√	√		0.861	0.646
√	√	√	APCE	0.866	0.648
√	√	√	Our	**0.873**	**0.654**

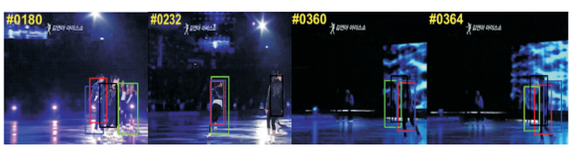
3.1.2　定性分析

图 3(a)为 Soccer 视频序列，该视频序列背景复杂，目标在运动过程中的多次出现被大量相似干扰物遮挡的情况，目标在第 109 帧中目标几乎不可见，Staple 和 ECO-HC 等算法无法准确跟踪，本文算法可以较好的跟踪且当目标刚从被遮挡区域恢复出来时仍能跟上目标。图 3(b)为 Dragon baby 视频序列，该视频中的目标和小恐龙对打，目标姿态变化丰富。目标快速运动过程中极易发生跟踪漂移，其他 3 种算法在跟踪过程中均产生了不同程度的误差，

本文算法有效缓解了因目标运动过快而造成的跟踪失败问题。图 3(c)为Skating 视频序列跟踪过程中背景光照发生变化,本文算法在低分辨率下能够很好跟踪目标。图 3(d)为 Biker 序列,在跟踪过程中目标发生旋转,要求算法具有一定旋转不变性。部分算法出现跟踪漂移,本文算法和 ECO-HC 算法能够较好地跟踪目标。

(a) Soccer

(b) Dragon bady



(c) Skating

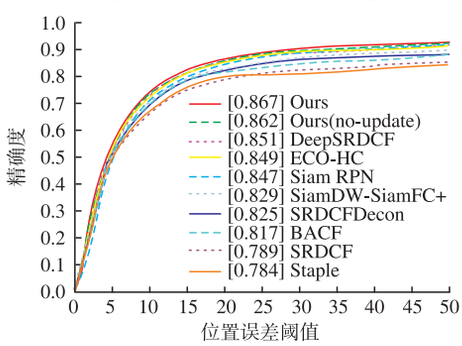
(d) Biker

图 3 视觉跟踪实验结果

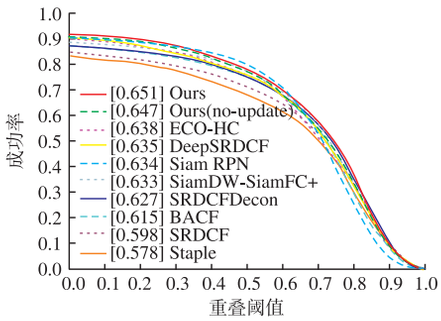
3.1.3 定量分析

1)整体性能

图 4 为本文算法与其他 8 种经典参照算法在 OTB100 数据集上评估结果。从图中可以看出,本文算法通过时序信息和空间信息的自适应融合,充分发挥了不同特征间判别力,相对 ECO-HC 算法,本文算法精度提升了 2.4%,成功率提升了 1.6%。



(a)精确度曲线



(b) 成功率曲线

图 4 不同算法精度曲线和成功率曲线

2)属性分析

在 OTB100 数据集测试了不同属性下的成功率曲线下面积(area under curve, AUC)并绘制了雷达图,如图 5 所示,运动类视频往往会出现图像低分辨,目标重影等情况。在该类视频中能够提取单帧图像的空间特征十分有限,本文算法充分利用了视频序列中帧与帧之间的时序信息。相对 ECO-HC 算法,在运动模糊视频中 AUC 提升 3.6%;快速运动视频中 AUC 提升 1.4%。在运动类视频序列中有明显优势。表明在应对复杂场景时本文算法具有很好的鲁棒性。

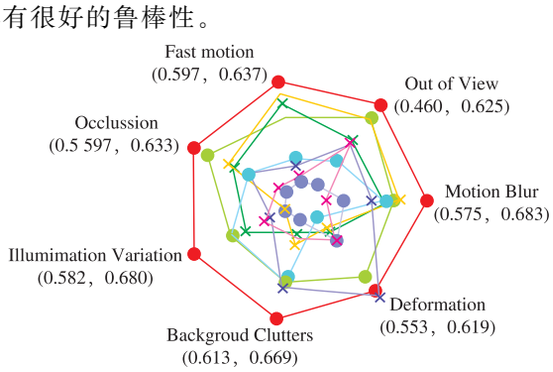


图 5 视觉跟踪实验结果

3.2 VOT2019 实验

为进一步评估本文算法在视觉跟踪中的有效性,本文在 VOT2019 数据集进行测试,并与其他 7 种经典的相关滤波类算法[27-29]进行比较,比较了 EAO(expected average overlap, EAO)、精度和鲁棒性 3 个指标,结果如表 4 所示。其中最优结果加黑表示,次优结果加下划线表示,第 3 优结果加虚下划线表示。相对 ECO-HC 算法,本文算法在 VOT2019 数据集中 EAO 提升了 1.7%,该结果进一步说明了基于时空信息和空间信息自适应融合算法在视觉跟踪中的有效性。

表 4 VOT2019 算法实验结果比较

Trackers	KCF	CISRDCF	BACF	ARCF	STRCF	ECOHC	ASRCF	Ours
EAO	0.110	**0.153**	0.116	0.135	0.114	0.132	0.145	0.149
精度	0.425	0.415	0.445	0.467	0.452	0.492	0.465	**0.493**
鲁棒性	1.279	0.632	0.657	**0.527**	0.700	0.868	0.700	0.788

4　结语

本文提出一种时序信息和空间信息自适应融合的视觉跟踪算法。利用目标的运动特征和表观特征的互补性，融合目标时序信息和空间信息，增强了目标时空特征的表征能力；其次，建立一种基于相似度的跟踪结果质量判别机制，根据跟踪结果实时调整运动特征在融合过程中的权重，以适应复杂动态场景下目标的变化。结果表明，本文算法有效提升了跟踪性能，在一些具有挑战性的视频中跟踪效果良好。但是，光流在提取过程中计算量较大，在提升跟踪性能的前提下，如何减少光流在跟踪过程中的计算量将是本文未来的工作。

参考文献

[1] LUO Y, YIN DANG WANG A. Pedestrian Tracking in Surveillance Video Based on Modified CNN[J]. Multimedia Tools and Applications, 2018, 77(18): 24041-24058.

[2] NIKOLOVA I. Performance Analysis of Robust Image Features Detection Algorithms[J]. Information Technologies & Control, 2014, 11(3): 2-15.

[3] 郭静静，侯志强，陈立琳，等. 一种长宽比自适应变化的目标尺度估计算法[J]. 空军工程大学学报：自然科学版，2021，22(1)：77-84.

[4] 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, 2014, 37(3): 583-596.

[5] DANELLJAN M, HAGER G, KHAN F S, et al. Accurate Scale Estimation for Robust Visual Tracking[C]//British Machine Vision Conference. Nottingham, UK: BMVA Press, 2014.

[6] 金泽芬芬，侯志强，余旺盛，等. 多特征博弈的目标跟踪算法[J]. 空军工程大学学报：自然科学版，2017，18(1)：50-56.

[7] MA C, HUANG J B, YANG X. Hierarchical Convolutional Features for Visual Tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 3074-3082.

[8] BHAT G, JOHNANDER J, DANELLJAN M, et al. Unveiling the Power of Deep Tracking[C]//proceedings of the European Conference on computer Vision. 2018: 483-498.

[9] DANELLJAN M, BHAT G, KHAN F S. ECO: Efficient Convolution Operators for Tracking[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6931-6939.

[10] TEED Z, DENG J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow[C]//European Conference on Computer Vision. Cham: Springer, 2020: 402-419.

[11] GKIOXARI G, MALIK J. Finding Action Tubes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2015: 759-768.

[12] SIMONYAN K, ZISSERMAN A. Two-Stream Convolutional Networks for Action Recognition in Videos[J]. Advances in Neural Information Processing Systems, 2014, 27: 1-9.

[13] GLADH S, DANELLJAN M, KHAN F S. Deep Motion Features for Visual Tracking[C]//2016 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE, 2016: 1243-1248.

[14] BOLME D S, BEVERIDGE J R, DRAPER B A. Visual Object Tracking Using Adaptive Correlation Filters[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: [S. l.]: IEEE, 2010: 2544-2550.

[15] 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. Honolulu, HI, USA: IEEE, 2017: 4021-4029.

[16] DANELLIAN M, ROBINSON A, KHAN F S. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking[C]//European Conference on Computer Vision. Cham: Springer, 2016: 472-488.

[17] CHO K, VAN M B, BAHDANAU D. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches[EB/OL]. ArXiv Preprint ArXiv: 1409.1259, 2014.

[18] WU Y, LIM J, YANG M H. Object Tracking Benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

[19] KRISTAN M, MATAS J, LEONARDIS A. The Seventh Visual Object Tracking vot2019 Challenge Results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul, Korea(South): IEEE, 2019.

[20] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary Learners for Real-Time Tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 1401-1409.

[21] 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. Long Beach, CA, USA: IEEE, 2019: 4591-4600.

(下转第 106 页)

[16] CEB. Concrete Structures Under Impact and Impulsive Loading[R]. CEB Bulletind' Information, vol. 187. Lausanne, France. Committee Euro-International du Beton, 1998.

[17] TEDESCO J W, ROSS C A. Strain-Rate-Dependent Constitutive Equations for Concrete[J]. Journal of Pressure Vessel Technology, 1998, 120(4): 398-405.

[18] GROTE D L, PARK S W, ZHOU M. Dynamic Behavior of Concrete at High Strain Rates and Pressures: I. Experimental Characterization[J]. International Journal of Impact Engineering, 2001, 25(9):869-886.

[19] 黄仕超,彭刚,邹三兵,等. 不同龄期混凝土动态力学性能研究[J]. 长江科学院院报, 2015, 32(12): 129-133, 143.

[20] 许金余,李为民,范飞林,等. 地质聚合物混凝土的冲击力学性能研究[J]. 振动与冲击, 2009, 28(1): 46-50, 194.

[21] 王世鸣. 冲击荷载下早龄期混凝土力学和损伤特性的试验研究[D]. 长沙: 中南大学, 2014.

[22] 刘鹏,余志武,陈令坤. 养护龄期对水泥混凝土性能和微观结构的影响[J]. 建筑材料学报,2012,15(5): 717-723.

[23] 杨立荣,王春梅,封孝信,等. 粉煤灰/矿渣基地聚合物的制备及固化机理研究[J]. 武汉理工大学学报, 2009, 31(7): 115-119.

[24] 黄华,郭梦雪,张伟,等. 粉煤灰-矿渣基地聚物混凝土力学性能与微观结构[J]. 哈尔滨工业大学学报, 2022,54(3):74-84.

[25] BISCHOFF P H, PERRY S H. Compressive Behaviour of Concrete at High Strain Rates[J]. Materials and Structures, 1991,24(6):425-450.

(编辑:杜娟)

(上接第 63 页)

[22] DANELLJAN M, HAGER G, SHAHBAZ K F. Learning Spatially Regularized Correlation Filters for Visual Tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 4310-4318.

[23] KIANI G H, FAGG A, LUCEY S. Learning Background - Aware Correlation Filters for Visual Tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE,2017: 1135-1143.

[24] DANELLJAN M, HAGER G, SHAHBAZ K F. Adaptive Decontamination of the Training Set: A Unified Formulation for Discriminative Visual Tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA :IEEE,2016: 1430-1438.

[25] LI B, YAN J, WU W. High Performance Visual Tracking with Siamese Region Proposal Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA :IEEE,2018: 8971-8980.

[26] DANELLJAN M, HAGER G, SHAHBAZ K F. Convolutional Features for Correlation Filter Based Visual Tracking[C]//Proceedings of the IEEE International Conference on computer Vision Workshops. [S. l.]:IEEE, 2015: 58-66.

[27] DAI K, WANG D, LU H. Visual Tracking via Adaptive Spatially - Regularized Correlation Filters[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA :IEEE,2019: 4670-4679.

[28] LI F, TIAN C ,ZUO W. Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA:IEEE, 2018: 4904-4913.

[29] HUANG Z, FU C, LI Y. Learning Aberrance Repressed Correlation Filters for Real-Time UAV Tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South) :IEEE,2019: 2891-2900.

(编辑:徐楠楠)