高技术通讯 2020 年 第30卷 第10期:1085 ~ 1092

doi:10.3772/j.issn.1002-0470.2020.10.011

结合卷积神经网络多特征融合的相关滤波跟踪①

杨海清② 许倩倩 唐怡豪 孙道洋

(浙江工业大学信息工程学院 杭州 310023)

摘 要 针对相关滤波跟踪中的多特征融合问题,本文提出了基于多通道相关滤波框架结合卷积神经网络(CNN)多特征融合的跟踪算法。首先引入梯度直方图和颜色名特征,利用传统的特征提取方法将提取的特征进行简单的矢量相加。然后采用在 ImageNet 上训练的卷积神经网络进行特征提取,使用 conv5-4 卷积层的输出作为特征,再分别训练各自的相关滤波器,对特征响应进行可靠性加权求和获得目标位置。最后,通过最大响应值和平均峰值相关能量的变化来判断是否更新模型。在标准测试集(OTB-100)上进行实验测试,与5种基于相关滤波的主流算法进行性能对比。实验结果表明,本文算法在光照变化、尺度变化及遮挡等复杂情况下的鲁棒性和跟踪精度都优于其他算法。

关键词 目标跟踪;卷积神经网络(CNN);相关滤波;特征提取;可靠性加权

0 引 言

视觉跟踪是计算机视觉中众多应用的基本问题之一[1],其典型场景是跟踪由第1帧中的边界框指定的未知目标对象。视觉跟踪在近几十年来取得了重大进展,但由于遮挡、变形、突然运动、光照变化和背景杂乱等引起的外观变化,对跟踪精度仍具有挑战性。

目前,基于相关滤波器的目标跟踪因使用快速傅里叶变换具有高效计算而引起了广泛的关注,其思想是将所有输入特征的循环版本回归到目标高斯函数,因此不需要目标外观的硬阈值样本。Bolme 等人[2]采用在亮度通道上的平方误差滤波器的最小输出和以进行快速跟踪,已经提出了几个扩展来提高跟踪精度,包括核化相关滤波器[3]、多维特征[4]、上下文学习[5]、尺度估计[6]、基于多特征融合的尺度自适应跟踪[7]和互补特征学习实时跟踪[8]。最近,基于卷积神经网络(convolutional neural network, CNN)[9]的特征在视觉跟踪中体现出很好的效果[10]。Wang 等人[11]提出要在视频库[12]学习一个双层神经网络。Hong 等人[13]在目标对象的不同实例上构造多个 CNN 分类器,以在模型更新期间排除噪声,从二进制样本中学习2层 CNN 分类器,不需要预训练过程。基于卷积特征的相关滤波算法(hierarchical convolutional features for visual tracking, HCF)[14]利用 CNN 中多个卷积层来提取目标特征,将高层高语义特征与低层高分辨特征进行有效的结合,提高了跟踪精度。

但由于提取目标的特征不强,会使目标发生丢失、漂移。针对相关滤波不同特征的提取和融合的问题,本文提出学习多维特征的相关滤波器[15],使用卷积特征与传统手工特征中的方向梯度直方图(histogram of oriented gridients, HOG)[16]和颜色名(color name, CN)[17]特征的结合,并且构造多个相关滤波器,区别于单一滤波器,利用传统的特征提取方法,将提取的特征(HOG + CN)进行简单的矢量相加,使用文献[18]中的方法提取 CNN 特征,在大型

① 浙江省科技计划(2017C37054)资助项目。

② 男,1971 年生,博士,副教授;研究方向:计算机视觉及应用;联系人,E-mail:yanghq@zjut.edu.cn

(收稿日期:2019-10-23)

ImageNet 数据集[19]上使用类别级标签对其进行训练,采用 conv5-4 卷积层的输出作为特征。在训练阶段,根据特征响应值计算可靠权重;在定位阶段,对特征检测响应值进行可靠加权后得到目标的位置;最后通过最大响应值和平均峰值相关能量(average peak-to correlation energy, APCE)[20]判断是否更新模型。采用跟踪基准(object tracking benchmark, OTB-100)[21]对多个视频序列进行测试,并与 5 种基于相关滤波主流跟踪算法作性能对比分析,实验结果充分证明本文算法在鲁棒性和跟踪精度上均优于其他算法。

1 多通道相关滤波器

学习阶段:记 d 通道目标外观模板为 f,期望输出为 y,然后通过求解最小化问题来学习与 f 具有相同大小的相关滤波器 h:

$$h* = \| r(h) - y \|^2 + \lambda \sum_{d=1}^{D} \| h \|^2 \quad (1)$$

$$r(h) = \sum_{d=1}^{D} f_d * h_d \quad (2)$$

其中,$r(h)$ 表示训练样本相关响应,$*$ 是空间域的循环卷积,λ 是正则化参数($\lambda \geqslant 0$)。式(1)中的最小化问题类似于训练[22]中的矢量相关滤波器,并且可以使用快速傅里叶变换在每个单独的特征信道中求解,d 通道上的频域学习滤波器可写为

$$H^d = \frac{Y \odot \bar{F}^d}{\sum_{i=1}^{D} F^i \odot F^i + \lambda} \quad (3)$$

式中,Y 是 y 傅里叶变换形式,$\bar{X}$ 表示复共轭,操作符$\odot$是点乘。

检测阶段:

$$r_h(x) = \max(F^{-1}(\sum_{d=1}^{D} H_d \odot \bar{X}^d)) \quad (4)$$

式中,$\bar{X}^d$ 表示待检测图像块提取的 d 通道特征图的离散傅里叶变换,$F^{-1}()$ 表示逆傅里叶变换。通过求解式(4)中最大响应值的位置确定下一帧跟踪目标的位置。

在目标跟踪过程中,目标的外观会发生变化,为了能持续跟踪目标,需要在线更新滤波器,在第 t 帧图像上进行目标跟踪时,定义学习到的目标模型图像为 $\bar{X}^t$、相关滤波器 h 模型的更新公式为

$$\begin{cases} \bar{X}^t = (1-\eta)\bar{X}^{t-1} + \eta \bar{X}^t \\ H_t^d = (1-\eta)H_t^{d-1} + \eta H^d \end{cases} \quad (5)$$

式中,η 是学习率。

2 多特征融合与目标跟踪

本文算法基于多通道相关滤波算法框架,采用卷积特征、传统手工特征(HOG + CN)多特征融合进行目标跟踪,用多个滤波器学习并进行卷积以生成各自响应图。

2.1 特征可靠性估计

假设特征通道相互独立,响应是所有特征通道之和,跟踪中加入通道可靠性权重 w_k,最终响应由加权后的特征通道和计算:

$$y(h) = \sum_{k=1}^{K} f_k * h_k \cdot w_k \quad (6)$$

其中,$f_k * h_k$ 表示特征通道响应,w_k 表示对应权重,w_k 由通道学习可靠权重和通道检测可靠权重计算,通道学习可靠权重在滤波器学习阶段由通道滤波器最大响应计算。

$$w_k^{lea} = \max(r_k) \quad (7)$$

通道检测可靠权重在检测阶段由响应图中 2 个最大峰值比值计算。当相似物体出现在目标附近时,会出现多峰,在这种情况下将比率约束为 0.5。

$$w_k^{det} = 1 - \min\left(\frac{\rho_{\max 2}}{\rho_{\max 1}}, \frac{1}{2}\right) \quad (8)$$

其中,$\rho_{\max 1}$、$\rho_{\max 2}$ 表示响应图中 2 个最大的峰值。由此可得通道可靠权重:

$$w_k = w_k^{lea} w_k^{dec} \quad (9)$$

其中 $\sum_k w_k = 1$,即所有通道系数和为 1,为了保证时间的鲁棒性,通道权重更新公式如下:

$$\boldsymbol{w}^t = (1-\eta)\boldsymbol{w}^{t-1} + \boldsymbol{w} \quad (10)$$

其中 $\boldsymbol{w} = [w_1, \cdots, w_K]^{\mathrm{T}}$。

2.2 遮挡机制的判断

响应图的峰值和波动在一定程度上反映了跟踪结果的置信度。当检测到的目标与实际目标极为匹配时,理想的响应图只有一个尖峰,所在其他区域平滑,相关峰越尖锐,定位精度越好。否则,整个响应

图将剧烈波动,如果继续使用不确定的样本来更新跟踪模型,它将大部分损坏,导致跟踪失败。因此,本文提出了具有2个标准的高置信度反馈机制。

第1个标准是响应图 $F(s, y; w)$ 的最大响应分数 $F_{\max 1}$,本文采用的是 HOG + CN 特征融合后的响应分数,定义为

$$F_{\max 1} = \max F(s, y; w) \tag{11}$$

第2个标准称为平均峰值相关能量(APCE)测量,定义为

$$APCE = \frac{|F_{\max} - F_{\min}|^2}{\text{mean}(\sum_{m,n}(F_{m,n} - F_{\min})^2)} \tag{12}$$

其中,$F_{\max}$ 和 $F_{\min}$ 表示3种特征融合后得到的最大响应值和最小响应值,m 和 n 代表响应图的宽和高。*APCE* 表示响应图的波动程度和检测目标的置信水平。目标明显出现在检测范围内,峰值越尖锐、噪声越小,*APCE* 值将相对变大,即响应图只有一个尖锐的峰值且其他处呈现平滑的状态。当对象被遮挡或丢失时,*APCE* 值将显著减小。

当 *APCE* 和 $F_{\max 1}$ 均低于一定阈值时,即发生遮挡,模型会停止更新,既避免原模型的污染又降低了模型的更新次数,在某种程度上加速了算法。若目标发生连续多帧遮挡时,再考虑初始化该跟踪算法。

2.3 多特征响应自适应融合与相关滤波器跟踪

本文使用来自 CNN(VGG-Net-19)的卷积特征映射来编码目标外观,与 CNN 前向传播一起,不同类别的对象语义区分被加强,并且精确定位的空间分辨率逐渐减小。对于视觉目标跟踪,目的是找出目标对象的准确位置,卷积神经网络高层的卷积特征更加抽象,具有丰富的语义信息,能够很好地解决非刚性形变、遮挡等问题,并且能够对目标进行类间判别,对于目标的外形变化是鲁棒的。不足的是空间分辨率低,对平移和尺度都有不变性,无法精确定位目标,会造成目标漂移和跟踪失败。

HOG 和 CN 特征分别描述了目标的梯度和颜色特征,在图像的每个局部区域内,通过计算梯度方向直方图来提取 HOG 特征,其描述了目标的边缘梯度信息。在目标的每个像素上进行非线性映射来提取 CN 特征,相比灰度特征,其能描述更丰富的目标颜色信息,HOG 特征保留了目标的位置信息,CN 保留了颜色的位置信息。虽然 HOG 特征具有一定的平移、光照不变性,但这种单一的特征难以适应跟踪中出现的多种挑战因素,而 CN 特征具有对图像大小和方向不敏感的特点,所以将 HOG 和 CN 2种特征进行融合,然后对目标进行描述,可实现优势互补,提高了分类器的性能,其缺点是不变性差。目标稍微形变就很难识别目标,尤其是旋转,即鲁棒性很差。

因此,本文提出将高分辨率、低鲁棒性的传统特征(HOG + CN)与高语义、高鲁棒性、低分辨率的卷积特征相结合的算法,以达到跟踪优势互补的效果。用传统特征提取方法提取 HOG + CN 特征融合再通过相关滤波器学习得到相关响应 $y_{\text{HOG+CN}}$,与此同时对图片块用训练好的卷积神经网络(VGG-Net-19)提取特征。为保证精确的跟踪文中删除了空间分辨率较低的完全连接层,使用 VGG-Net-19 最后一层卷积层(conv5-4)特征,再用一个核相关滤波器学习得到相关响应 y_{CNN};在训练阶段,根据式(9)特征响应值计算可靠权重 w_k;在模板特征响应图层进行自适应特征融合,由式(6)得到融合后的输出响应:

$$y_t = w_1 \times y_{\text{HOG+CN}} + w_2 \times y_{\text{CNN}} \tag{13}$$

其中 w_1 和 w_2 由式(10)更新且 $w_1 + w_2 = 1$;通过 y_t 的峰值得到目标位置,最后再根据响应图的最大响应分数 $F_{\max 1}$ 和 *APCE* 是否低于一定的阈值判断遮挡从而分别更新各自的滤波器。

2.4 算法流程

算法流程如图1所示,算法步骤如下。

(1)在第 t 帧的目标估计位置处分别提取 CNN、HOG 和 CN 特征,将提取的 HOG 和 CN 特征融合进行简单的矢量相加,调整融合后特征的尺寸使其与 CNN 卷积特征尺寸相同,训练滤波器模型 $H_{\text{HOG+CN}}$ 和 H_{CNN},通过式(5)更新模型,由式(7)计算学习权重 $w_k^{(lea)}$(遮挡不更新),使用余弦窗(cosine window)消除边界响应。

(2)在 $t+1$ 帧的目标估计位置处提取 HOG + CN 和 CNN 特征位置侯选样本,将训练得到滤波器模型 $H_{\text{HOG+CN}}$、H_{CNN},分别学习并进行卷积以生成各自响应值,记为 $y_{\text{HOG+CN}}$、y_{CNN},由式(8)计算检测权重 $w_k^{(dec)}$。

(3)将通道检测响应值 y_{HOG+CN} 和 y_{CNN} 进行权重通道可靠性加权融合;由式(6)得到响应值 y_t,其中 w_1 和 w_2 由通道可靠权重 $w_k = w_k^{(lea)} w_k^{(dec)}$ 得到,通道可靠权重由式(10)更新。

(4)根据通道可靠加权后得到的响应值 y_t 估计出目标的位置,并根据融合特征(HOG + CN)后的响应图中最大响应分数 F_{max1} 和 $APCE$ 值是否低于一定的阈值来判断遮挡,从而通过式(5)和式(10)更新下一帧的滤波器模型以及通道可靠性权重(遮挡不更新模型)。

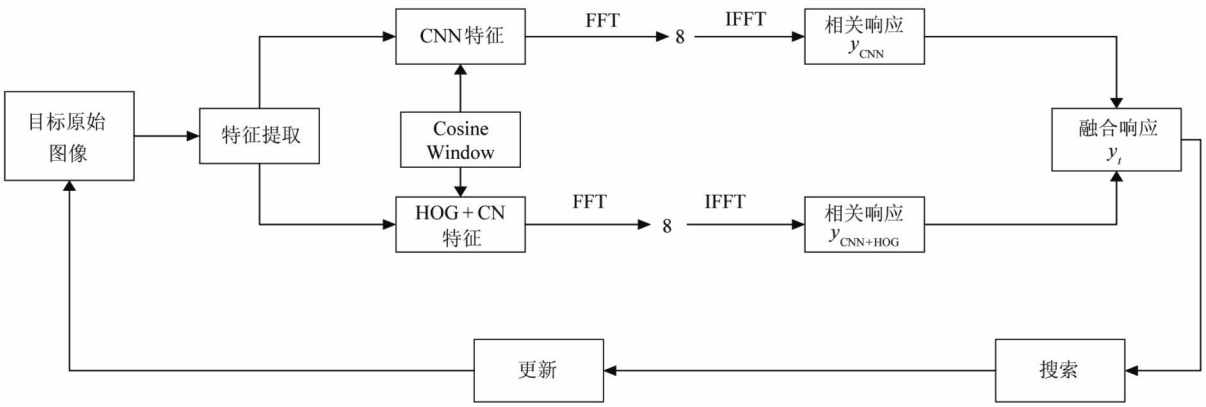

图 1　算法流程

3　实验结果分析

实验环境基于 Window7 系统和 Matlab 2016b,使用目标跟踪基准(OTB-100)部分视频序列评估了本文算法的性能,其中包含的复杂情况有光照变化、平面内旋转、快速移动、运动模糊、背景干扰、遮挡、尺度变化和非刚性形变,并与最先进的算法进行了比较。主要使用距离精度(distance precision,DP)、重叠精度(overlap precision,OP)、平均帧率(frame per second,FPS)3 种指标进行评价,其中 DP 描述了跟踪算法估计目标位置(bounding box)的中心点与人工标注(ground-truth)目标的中心点,这两者的距离小于给定阈值的视频帧的百分比,阈值设定为 20 个像素点,评价了算法的鲁棒性;OP 指得分大于某一阈值的帧数占跟踪总帧数的百分比,评价了算法的准确性。根据 PASCAL 评价指标[23],本文选择重叠率阈值为 0.5,遵循文献[24]中的协议,对所有视频序列和所有灵敏度分析使用相同的参数值,在 Matlab 中的 Intel i5-4770 上实现了跟踪器。

3.1　定性分析

测试的算法除了本文算法以外,还增加了 KCF[4]、CN[17]、SAMF[7]、CNN + SVM[13] 和 HCF[14] 等近几年提出的主流相关滤波跟踪算法,为了相对直观地证明本文算法的优越性,选用 3 组视频序列 girl2、coke、singer1 在 6 种跟踪算法中的比较,如图 2、图 3、图 4 所示,图中序列分别测试了遮挡、快速移动、光照变化以及尺度变化 4 种情况。图片从左到右算法依次为本文算法(OUR)、CN、KCF、HCF、SAMF、CNN + SVM,图2中序列118帧目标发

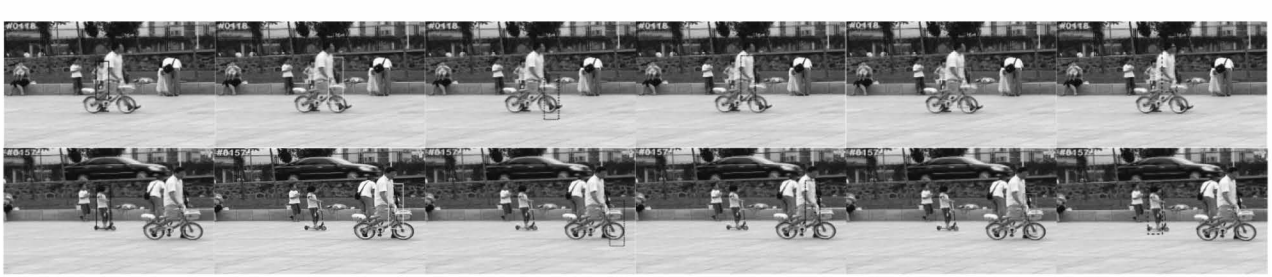

图 2　遮挡情况下 6 种跟踪算法对比

图3　快速运动与光照变化情况下6种算法对比

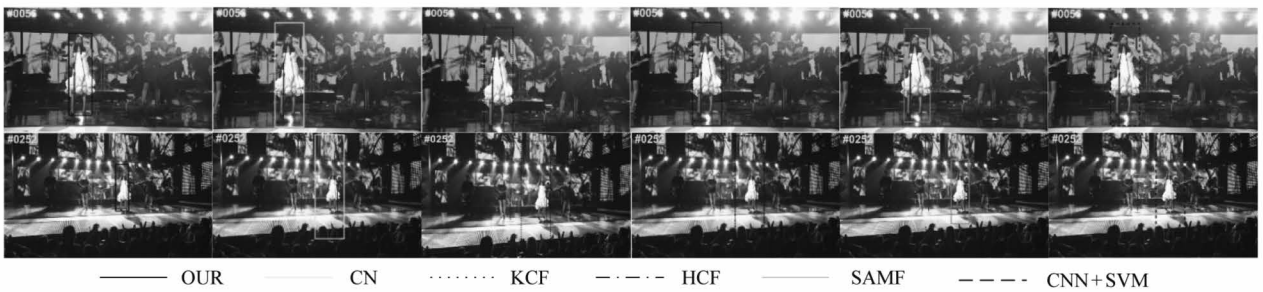

图4　尺度变化情况下6种算法对比

生遮挡,其他算法跟踪误差逐渐积累,从157帧中可以看出其他算法均跟丢目标;从图3可以看出,由于目标快速移动和在强烈光照影响下,其他算法表现比较差;图4中序列目标发生了尺度变化,其他算法不能很好适应尺度变化,会出现漂移的现象。而本文算法对目标的位置和目标尺度变化均做出了很好的处理,表现出很高的鲁棒性,在整个跟踪过程中能适应各类复杂场景并且能准确跟踪目标。

3.2　定量分析

本实验中,为了分析多特征融合方法的有效性,将OUR _ OCC(HOG + CN + CNN)、OUR _ OCC _ cnn(CNN)和OUR _ OCC _ hogcn(HOG + CN)3种算法在OTB-100的部分视频序列上做了实验,主要从一次性通过评估(one-pass evaluation,OPE)方面进行测评。从表1和图5可以看出,OUR _ OCC的精度和成功率都高于传统特征算法和卷积特征的算法,精度分别提升了9%和2.8%,成功率分别提升了5.9%和2.9%。由图6可以看出在快速移动和

表1　融合特征前后3种算法性能对比表

评价指标	本文算法	传统特征	卷积特征
精度	90.4%	81.4%	87.6%
成功率	84.7%	78.8%	81.8%

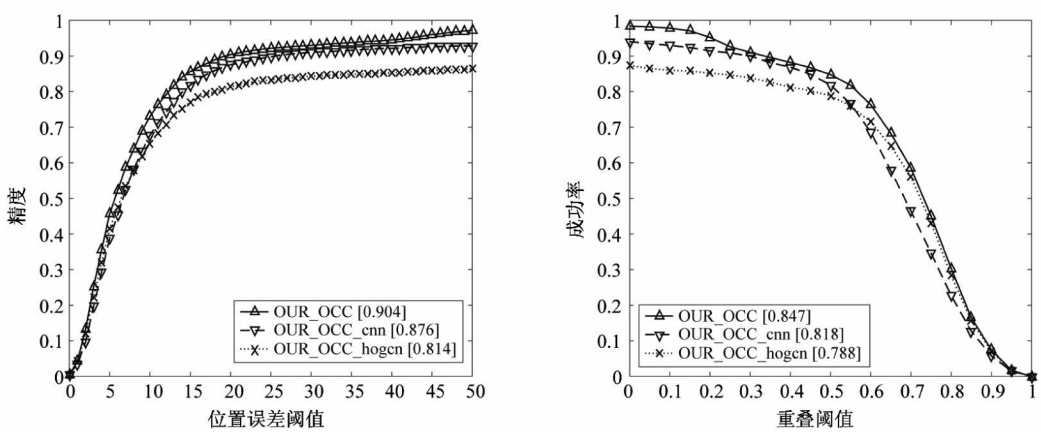

图5　融合特征前后3种算法的成功率图和精度图

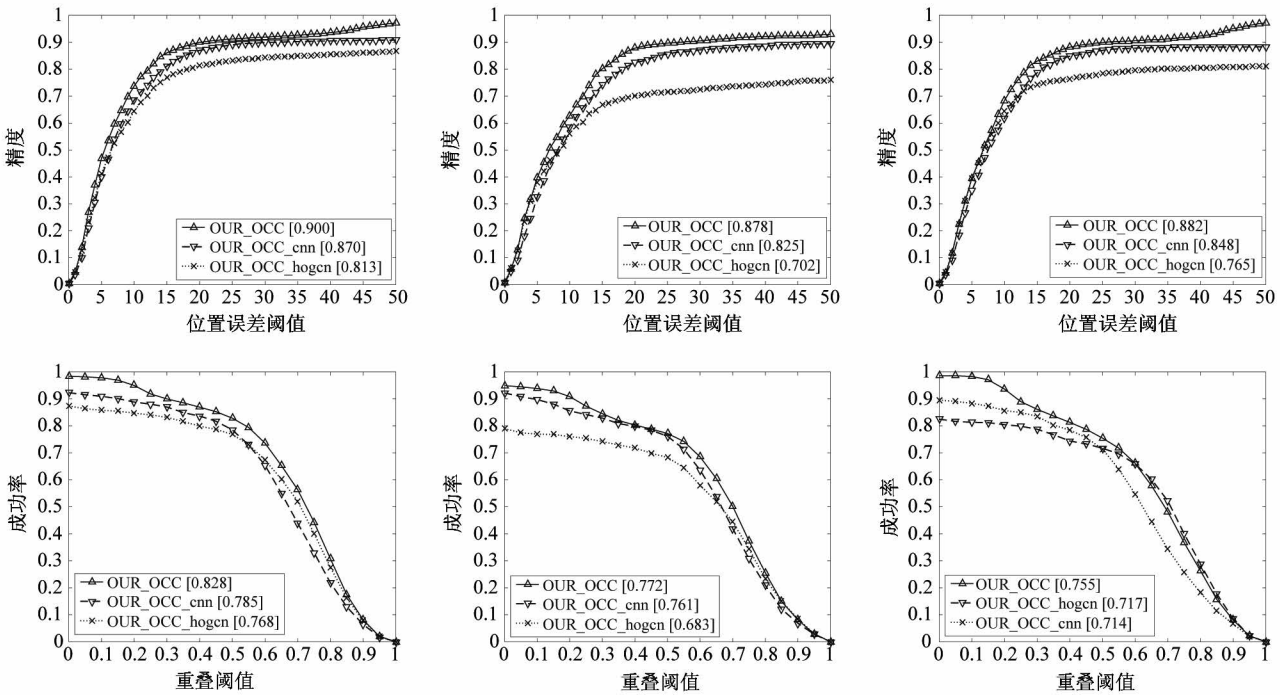

图 6　融合特征前后 3 种算法在复杂场景下精度图和成功率图

尺度变化情况下，OUR _ OCC _ cnn 的跟踪精度领先于 OUR _ OCC _ hogcn，但是鲁棒性相对较差；目标遮挡时会出现漂移的现象，而融合后的 OUR _ OCC 算法均达到了最佳性能。

对算法准确率和成功率进行绘图，本文算法在曲线图中的名称为 OUR _ OCC，如图 7 和表 2所示，在准确率和成功率方面都优于其他算法，分别达到了 90.4% 和 84.7%。图 8 显示了在遮挡、尺度变化和运动模糊情况下本文算法的精度分别达到了 93.9%、88.2%、88.7%，成功率为 92.1%、75.5% 和 80%，均优于其他算法，说明该算法能根据不同跟踪场景下特征描述目标的能力，自适应调整融合权重以及遮挡判断与更新模型，减少了目标跟丢的可能性，提高了算法的鲁棒性。

表 2　本文算法与其他算法的性能对比表

评价指标	OUR	KCF	CN	SAMF	HCF
速度(FPS)	0.9	178	74	16	0.8
精度(%)	90.4	73.7	65.3	79.6	88.6
成功率(%)	84.7	63.7	53.2	76.3	75.9

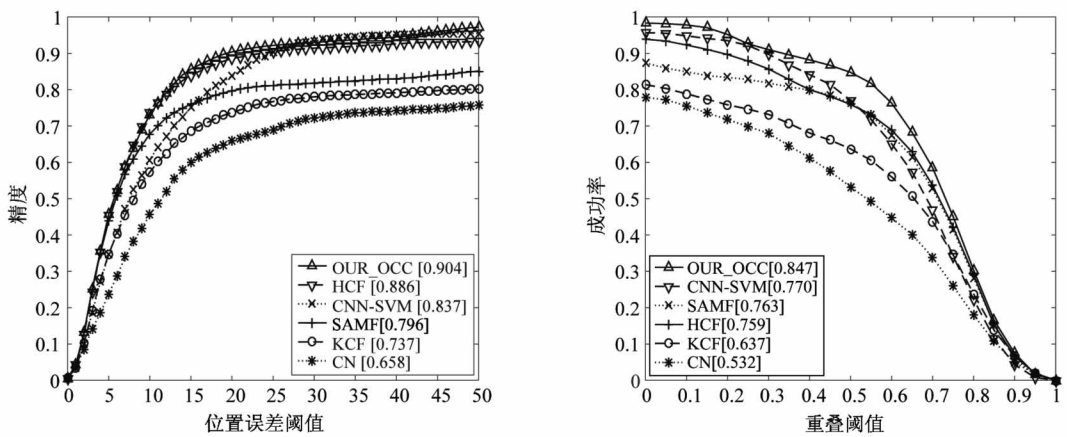

图 7　6 种算法跟踪测试基准的精度图和成功率图

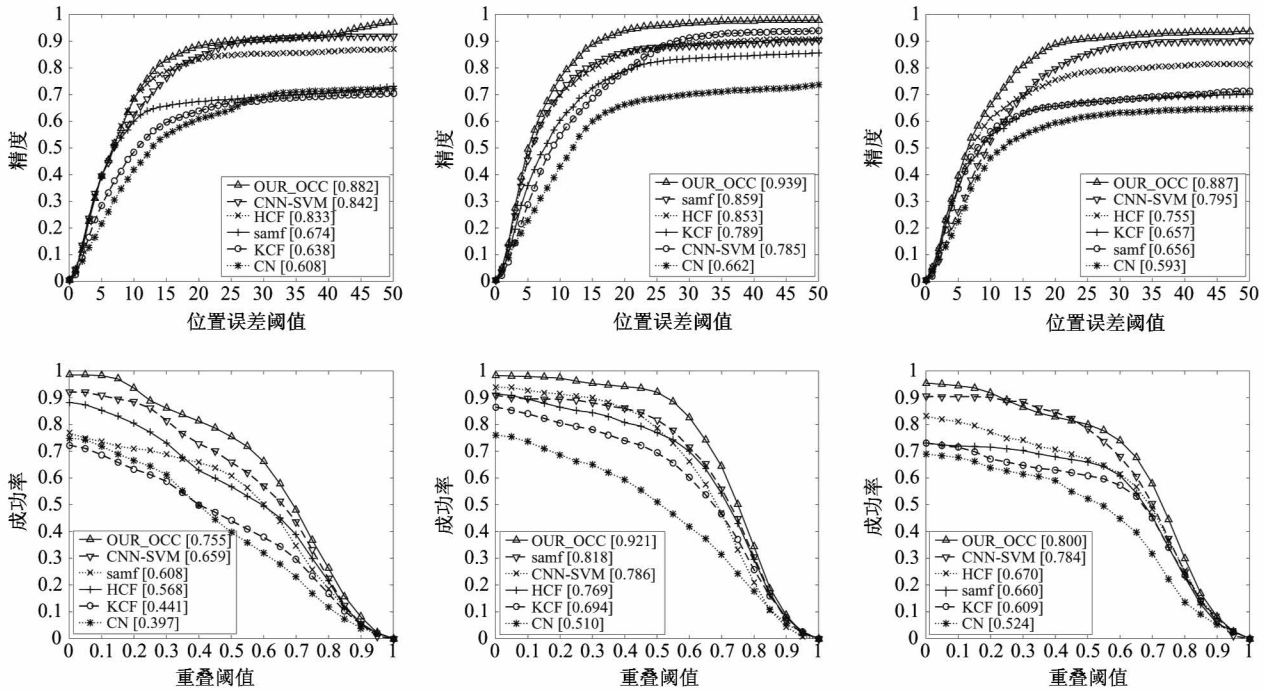

图 8　6 种算法复杂场景下跟踪测试基准的精度图和成功率图

4　结　论

本文算法提出了传统手工特征(HOG + CN)和深度学习领域中的卷积(VGG-Net-19)多特征融合,构造多个相关滤波器区别于使用单个滤波器的跟踪算法。HOG 和 CN 特征使用传统的特征提取方法进行简单的矢量相加,采用卷积神经网络(VGG-Net-19)conv5-4 层的输出作为特征,调整融合后特征的尺寸使得与 CNN 卷积特征尺寸相同,训练滤波器模型,分别学习并进行卷积以生成各自响应值。在训练阶段,根据通道的响应值计算可靠权重;在定位阶段,对通道检测响应值进行可靠加权从而估计目标的位置,提高跟踪精度。最后根据 APCE 值和 $F_{\max 1}$ 值的变化判断是否更新模型,当二者均低于一定的阈值时即目标发生遮挡,模型停止更新,反之更新模型。本文采用跟踪基准(OTB-100)对多个视频序列进行测试,并与 5 种基于相关滤波主流跟踪算法作性能对比分析。实验结果表明,HOG、CN 和卷积特征的融合,能够取得较好的精度和较强的鲁棒性,并且在遮挡、尺度变化、运动模糊等复杂情况下,也能对目标进行很好的跟踪。但是,本文算法仍存在不足之处,如表 2 所示,读取每帧图片的速度为 0.9 FPS,而 KCF 算法在相同实验条件下,达到了 178 FPS,主要是本文算法基于卷积神经网络在特征提取时比较耗时,虽然精度提升了很多,但速度有待进一步提高。在之后的研究工作可以在 GPU 上对算法进行优化,进而提升速度。

参考文献

[1] Smeulders A W M, Chu D M, Cucchiara R, et al. Visual tracking: an experimental survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(7):1442-1468

[2] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 2544-2550

[3] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]// European Conference on Computer Vision, Berlin, Germany, 2012: 702-715

[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*, 2015, 37(3):583-596

[5] Zhang K H, Zhang L, Liu Q S, et al. Fast visual tracking via dense spatio-temporal context learning[C]// European Conference on Computer Vision, Zurich, Switzerland, 2014: 127-141

[6] 张雷, 王延杰, 孙宏海,等. 采用核相关滤波器的自适应尺度目标跟踪[J]. 光学精密工程, 2016, 24(2):

448-459
[7] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[C]//European Conference on Computer Vision, Zurich, Switzerland, 2014: 254-265
[8] 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, USA, 2016: 1401-1409
[9] Roska T, Chua L O. The CNN universal machine: an analogic array computer[J]. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 1993, 40(3):163-173
[10] 胡硕, 赵银妹, 孙翔. 基于卷积神经网络的目标跟踪算法综述[J]. 高技术通讯, 2018, 28(3):207-213
[11] Wang L, Liu T, Wang G, et al. Video tracking using learned hierarchical features[J]. *IEEE Transactions on Image Processing*, 2015, 24(4):1424-1435
[12] Zou W Y, Zhu S H, Ng A Y, et al. Deep learning of invariant features via simulated fixations in video[C]//Advances in Neural Information Processing Systems, New York, USA, 2012: 3203-3211
[13] Hong S, You T, Kwak S, et al. Online tracking by learning discriminative saliency map with convolutional neural network[C]//International Conference on Machine Learning, Lile, France, 2015: 597-606
[14] 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, Santiago, Chile, 2015: 3074-3082
[15] Danelljan M, Shahbaz Khan F, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA, 2014: 1090-1097
[16] Surasak T, Takahiro I, Cheng C, et al. Histogram of oriented gradients for human detection in video[C]//International Conference on Business and Industrial Research (ICBIR), Bankok, Thailand, 2018: 172-176
[17] Van d W J, Schmid C, Verbeek J, et al. Learning color names for real-world applications[J]. *IEEE Transactions on Image Processing*, 2009, 18(7):1512-1523
[18] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 1646-1654
[19] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database[C]//IEEE Conference on Computer Vision and Pattern Recognition, Florida, USA, 2009: 248-255
[20] Sainath T N, Kingsbury B, Saon G, et al. Deep convolutional neural networks for large-scale speech tasks[J]. *Neural Networks*, 2015, 64: 39-48
[21] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA, 2013: 2411-2418
[22] Naresh Boddeti V, Kanade T, Vijaya Kumar B V K. Correlation filters for object alignment[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA, 2013: 2291-2298
[23] Everingham M, Zisserman A, Williams C K I, et al. The 2005 PASCAL visual object classes challenge[J]. *Lecture Notes in Computer Science*, 2007, 111(1): 98-136
[24] Wang N, Yeung D Y. Learning a deep compact image representation for visual tracking[C]//Advances in Neural Information Processing Systems, New York, USA, 2013: 809-817

Correlation filter tracking based on multi-feature fusion of convolutional neural networks

Yang Haiqing, Xu Qianqian, Tang Yihao, Sun Daoyang
(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

Abstract

To solve the multi-feature fusion problem in correlated filter tracking, this paper proposes a tracking algorithm combined with multi-channel correlation filter framework and multi-feature fusion convolution neural network (CNN). Firstly, this work introduces the gradient histogram and color name features, and uses the traditional feature extraction method to let the extracted features are simply vector-added. Secondly, the convolution neural network trained on ImageNet is used to extract the feature and the output of the conv5-4 convolutional layer is used as a feature. Afterwards, this work trains the correlation filters respectively. The target position is obtained by the reliable weighted sum of the feature response. Finally, by the change of both the maximum response value and the average peak-to correlation energy, it considers whether or not to update the prediction model. Experiments are carried out on the object tracking benchmark (OTB-100), and compared with five mainstream algorithms based on correlation filtering. The experimental results show that the robustness and tracking accuracy of the proposed algorithm are superior to the other algorithms' in the complex conditions of illumination changes, scale variation and occlusion.

Key words: target tracking, convolutional neural network (CNN), correlation filtering, feature extraction, reliability weighting