

一种基于多相关滤波器组合的目标跟踪方法

潘迪夫[†],李耀通,韩锬

(中南大学 交通运输工程学院,湖南长沙 410075)

摘要:针对复杂跟踪环境条件下目标的跟踪失败问题,提出一种基于多相关滤波器组合的目标跟踪方法.首先 2 个分别采用颜色属性(Color Name,CN)特征和方向梯度直方图(Histogram of Oriented Gradient,HOG)特征的核相关滤波器(Kernelized Correlation Filter,KCF)通过自适应融合手段进行响应图信息融合,确定目标的预测位置;然后通过以目标区域为基础进行多尺度采样,提取 CN-HOG 拼接特征构建尺度相关滤波器,得到目标的最佳尺度;最后设计了模型的自适应更新策略,通过判断目标是否发生遮挡来决定是否在当前帧进行模型更新.在 50 组视频序列上对所提算法与 6 种当前主流的相关滤波跟踪算法进行了实验.实验结果表明,在复杂的跟踪环境条件下,所提算法取得了最好的跟踪精度和成功率,能够有效处理目标遮挡和尺度变化等问题,且具有较快的跟踪速度.

关键词:目标跟踪;相关滤波;尺度评估;模型自适应更新

中图分类号:TP391

文献标志码:A

A Target Tracking Method Based on Multi-correlation Filter Combination

PAN Difu[†], LI Yaotong, HAN Kun

(School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China)

Abstract: To cope with the problem of object tracking failure in the challenging environment, a target tracking method based on multi-correlation filter combination was proposed. Firstly, two kernelized correlation filters(KCF) based on color name (CN) features and histogram of oriented gradient (HOG) features, respectively, integrated the map information through adaptive fusion method, and were used to determine the prediction position of the target. Then, through the multi-scale sampling based on the target region, CN-HOG composite feature was extracted to construct a scale correlation filter to obtain the optimal scale of target. Finally, the adaptive updating strategy of the model was designed to determine whether the model was updated in the current frame through determining whether the target was occluded. The proposed algorithm and 6 state-of-the-art methods were tested on 50 video sequences. The experiment results indicate that the proposed algorithm gains the best precision and success rate in the challenging environment, it can effectively deal with the problem of object occlusion and scale change, and it has a fast tracking speed.

Key words: object tracking; correlation filter; scale estimate; model adaptive updating

* 收稿日期:2018-04-18

基金项目:国家自然科学基金资助项目(51305467),National Natural Science Foundation of China(51305467);湖南省自然科学基金资助项目(12JJ4050,2016JJ4117),Natural Science Foundation of Hunan Province(12JJ4050,2016JJ4117)

作者简介:潘迪夫(1957—),男,广东兴宁人,中南大学教授,博士生导师

[†] 通讯联系人,E-mail:difupan@cus.edu.cn

视觉跟踪在计算机视觉应用领域中扮演着重要的角色^[1-3],例如视频监控、人机交互、机器人技术和增强现实等。目前,基于相关滤波(Correlation Filters, CF)的跟踪算法因其高精度、高鲁棒性、速度快的特点^[4],引起了相关学者们的广泛关注和研究。针对目标表征建模的相关滤波跟踪算法主要分为两类:单模型和多模型。单模型是指采用单特征来训练分类器的相关滤波跟踪算法,主要包括采用灰度特征的核循环结构跟踪器(CSK)^[5]、使用 CN 特征扩展稠密采样跟踪器(CN)^[6]、采用 HOG 特征的核相关跟踪器(KCF)^[6]、采用 HOG 特征的尺度自适应跟踪器(DSST)^[7]。多模型是指采用多特征来训练分类器的相关滤波跟踪算法,主要包括通过串接 CN 特征和 HOG 特征的尺度自适应跟踪器(SAMF)^[8]、基于 CN 特征和 HOG 特征贡献度进行加权组合的尺度自适应跟踪器(Staple)^[9]、根据 CN 特征和 HOG 特征的响应图进行加权融合模型自适应更新跟踪器(FAUT)^[10]、CN 特征和 HOG 特征串接的多尺度跟踪器(FMFPSA)^[11]、CN 特征和 HOG 特征自适应融合的跟踪器(FAKCF)^[12]。经过大量的实验结果分析可知,由于多模型相关滤波跟踪算法采用多特征描述目标,提高了跟踪算法的判别力,其跟踪性能高于单模型。

尽管上述的跟踪算法已取得较好的跟踪效果,但现实环境中依然存在很多挑战性问题,主要包括光照变化、背景复杂、目标旋转以及目标遮挡等情况造成目标表征发生变化而导致目标跟踪失败。还有在上述的多模型相关滤波跟踪算法中,对于多特征的使用方法仍然存在一些不合理之处,并且关于目标尺度评估的准确性也有待提高。

基于上述分析,本文提出一种基于多相关滤波器组合的目标跟踪方法(Multi-Correlation Filter Combination based Tracking Method, MCFC),其中2个分别采用单特征训练的 KCF 跟踪器用于目标定位,通过本文提出的以历史响应峰值的均值为基础的响应图信息融合判断机制进行响应图融合,得到最终的目标位置,以提高算法的定位精度;然后1个采用组合特征训练的相关滤波器用于目标尺度评估,以增加算法尺度估计的准确性;最后设计一种以历史响应峰值的均值和峰值旁瓣比为基础的模型自适应更新策略,以提升算法的鲁棒性。在若干视频序列上进行对比实验,证明了本文所提算法具有较高的整体跟踪性能。

1 KCF 跟踪器基本原理

KCF 跟踪器是一种通过引入核函数对相关滤波理论进行改进的核相关滤波方法,核心思想是通过将图像块进行循环密集采样构造循环矩阵,利用循环矩阵的性质与傅里叶域建立关系,通过离散傅里叶变换在频域中快速实现分类器训练和目标检测,以达到快速跟踪目标的目的^[6]。相关滤波目标跟踪方法采用岭回归分类器来预测目标位置^[4-6,13]。对循环采样得到的训练样本 X 和高斯函数得到的理想输出 Y 构造目标优化函数模型

$$\min_W \left(\sum_i (y_i - f(x_i))^2 + \lambda \|w\|^2 \right) \quad (1)$$

求解分类器系数 W 。通过映射函数 $\varphi(\cdot)$ 将低维空间的特征 x 映射到 Hilbert 空间得到 $\varphi(x)$, 此时分类器的解可表示为 $W = \sum_i (\alpha_i \varphi(x_i))$, 则 $f(x_j) = \langle w, \varphi(x_j) \rangle = \sum_i \alpha_i \langle \varphi(x_i), \varphi(x_j) \rangle$, 引入核函数 $k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$, 构造循环矩阵 K , 矩阵 K 中元素 $k_{ij} = k(x_i, x_j)$, 对式(1)进行求解可得

$$\alpha = (K + \lambda I)^{-1} Y \quad (2)$$

式中: I 为单位矩阵, λ 为防止过拟合的正则项^[14]。利用循环矩阵 K 的性质通过离散傅里叶变换(DFT)将问题变换到频域进行求解

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (3)$$

式中: \hat{k}^{xx} 表示样本特征 x 的自相关核向量, y 为样本 x 对应的标签向量, 符号 $\hat{\cdot}$ 表示向量的离散傅里叶变换(此表示方法应用于全文)。对于核函数, 本文选择图像处理中常用的高斯核, 其计算公式为

$$\hat{k}^{xx'} = \exp \left(-\frac{1}{\sigma^2} (\|x^2\| + \|x'\|^2 - 2F^{-1}(\hat{x}^* \odot \hat{x}')) \right) \quad (4)$$

式中: 符号 \odot 表示点乘操作, F^{-1} 表示离散傅里叶逆变换, \hat{x}^* 为 \hat{x} 的复共轭。

为获取第 t 帧图像中目标的位置, 以第 $t-1$ 帧的目标位置在第 t 帧中提取检测样本特征 Z , 求解分类器的输出响应

$$y = F^{-1}(\hat{k}^{xz} \odot \hat{\alpha}) \quad (5)$$

式中: k^{xz} 表示训练样本特征 x 和检测样本特征 z 的

互相关核向量,可通过式(4)计算获得.输出响应 y 中的峰值位置即为当前帧图像中的目标位置. 根据计算的检测结果对分类器参数 α 和目标模板 x 进行更新

$$\begin{cases} \alpha_t = (1 - \eta)\alpha_{t-1} + \eta\alpha'_t \\ x_t = (1 - \eta)x_{t-1} + (1 - \eta)z_t \end{cases} \quad (6)$$

其中, η 为学习率,以适应目标外观表征变化,提高跟踪器的鲁棒性和准确性.

2 MCFC 目标跟踪方法

2.1 位置跟踪的相关滤波器组合

MCFC 方法将采用基于颜色属性 (Color Name, CN)特征和基于方向梯度直方图(Histogram of oriented Gradients, HOG) 特征的两个相关滤波器分别对目标进行位置跟踪,两个滤波器的训练和检测过程均独立完成,两者所输出的响应图会根据本文提出的判断机制进行有效融合,得到最终的目标位置.

2.1.1 目标特征选择分析

在视觉跟踪的应用中,CN 特征^[5]是一种比灰度特征更具判别能力的多通道全局表征特征^[5],其在运动物体快速运动和变形的情况下取得很好的跟踪效果,但在光照剧烈变化下处理效果不理想;而 HOG 特征^[6]作为描述图像局部区域轮廓、形状的局部特征,光照变化对其检测效果影响不大,但在目标快速运动和变形环境下跟踪能力有所下降. 不同环境下两种特征单独使用的跟踪效果见图 1.因此,本文分别选择 CN 特征和 HOG 特征构建独立的相

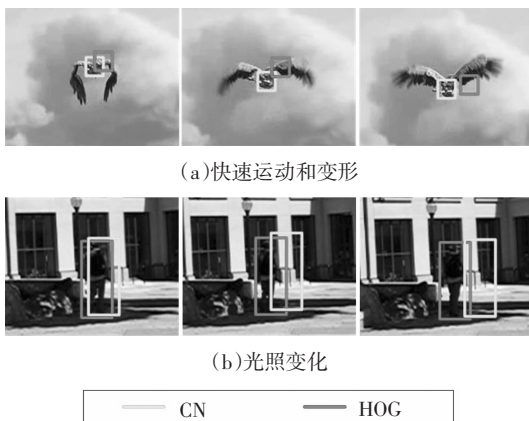


图 1 不同环境下单独使用 CN 和 HOG 特征的跟踪效果

Fig.1 The tracking effect of CN and HOG feature used separately in different environments

关滤波器,并将两者所输出的响应图进行有效融合,以此弥补在不同环境下单一滤波器跟踪性能的不足之处.

2.1.2 响应图信息融合判断机制

由 2.1.1 节分析可知,基于CN 特征和基于 HOG 特征的两个相关滤波器是互补的,两者结合能够适应更多复杂的跟踪场景,但是如何有效地将它们所输出的响应图信息融合仍然是值得深究的一个问题.目前,在视觉跟踪领域中基于特征信息融合的手段主要有两大方面,一是特征层面上的特征串接,即将提取的 CN 特征与 HOG 特征转化为二维空间大小相同的多通道特征矩阵,然后两个矩阵串接起来训练和更新滤波器,如 Li 等^[8]的 SAMF 方法和杨源等^[11]的 FMFPSA 方法,该种手段并没有对特征的贡献度进行分析,只是将两种特征简单叠加,易将错误的特征信息引入到模型中,从而导致跟踪失败;另一方面是决策层面的响应图融合,即根据颜色特征和 HOG 特征的响应值比值来给两者分配权重,两种特征的响应值与权重相乘再相加得到最终的响应图,如王晔等^[10]的 FAUT 方法和 Bertinetto 等^[9]的 Staple 方法,该种手段对特征信息的权重占比进行了分析,但也只是简单地对两种特征的峰值进行比较或者是以两者的理想输出与实际输出的差值之和的比值作为判断依据,当出现伪峰值或两种特征的响应峰值差异较大时,上述的判断依据便会失去准确性,如图 2 所示.

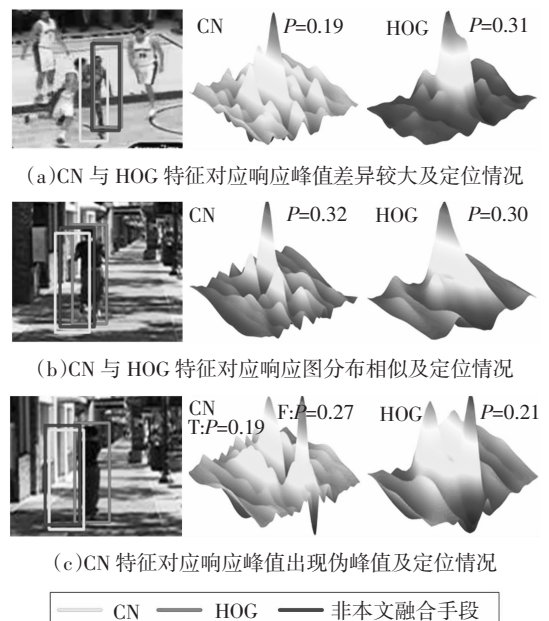


图 2 CN 与 HOG 特征对应响应值的不同分布及其定位情况
Fig.2 The different distribution and location of response

values of CN and HOG feature

图2(a)中CN特征对应的峰值较HOG特征的小,但是CN特征定位精度比HOG特征的高,此时若以两者峰值的比值作为权重的分配依据则会带来较大的定位误差;图2(b)中CN和HOG特征对应的响应图分布情况相似,但此时HOG特征定位精度远高于CN特征,若以理想输出与实际输出的差值之和的比值作为特征融合依据,也会导致目标定位误差;图2(c)中CN特征出现伪峰值,此时若采用上述任一判断依据均会造成目标定位失败.基于上述分析可知,决策层面的响应图信息融合更具有合理性,但是其融合判断手段的可靠性和准确性不足.经过大量实验分析可知响应峰值、定位精度以及历史峰值的均值存在一定的关系.因此,本文以历史峰值的均值为基础提出一种新的响应图信息融合判断机制,步骤如下.

Step 1. 已知信息:第 $t-1$ 帧图像的CN滤波器和HOG滤波器参数、目标模板以及目标位置 p_{t-1} ;两个滤波器从第1帧到第 $t-1$ 帧历史峰值的均值 AF_{t-1}^{cn} 和 AF_{t-1}^{hog} ;还有相应的阈值 $T_l^{\text{cn}}, T_u^{\text{cn}}, T_l^{\text{hog}}, T_u^{\text{hog}}$,其中 $T_l^{\text{cn}} < T_u^{\text{cn}}, T_l^{\text{hog}} < T_u^{\text{hog}}$.

Step 2. 分别求取第 t 帧图像CN滤波器和HOG滤波器的响应图 $R_t^{\text{cn}}, R_t^{\text{hog}}$,响应峰值 $MF_t^{\text{cn}}, MF_t^{\text{hog}}$,并以式(7)求取参数 $w_t^{\text{cn}}, w_t^{\text{hog}}$.

$$\begin{cases} w_t^{\text{cn}} = \frac{MF_t^{\text{cn}} - 0.5(T_l^{\text{cn}} + T_u^{\text{cn}})AF_{t-1}^{\text{cn}}}{(T_l^{\text{cn}} - T_u^{\text{cn}})AF_{t-1}^{\text{cn}}} \\ w_t^{\text{hog}} = \frac{MF_t^{\text{hog}} - 0.5(T_l^{\text{hog}} + T_u^{\text{hog}})AF_{t-1}^{\text{hog}}}{(T_l^{\text{hog}} - T_u^{\text{hog}})AF_{t-1}^{\text{hog}}} \end{cases} \quad (7)$$

Step 3. 按以下式(8)确定第 t 帧图像的响应图 R_t .

$$\begin{cases} R_t = R_t^{\text{cn}}, \text{ if } MF_t^{\text{cn}} > T_u^{\text{cn}} AF_{t-1}^{\text{cn}} \text{ and } MF_t^{\text{hog}} < T_l^{\text{hog}} AF_{t-1}^{\text{hog}}; \\ R_t = R_t^{\text{hog}}, \text{ if } MF_t^{\text{cn}} < T_l^{\text{cn}} AF_{t-1}^{\text{cn}} \text{ and } MF_t^{\text{hog}} > T_u^{\text{hog}} AF_{t-1}^{\text{hog}}; \\ R_t = 0.5R_t^{\text{cn}} + 0.5R_t^{\text{hog}}, \text{ if } MF_t^{\text{cn}} > T_u^{\text{cn}} AF_{t-1}^{\text{cn}} \\ \text{ and } MF_t^{\text{hog}} > T_u^{\text{hog}} AF_{t-1}^{\text{hog}}; \\ R_t = \frac{w_t^{\text{cn}}}{w_t^{\text{cn}} + w_t^{\text{hog}}} R_t^{\text{cn}} + \frac{w_t^{\text{hog}}}{w_t^{\text{cn}} + w_t^{\text{hog}}} R_t^{\text{hog}}, \\ \text{ if } \begin{cases} T_l^{\text{cn}} AF_{t-1}^{\text{cn}} < MF_t^{\text{cn}} < T_u^{\text{cn}} AF_{t-1}^{\text{cn}}; \\ T_l^{\text{hog}} AF_{t-1}^{\text{hog}} < MF_t^{\text{hog}} < T_u^{\text{hog}} AF_{t-1}^{\text{hog}}. \end{cases} \end{cases} \quad (8)$$

由上述步骤得到的响应图 R_t 中的最大值处即为目标位置 p_t ,当式(8)中所有的条件均不能满足时,那么取第 $t-1$ 帧图像的目标位置作为第 t 帧图像的目标位置,即 $p_t = p_{t-1}$.

2.2 尺度评估的相关滤波器

目前,在视觉跟踪领域中针对目标尺度变化问题的主要解决方法是对目标建立尺度金字塔搜索模型,在跟踪过程中通过寻找金字塔模型中输出响应最高的尺度来作为目标的真正尺度^[7-8,11,17],其中最具代表性的是Li等^[8]SAMF方法和Danelljan等^[7]的DSST方法.SAMF方法是位置跟踪与尺度评估结合起来同时计算,即在设定的所有尺度中寻找最大的输出响应值对应的位置作为目标位置,同时该最大响应值所在的尺度作为目标尺度;DSST方法是首先确定目标的位置,在该位置上再通过尺度搜索来确定目标的尺度.这两种方法均取得了很好的跟踪效果,但DSST方法的跟踪速度要比SAMF方法的更快.因此,本文借鉴DSST方法中的尺度评估手段,并对其加以改进.尺度评估采用的是线性相关滤波器,其构建方式与KCF相似,只是将高斯核替换为线性核,在这里不再赘述其基本原理,只给出相应的计算公式,如式(9)和式(10)所示,具体细节请参考文献[7].

$$y_s = F^{-1} \left\{ \frac{\sum_{l=1}^d (A^l)^* z_s^l}{B + \lambda_s} \right\} \quad (9)$$

$$\begin{cases} A^l = (1 - \eta_s) A_{t-1}^l + \eta_s \hat{g}_t^* \hat{f}_t^l \\ B_t = (1 - \eta_s) B_{t-1} + \eta_s \sum_{k=1}^d (\hat{f}_k^l)^* \hat{f}_k^l \end{cases} \quad (10)$$

式中:式(9)求取尺度评估滤波器的输出响应 y_s ; y_s 中最大响应值对应的尺度即为目标尺度.式(10)更新尺度评估滤波器参数 A^l 和 B . f 为具有 d 通道的训练样本特征, g 为高斯函数产生的理想输出, z_s 为检测样本特征, λ_s 和 η_s 分别为正则项和学习速率.

DSST方法的尺度滤波器采用的特征是HOG特征,由2.1.1节分析可知,HOG特征在目标快速运动和变形情况下跟踪效果并不理想,CN特征在光照剧烈变化环境下会造成定位误差,同样在尺度评估中两种特征遇到相同的问题也会导致尺度估计偏差,具体情形见图3.

由图3分析可知,在不同环境下仅使用单一特征对目标进行尺度评估会造成一定的尺度偏差甚至跟踪漂移.因此,本文结合两种特征对目标进行尺度评估,即将CN和HOG特征拼接起来,增加目标的表征信息,以此代替DSST方法尺度滤波器的HOG特征.如图3所示,本文尺度滤波器对目标尺度的评估结果在图3(a)和(b)中均为最优,在图3(c)中次

优,比仅使用单一特征尺度滤波器的性能有了较大的提升,同时其对跟踪环境的适用性更加广泛.

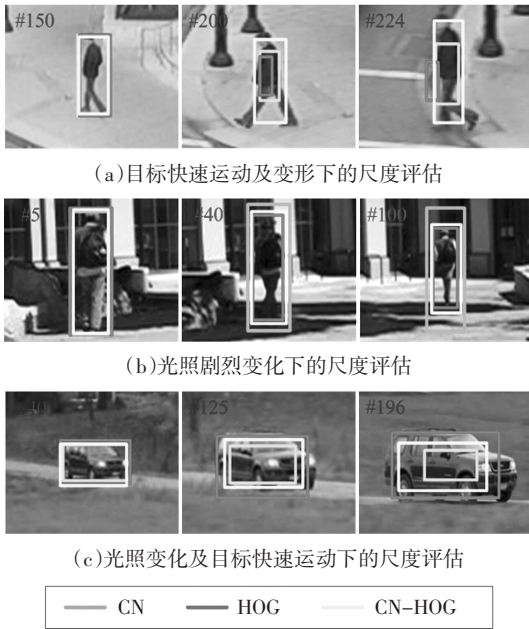


图3 不同环境 CN、HOG 和 CN-HOG 特征的尺度评估结果

Fig.3 The scale estimation results of CN, HOG and CN-HOG features in different environments

2.3 模型自适应更新

在目标跟踪过程中,由于跟踪环境的多样性和复杂性,常常会遇到一些挑战性问题,如光照剧烈变化、背景复杂和目标遮挡等,特别是在发生目标遮挡时,若此时还是继续以式(5)对滤波器参数和目标模板进行更新,则会将错误的目标表征信息引入到目标模板中,同时滤波器参数会出现偏差,最终导致后续的跟踪失败.因此必须制定一个更新策略,让跟踪算法自行判断是否发生目标遮挡,若是则不进行模型更新.经过实验分析可知,滤波器输出的响应图能够有效反映目标当前的环境状态,如背景复杂或发生遮挡等,见图4.

图4(a)显示在初始背景下滤波器输出的响应图是单峰的,且峰值很大,分布情况与理想高斯函数输出响应接近;图4(b)显示在复杂背景下滤波器输出的响应图虽然还是单峰的,但是峰值已经明显减小;图4(c)显示当目标发生部分遮挡时,滤波器输出的响应图是多峰的,且峰值已经降到很低,但依然能够成功定位目标;图4(d)显示当目标发生完全遮挡时,滤波器输出的响应图是多峰的,峰值很低,此时最大峰值位置可能并不是目标位置,即出现伪峰值,导致后续的跟踪失败.

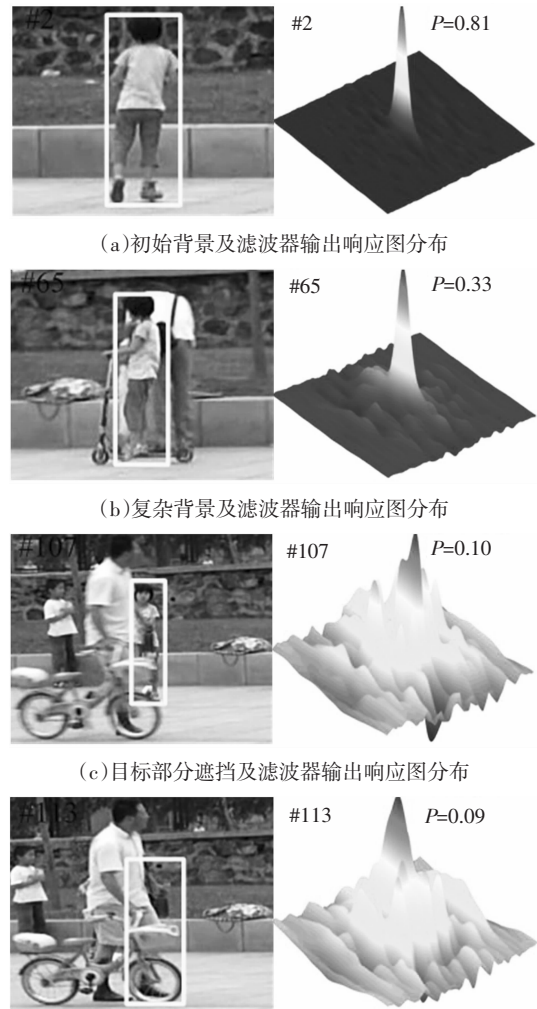


图4 不同背景下滤波器输出响应图的分布情况
Fig.4 The distribution of the response map of the filter in different backgrounds

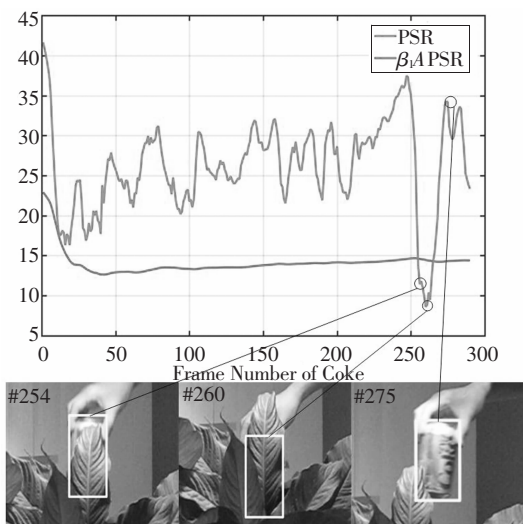
由图4分析可知,在目标跟踪过程中,可以根据滤波器输出响应图的分布情况来判断目标是否发生遮挡.为了将响应图的分布情况转化为一个量化指标,本文参考文献[13],引入峰值旁瓣比(Peak to Sidelobe Ratio, PSR)对响应图的分布情况进行统计,并将PSR作为模型更新的一个置信度指标,PSR的计算方法如下:

$$PSR_i = \frac{p - \mu}{\sigma} \tag{11}$$

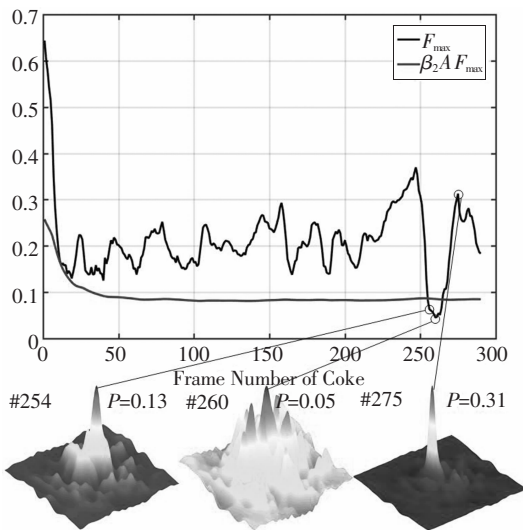
式中: p 为响应图最大值(峰值),峰值周围的区域定义为旁瓣区域; μ 和 σ 分别为旁瓣区域所有响应值的均值和标准差.实验中旁瓣区域定义为峰值周围 11×11 的区域.PSR越大,表明跟踪质量越好.为了使模型更新策略更可靠,本文将响应图输出的最大值

F_{\max} 作为另一个置信度指标,只有当这两个指标均大于其历史均值 $APSR$ 和 AF_{\max} 的一定比例 β_1 和 β_2 ,如式(12)所示,才认为是高置信度,并以式(6)对滤波器参数和目标模板进行更新,具体过程见图 5 及其分析结果.

$$\begin{cases} F_{\max}^t > \beta_1 AF_{\max}^{t-1} \\ PSR_t > \beta_2 APSR_{t-1} \end{cases} \quad (12)$$



(a)峰值旁瓣比与其历史均值比较结果



(b)最大输出响应值与其历史均值比较结果

图 5 峰值旁瓣比与其均值的比较结果和最大输出响应值与其均值的比较结果(Coke)

Fig.5 The results of PSR compared with its mean value and the maximum response value compared with its mean value (Coke)

由图 5 分析可知,Coke 序列的第 254 帧和 260 帧分别发生目标部分遮挡和完全遮挡,此时响应图的 PSR 和最大响应值均小于其历史均值的 β_1 和 β_2

倍,因此不对模型进行更新,避免将错误的目标表征信息引入到模型中;第 275 帧中,当目标不被遮挡,由于模型没有被污染,则又能重新正确跟踪目标,此时其响应图的 PSR 和最大响应值均大于其历史均值的 β_1 和 β_2 倍,可以对模型进行更新.

2.4 方法整体流程

综合上述,MCFC 方法的整体流程图如图 6 所示,方法的主要跟踪步骤如下.

输入:第 t 帧图像 I_t ;第 $t-1$ 帧图像的目标位置 p_{t-1} 和尺度大小 s_{t-1} ,CN 和 HOG 特征位置跟踪滤波器参数 $\alpha_{t-1}^{cn}, \alpha_{t-1}^{log}$ 和目标模板 $x_{t-1}^{cn}, x_{t-1}^{log}$, 尺度评估滤波器参数 A_{t-1}, B_{t-1} ;

输出:第 t 帧图像目标位置 p_t 和目标尺度 s_t ,CN 和 HOG 特征位置跟踪滤波器参数 $\alpha_t^{cn}, \alpha_t^{log}$ 和目标模板 x_t^{cn}, x_t^{log} , 尺度评估滤波器参数 A_t, B_t .

目标定位:

Step 1. 以第 $t-1$ 帧的目标位置 p_{t-1} 和尺度 s_{t-1} 为基础,在图像 I_t 中提取 CN 和 HOG 特征检测样本 z_t^{cn}, z_t^{log} ;

Step 2. 利用 z_t^{cn} 和 z_t^{log} 通过式(4)分别计算 CN 和 HOG 特征位置跟踪滤波器的输出响应 y_t^{cn}, y_t^{log} , 并根据式(7)将 y_t^{cn} 和 y_t^{log} 融合为 y_t ;

Step 3. 确定 y_t 中最大值位置即为目标位置 p_t .

尺度评估:

Step 4. 以第 t 帧目标位置 p_t 和第 $t-1$ 帧目标尺度 s_{t-1} 为基础,在图像 I_t 中分别提取 CN 和 HOG 特征,并将两者拼接为 CN-HOG 特征检测样本 $z_{s,t}$;

Step 5. 利用 $z_{s,t}$ 通过式(8)计算尺度评估滤波器的输出响应 $y_{s,t}$;

Step 6. 确定 $y_{s,t}$ 中最大值对应的尺度即为目标尺度 s_t .

模型更新:

Step 7. 以第 t 帧目标位置 p_t 和尺度 s_t 为基础,在图像 I_t 中分别提取 CN 和 HOG 特征位置跟踪滤波器的训练样本 x_t^{cn}, x_t^{log} , 尺度评估滤波器的训练样本 f_t ;

Step 8. 根据式(10),利用 x_t^{cn} 和 z_t^{cn}, x_t^{log} 和 z_t^{log} 通过式(5)分别将 CN 和 HOG 特征位置滤波器参数更新为 α_t^{cn} 和 α_t^{log} , 目标模板更新为 x_t^{cn} 和 x_t^{log} ;

Step 9. 根据式(10),利用 f_t 通过式(9)将尺度评估滤波器参数更新为 A_t, B_t .

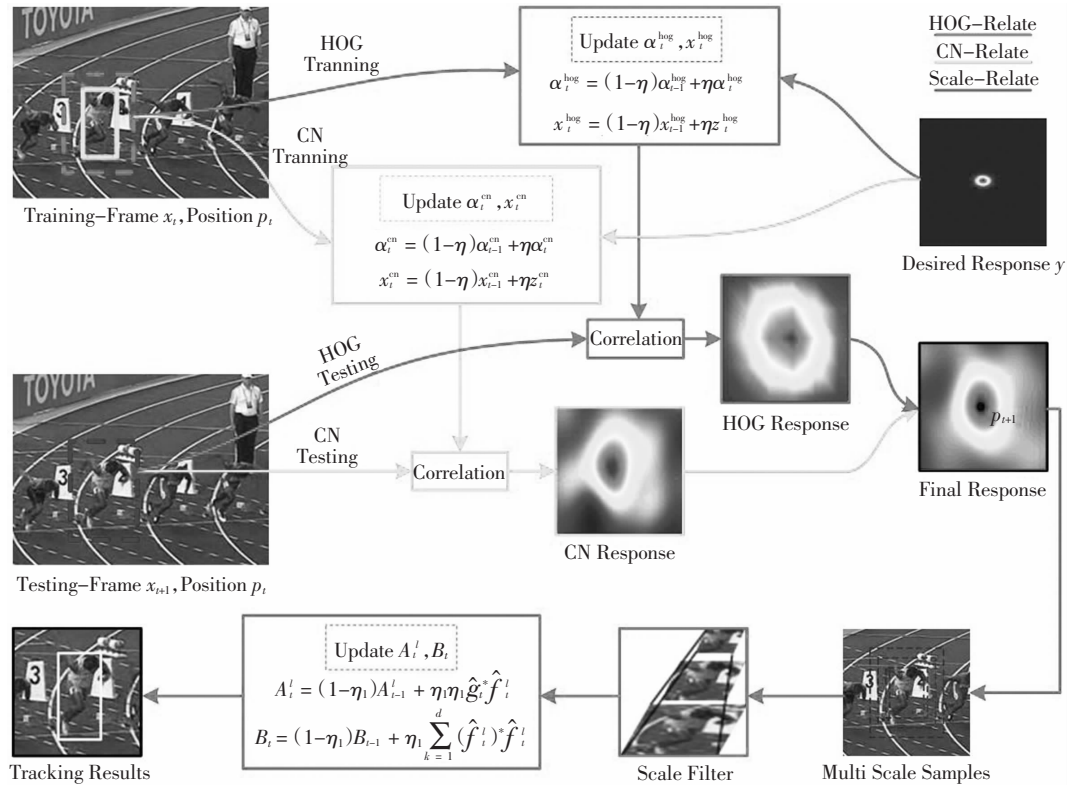


图 6 MCFC 方法流程图

Fig.6 The flow diagram of MCFC method

3 实验

本文的仿真实验均在 CPU 为 Intel Corei5-6400、主频为 2.70 GHz、8 GB 内存的计算机环境下通过 MATLAB2016a 软件平台进行实现。算法的参数设置具体如下：目标特征提取区域的大小设置为目标大小的 2.5 倍；相关滤波器的正则化参数 $\lambda_p^{\text{cn}} = 10^{-2}, \lambda_p^{\text{hog}} = 10^{-4}, \lambda_s = 10^{-2}$ ；高斯核宽 $\sigma_p^{\text{cn}} = 0.2, \sigma_p^{\text{hog}} = 0.5$ ；学习速率 $\eta_p^{\text{cn}} = 0.075, \eta_p^{\text{hog}} = 0.02, \eta_s = 0.02$ ；尺度步长 $a = 1.02$ 和尺度采样个数 $S = 33$ 。通过大量的实验对比分析，当响应图融合判断阈值 $T_l^{\text{cn}} = 0.3, T_u^{\text{cn}} = 0.75, T_l^{\text{hog}} = 0.25, T_u^{\text{hog}} = 0.65$ ，模型自适应更新比例系数 $\beta_1 = 0.3, \beta_2 = 0.45$ 时，算法能得到最好的跟踪性能。在实验过程中，所有的参数设置固定不变。

为了充分验证本文 MCFC 跟踪方法的跟踪鲁棒性和准确性，我们将使用 OTB2013^[19]测试数据集 50 个视频序列，其中包括具有尺度变化的 28 组视频和目标遮挡的 27 组视频等，并与当前主流的相关滤波跟踪算法进行对比。选取的算法有：CSK^[4]、KCF^[6]、CN^[5]、Staple^[9]、SAMF^[8]、DSST^[7]。其中，使用单特

征的有：CSK(灰度特征)、KCF(HOG 特征)、CN(CN 特征)、DSST(HOG 特征)；使用特征融合的有：Staple(颜色统计特征+HOG 特征)、SAMF(CN 特征+HOG 特征)；考虑目标尺度评估的有：Staple、SAMF、DSST。接下来，我们将根据实验结果从定性和定量两个方面分析 7 种跟踪算法的性能。

3.1 定性分析

图 7 展示的是实验中 7 种跟踪算法的部分跟踪结果。根据视频序列中不同的跟踪环境条件，我们将从以下 6 个方面对算法的性能进行定性分析：

1)光照变化：以“Shaking”和“Human7”为例。由图 7(a)和(b)分析可知，当运动目标所处环境发生剧烈的光照变化时，大部分的跟踪算法都发生了跟踪漂移，而 MCFC 方法却始终能够稳定地跟踪目标，表明该算法对光照变化具有较好的鲁棒性。

2)尺度变化：以“Skating1”和“Board”为例。由图 7(c)和(d)可知，在跟踪过程中，当运动目标的尺度发生变化时，大部分的跟踪算法都能成功定位目标，并且 SAMF、DSST、Staple 和 MCFC 算法都能实现目标尺度的自适应变化，但本文的 MCFC 算法的



图7 7种跟踪算法部分跟踪结果展示

Fig.7 Some presented tracking results of 7 tracking algorithms

尺度评估方法更加准确。

3) 目标旋转:目标旋转包括平面内旋转和平面外旋转.平面内旋转以“BlurOwl”和“Shaking”为例,在图7(a)和(e)中,目标发生了平面内旋转,此时要求算法采用的特征具有高度的旋转不变性,而本文采用的CN特征具有旋转不变性,所以能够很好适应这种情况;平面外旋转以“Board”和“Football”为例,在图7(d)和(f)中,目标发生平面外旋转,此前建立的目标模型会失效,而本文加入了模型自适应更新策略,当目标恢复之前状态,仅有本文MCFC算法成功定位目标。

4) 背景复杂:以“Couple”和“Football”为例.在背景复杂的跟踪环境下,如图7(f)和(h)所示,要求跟踪算法具有较高的准确性,而在这两个视频序列的整个跟踪过程中,只有MCFC算法能准确跟踪目标。

5) 目标遮挡:以“Girl2”和“Box”为例.在图7(i)

和(j)中,目标分别发生了部分遮挡和完全遮挡,由于本文的MCFC算法引入了模型自适应更新策略,有效减少了将错误的目标信息更新到目标模型中,对目标遮挡情况具有较高的鲁棒性,因此始终能准确跟踪目标。

6) 运动模糊:以“BlurOwl”和“BlurBody”为例.由图7(e)和(h)可知,当目标发生运动模糊时,由于本文有效的响应图信息融合判断机制,在两个视频序列的整个跟踪过程中,只有本文的MCFC算法能一直正确跟踪目标。

3.2 定量分析

本文将采用以下几个评价指标^[19]对跟踪算法进行定量分析。

1) 中心位置误差(center location error, CLE):跟踪算法检测的目标中心位置与标定的目标中心位置之间的平均欧式距离;

2) 距离精度 (distance precision, DP): 视频中 CLE 小于某一设定距离阈值的帧数与总帧数的百分比;

3) 重叠精度 (overlap precision, OP): 跟踪算法预测的目标框与标定的目标框之间的重叠面积与总面积的比值超过某一设定重叠阈值的视频帧数与总帧数的百分比;

4) 精度曲线: 横轴为给定的距离阈值, 纵轴为对应的 DP, 通常取阈值为 20 个像素时精度曲线值 (P20) 为典型值;

5) 成功率曲线: 横轴为给定的重叠阈值, 纵轴为对应的 OP, 将曲线下方的面积 (AUC) 作为该曲线的典型值.

本文共设置了三组实验, 第一组是 7 种算法的单个视频跟踪性能的定量分析; 第二组是 7 种算法的综合性能的定量分析; 第三组是对比 7 种算法的平均跟踪速度.

实验 1: 单个视频跟踪性能的定量分析

针对上述的 10 组视频序列, 采用 CLE、DP 和 OP 三个评价指标对 7 种算法的跟踪性能进行对比分析. 表 1 为 7 种跟踪算法在 10 组视频中的 CLE、

DP 和 OP 结果. 表中有下划线的数据表示最优结果, 有点式下划线的数据为次优结果 (下文采取同样表述方式). 取每种算法在 10 组视频中的平均 CLE、DP (距离阈值为 20 像素) 和 OP (重叠阈值为 0.5) 的平均值作为判断依据, 在这三个评价指标中, MCFC 算法均取得最优结果, 表明本文算法能够很好地适应不同的跟踪环境, 具有较好的跟踪性能.

实验 2: 算法综合跟踪性能的定量分析

为了综合评比 7 种算法的跟踪性能, 选用 OTB2013 的 50 组视频序列进行测试分析, 采用精度曲线和成功率曲线两个评价指标来进行综合跟踪性能的定量分析.

图 8 为 7 种算法在 50 组视频中的整体精度曲线和成功率曲线. 由图 8 分析可知, 采用多特征的 MCFC、Staple 和 SAMF 算法的精度和成功率明显高于采用单特征的 KCF、DSST、CN 和 CSK 算法. 这表明对目标使用多特征描述能够增强跟踪算法的判别能力, 提高跟踪性能. 在 MCFC、Staple 和 SAMF 三种算法中, MCFC 的精度 (P20) 和成功率 (AUC) 均是最高, 比第二名的 Staple 分别提高 7.5% 和 4.4%, 证明本文 MCFC 算法的有效性.

表 1 7 种算法在 10 组视频中的 CLE、DP 和 OP 结果

Tab.1 The results of CLE、DP、OP of 7 algorithms in 10 groups of video

评价指标	跟踪算法	Shaking	Human7	Skating1	Board	BlurOwl	Football	Couple	BlurBody	Girl2	Box	平均值
CLE/像素	CSK	16.7	18.7	7.62	120	794	16.5	144	68.5	4840	126	615.2
	SAMF	62.2	45.2	6.15	10.2	151	13.4	25.8	16.7	37.9	8.33	<u>37.7</u>
	KCF	113	48.2	7.66	35	183	14.9	47.2	64.2	265	88.8	86.7
	CN	15.1	26.5	7.92	25.7	130	16.4	123	36.5	42.4	101	52.5
	DSST	7.18	27.6	6.62	43.4	176	16.5	128	89.6	172	106	77.3
	Staple	139	2.49	70.7	31.2	124	13.1	34.4	6.33	114	90.5	62.6
	MCFC	8.72	5.84	6.21	11.4	10.7	6.06	9.78	5.8	8.89	9.97	<u>8.3</u>
DP/%	CSK	59.5	65.6	98.5	8.6	16.2	79.8	8.57	30.2	7.13	9.39	38.3
	SAMF	2.74	44.8	100	87.1	23.1	79.6	54.3	89.8	82.5	93.8	<u>65.8</u>
	KCF	1.92	47.2	100	70.1	22.7	79.6	25.7	58.4	7.07	41.5	45.4
	CN	69.9	56	100	67.6	20.9	79.8	10.7	53.9	74.7	39.4	57.3
	DSST	100	42.4	100	69.8	23.5	79.8	10.7	62.9	7.13	39.4	53.6
	Staple	1.64	100	53.8	74.3	46.9	79.8	67.9	99.4	7.6	41.4	57.3
	MCFC	95.3	100	100	88.2	92.7	100	86.4	99.4	95.7	90.2	<u>94.8</u>
OP/%	CSK	61.6	31.2	37	33	16.5	63.5	8.57	47.3	7.13	8.96	31.5
	SAMF	1.37	44.8	79.5	96.8	23.3	78.2	53.6	92.5	84.9	95.3	<u>65.0</u>
	KCF	1.37	40.8	37.3	85.4	22.8	66.6	24.3	58.7	7	35.7	37.9
	CN	67.1	31.2	38	78.9	21.1	63.8	10.7	80.8	65.1	34	49.1
	DSST	100	40.4	91.5	71.2	23.6	62.2	10.7	63.2	7.27	39.6	50.9
	Staple	1.64	100	51.2	77.6	50.6	78.2	67.9	99.7	7.4	41.5	57.6
	MCFC	94.2	89.2	99.8	93	93.3	80.4	68.6	99.7	88.3	94.4	<u>90.1</u>

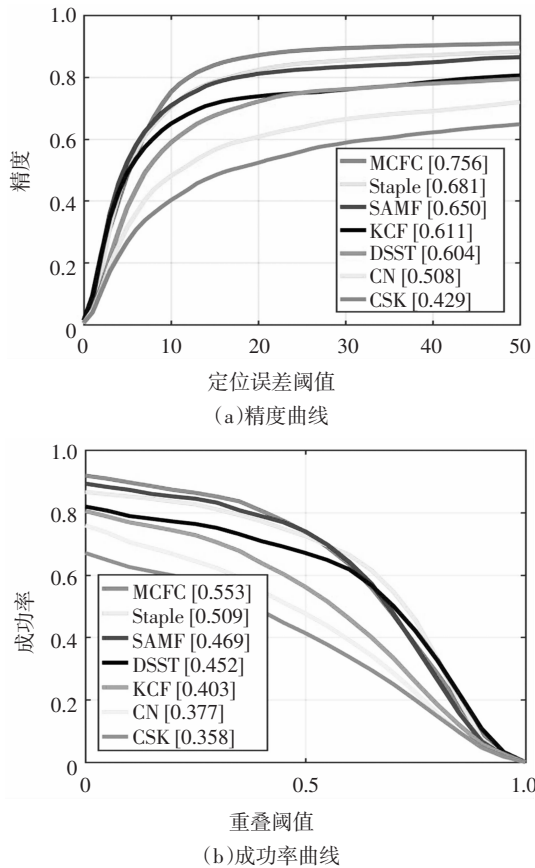


图 8 7 种算法的精度曲线和成功率曲线
Fig.8 The accuracy curve and success rate curve of 7 algorithms

为了进一步对比 7 种算法在不同的跟踪环境条件下的跟踪性能, 在 50 组视频中分别求取每种算法在 11 种跟踪环境条件^[9]下的 DP 和 OP, 结果如表 2 和表 3 所示. 11 种跟踪环境包括: 光照变化 (Illumination Variation, IV)、平面外旋转 (Out-of-Plane Rotation, OPR)、尺度变化 (Scale Variation, SV)、遮挡 (Occlusion, OCC)、变形 (Deformation, DEF)、运动模糊 (Motion Blur, MB)、快速运动 (Fast Motion, FM)、平面内旋转 (In-Plane Rotation, IPR)、不在视野内 (Out-of-View, OV)、背景复杂 (Background Clutters, BC)、低分辨率 (Low Resolution, LR)。

由表 2 数据可知, MCFC 算法的距离精度除了在 LR 跟踪环境下的结果为次优, 其余的均为最优结果, 如 SV、OCC 等; 由表 3 数据可知, MCFC 算法的重叠精度在大部分跟踪环境下均为最优结果, 如 SV、OCC 等, 其余也为次优结果. 综合表 3 和表 4 的数据, 结果表明了 MCFC 跟踪方法在目标遮挡、目标变形等跟踪环境下具有较高的鲁棒性和准确性, 充分验证了本文所提的响应图信息融合判断机制、尺度评估方法以及模型自适应更新策略的有效性。

实验 3: 7 种算法的平均跟踪速度对比

为了验证本文 MCFC 方法的实时性, 分别求取 7 种算法在 50 组视频中的平均跟踪速度, 结果如表 4 所示. MCFC 算法在取得最好跟踪精度和成功率的

表 2 不同跟踪环境下 7 种算法的 DP 结果 (距离阈值为 20 像素)

Tab.2 The DP results of 7 algorithms in different tracking environments (the distance threshold is 20 pixels)

算法	IV	OPR	SV	OCC	DEF	MB	FM	IPR	OV	BC	LR
CSK	0.508	0.468	0.485	0.456	0.360	0.337	0.386	0.534	0.253	0.528	0.577
SAMF	0.745	<u>0.740</u>	0.744	<u>0.752</u>	0.711	0.657	0.673	<u>0.761</u>	<u>0.683</u>	<u>0.683</u>	<u>0.876</u>
KCF	0.671	0.620	0.644	0.628	0.566	0.544	0.584	0.672	0.533	0.665	0.755
CN	0.549	0.577	0.556	0.528	0.484	0.438	0.441	0.63	0.400	0.594	0.740
DSST	0.744	0.638	0.675	0.654	0.611	0.502	0.547	0.698	0.500	0.682	0.749
Staple	<u>0.751</u>	0.671	<u>0.747</u>	0.733	<u>0.727</u>	<u>0.701</u>	<u>0.731</u>	0.738	0.675	0.670	0.775
MCFC	<u>0.778</u>	<u>0.770</u>	<u>0.783</u>	<u>0.760</u>	<u>0.740</u>	<u>0.754</u>	<u>0.761</u>	<u>0.789</u>	<u>0.762</u>	<u>0.821</u>	<u>0.796</u>

表 3 不同跟踪环境下 7 种跟踪算法的 OP 结果 (重叠阈值为 0.5)

Tab.3 The OP results of 7 tracking algorithms in different tracking environments (the overlapping threshold is 0.5)

算法	IV	OPR	SV	OCC	DEF	MB	FM	IPR	OV	BC	LR
CSK	0.369	0.351	0.348	0.344	0.268	0.318	0.329	0.409	0.267	0.409	0.313
SAMF	0.596	<u>0.575</u>	0.560	<u>0.591</u>	0.548	0.562	0.548	<u>0.596</u>	<u>0.597</u>	<u>0.569</u>	<u>0.554</u>
KCF	0.469	0.451	0.442	0.471	0.412	0.472	0.471	0.492	0.491	0.520	0.400
CN	0.398	0.428	0.391	0.396	0.366	0.396	0.369	0.484	0.382	0.465	0.380
DSST	0.563	0.517	0.536	0.530	0.502	0.460	0.481	0.556	0.473	0.566	0.413
Staple	<u>0.605</u>	0.539	<u>0.579</u>	0.589	<u>0.585</u>	<u>0.594</u>	<u>0.609</u>	<u>0.590</u>	0.576	0.536	0.438
MCFC	<u>0.601</u>	<u>0.597</u>	<u>0.582</u>	<u>0.594</u>	<u>0.565</u>	<u>0.658</u>	<u>0.618</u>	<u>0.596</u>	<u>0.667</u>	<u>0.663</u>	<u>0.453</u>

同时,其平均跟踪速度值为 15.6 帧/s,仍具有较快的跟踪速度。

表 4 7 种跟踪算法在 50 个视频序列上的平均跟踪速度

Tab.4 The average tracking speed of 7 tracking algorithms on 50 video sequences

跟踪算法	CSK	SAMF	KCF	CN	DSST	Staple	MCFC
平均跟踪速度/(帧·s ⁻¹)	386.6	18.1	196.7	182.4	16.8	67.4	15.6

4 结论

本文提出了一种多相关滤波器组合的目标跟踪方法,位置跟踪相关滤波器用于目标的精确定位,尺度评估相关滤波器用于目标的尺度估计。主要改进的地方有 3 处:提出一种以历史响应峰值的均值为基础的响应图信息融合判断机制,提高了目标跟踪算法的定位精度;将 CN 和 HOG 拼接作为尺度评估滤波器的使用特征,提高了跟踪算法尺度估计的准确性;结合历史响应峰值的均值和 PSR 的模型自适应更新策略,有效解决目标遮挡、目标平面外旋转等问题,提高了跟踪算法的鲁棒性。

由于本文 MCFC 算法在目标定位和尺度评估上均使用了 CN 和 HOG 特征,造成了跟踪速度下降,但在获得更高跟踪性能的同时,仍具有较快的跟踪速度。将来可以通过特征降维等手段减少算法的计算量,在保证跟踪性能的基础上提高跟踪速度。同时可以结合深度学习方法来融合多分辨率的特征图,以此提高本文 MCFC 算法的跟踪性能。

参考文献

- [1] 高文,朱明,贺柏根,等.目标跟踪技术综述[J].中国光学,2014,7(3):365—375.
GAO W,ZHU M,HE B G,*et al.* Overview of target tracking technology[J].Chinese Optics,2014,7(3):365—375.(In Chinese)
- [2] LI X,HU W,SHEN C,*et al.* A survey of appearance models in visual object tracking[J].ACM Transactions on Intelligent Systems & Technology,2013,4(4):1—58.
- [3] SMEULDERS A W M,CHU D M,CUCCHIARA R,*et al.* Visual tracking: An experimental survey[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2013,36(7):1442—1468.
- [4] RUI C,MARTINS P,BATISTA J. Exploiting the circulant structure of tracking-by-detection with kernels [C]//European Conference on Computer Vision. Springer-Verlag,2012:702—715.
- [5] DANELLJAN M,KHAN F S,FELSBURG M,*et al.* Adaptive color attributes for real-time visual tracking [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society,2014: 1090—1097.
- [6] HENRIQUES J F,CASEIRO R,MARTINS P,*et al.* High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2015,37(3):583—596.
- [7] DANELLJAN M,HAGER G,KHAN F S. Accurate scale estimation for robust visual tracking [C]//British Machine Vision Conference. 2014:65.1—65.11.
- [8] LI Y,ZHU J K. A scale adaptive kernel correlation filter tracker with feature integration [C]//European Conference on Computer Vision. Springer, Cham,2014:254—265.
- [9] BERTINETTO L,VALMADRE J,GOLODETS S,*et al.* Staple: complementary learners for real-time tracking [C] //Computer Vision and Pattern Recognition. IEEE,2016:1401—1409.
- [10] 王暉,王春平,李军,等.特征融合和模型自适应更新相结合的相关滤波目标跟踪[J].光学精密工程,2016,24(8):2059—2066.
WANG W,WANG C P,LI J,*et al.* Correlation filter tracking based on feature fusing and model adaptive updating [J]. Optics and Precision Engineering,2016,24(8):2059—2066.(In Chinese)
- [11] 杨源,涛涛,查宇飞,等.快速多特征金字塔的尺度目标跟踪方法[J].西安交通大学学报,2016,50(10):49—56.
YANG Y,KU T,ZHA Y F,*et al.* Fast multi-feature pyramids for scale-adaptive object tracking [J]. Journal of Xi'an Jiaotong University,2016,50(10):49—56.(In Chinese)
- [12] 熊昌镇,赵璐璐,郭芬红.自适应特征融合的核相关滤波跟踪算法[J].计算机辅助设计与图形学学报,2017,29(6):1068—1074.
XIONG C Z,ZHAO L L,GUO F H. Kernelized correlation filters tracking based on adaptive feature fusion[J]. Journal of Computer-Aided Design & Computer Graphics,2017,29(6):1068—1074.(In Chinese)
- [13] BOLME D S,BEVERIDGE J R,DRAPER B A,*et al.* Visual object tracking using adaptive correlation filters [C] //Computer Vision and Pattern Recognition. IEEE,2010: 2544—2550.
- [14] CORTIS C,VAPNIK V. Support-vector networks [J]. Machine Learning,1995,20(3):273—297.
- [15] WEIJER J V D,KHAN F S. An overview of color name applications in Computer Vision [C]//Computational Color Imaging. Springer International Publishing,2015:16—22.
- [16] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE,2005: 886—893.
- [17] 沈秋,严小乐,刘霖枫,等.基于自适应特征选择的多尺度相关滤波跟踪[J].光学学报,2017,37(5):166—175.
SHEN Q,YAN X L,LIU L F,*et al.* Multi-scale correlation filtering tracker based on adaptive feature selection[J]. Acta Optica Sinica,2017,37(5):166—175.(In Chinese)
- [18] 马晓楠,刘晓利,李银伢.自适应尺度的快速相关滤波跟踪算法[J].计算机辅助设计与图形学学报,2017,29(3):450—458.
MA X N,LIU X L,LI Y Y. Fast scale-adaptive correlation tracking [J]. Journal of Computer-Aided Design & Computer Graphics,2017,29(3):450—458.(In Chinese)
- [19] WU Y,LIM J,YANG M. Online object tracking: a benchmark [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2013:2411—2418.