

# 自适应特征融合和模型更新的相关滤波跟踪

常敏<sup>1,2\*</sup>, 沈凯<sup>1,2</sup>, 张学典<sup>1,2</sup>, 杜嘉<sup>1</sup>, 李峰<sup>1</sup>

<sup>1</sup>上海理工大学光电信息与计算机工程学院, 上海 200093;

<sup>2</sup>上海市现代光学系统重点实验室, 上海 200093

**摘要** 针对复杂场景下单个特征的稳健性差, 以及目标存在背景干扰和目标遮挡时跟踪失败的问题, 提出一种基于自适应特征融合和模型更新的相关滤波跟踪算法。该算法在核相关滤波的基础上, 通过对不同特征的响应图采用平均峰值-相关能量的方法进行加权求和, 实现了响应图层面的自适应特征融合。根据响应图的峰值特性计算自适应权重, 以其作为置信度确定模型的更新率, 进而设计自适应模型更新方法。实验结果表明, 该算法能够很好地适应背景干扰、目标遮挡、旋转运动等复杂场景, 与近年来优秀的相关滤波跟踪算法相比, 所提算法的平均距离精度比其中最优的算法提高了 2.64%, 平均重叠精度提高了 1.54%。

**关键词** 机器视觉; 目标跟踪; 相关滤波; 自适应特征融合; 模型更新

中图分类号 TP391

文献标识码 A

doi: 10.3788/AOS201939.0915001

## Correlation Filter Tracking Based on Adaptive Feature Fusion and Model Updating

Chang Min<sup>1,2\*</sup>, Shen Kai<sup>1,2</sup>, Zhang Xuedian<sup>1,2</sup>, Du Jia<sup>1</sup>, Li Feng<sup>1</sup>

<sup>1</sup> School of Optical-Electrical and Computer Engineering,

University of Shanghai for Science and Technology, Shanghai 200093, China;

<sup>2</sup> Shanghai Key Lab of Modern Optical System, Shanghai 200093, China

**Abstract** To address the poor robustness of single feature in a complex scene and tracking failure caused by background interference and object occlusion, this study proposes a correlation filter tracking algorithm that combines adaptive feature fusion and adaptive model update. Based on kernel correlation filtering, the proposed algorithm performs weighted summation on the response maps of different features by adopting the average peak-correlation energy method to realize adaptive feature fusion of response maps. The adaptive weight is calculated as the confidence according to the peak characteristics of the response maps to determine the update rate of the model, thereby realizing the design of an adaptive model updating method. Experimental results demonstrate that the algorithm can adapt to complex scene changes, such as background disturbance, object occlusion, and rotational motion. Compared to popular correlation filtering tracking algorithms, the proposed algorithm increases the average distance and overlapping precision by 2.64% and 1.54%, respectively.

**Key words** machine vision; object tracking; correlation filter; adaptive feature fusion; model updating

**OCIS codes** 150.0155; 100.4999; 110.1085

## 1 引 言

目标跟踪技术是计算机视觉领域的基础性研究, 在智能监控、人机交互和军事制导等领域具有广泛应用, 但在目标变形、目标遮挡、快速或不规则运动等干扰下, 目标精确定位仍然难以实现<sup>[1]</sup>。为此, 研究人员提出了大量的目标跟踪算法, Comaniciu

等<sup>[2]</sup>将均值漂移理论应用到目标跟踪领域, 采用将候选区域以概率密度函数的梯度方向移动到目标位置的局部最优方法, 取得了很好的跟踪效果, 但当目标快速运动时, 概率密度函数梯度方向不明显, 易导致跟踪失败; Nummiaro 等<sup>[3]</sup>利用粒子滤波原理, 将运动目标跟踪问题转化为贝叶斯估计问题, 使得跟踪性能得到明显提升, 但采用颜色统计直方图作为

收稿日期: 2019-03-05; 修回日期: 2019-03-22; 录用日期: 2019-05-05

基金项目: 国家重大仪器专项(2016YFF0101400)

\* E-mail: changmin@usst.edu.cn

目标特征表达过于单一,结果受背景干扰的影响;Hare 等<sup>[4]</sup>提出的核化结构支持向量机跟踪方法在结构化输出空间中引入基于结构化损失的约束,使得样本选择更可靠,跟踪效果得到提升,但速度方面有待提高。

在基于学习的跟踪方法中,基于相关滤波与基于深度学习的方法成为主流跟踪算法。其中,基于相关滤波的跟踪方法具有速度快、精度高、稳健性好的特点<sup>[5]</sup>,因而相关研究较深入。Bolme 等<sup>[6]</sup>首次将相关滤波器引入到跟踪应用,并提出了最小均方误差和(MOSSE)滤波跟踪器,该算法利用图像灰度特征训练分类器,提高了跟踪的速度;Henriques 等<sup>[5]</sup>通过介绍改进 MOSSE 算法的核函数,提出了循环结构核(CSK),利用循环矩阵性质实现了密集采样,解决了训练样本不足的问题;此后,Henriques 等<sup>[7]</sup>又提出了核相关滤波(KCF)算法,在岭回归中引入核函数,利用多通道的方向梯度直方图(HOG)特征替代单通道的灰度特征,提高了跟踪的精度;Danelljan 等<sup>[8]</sup>利用颜色属性特征(CN)扩展 CSK 算法,采用主成分分析将 11 维颜色特征降至 2 维,提升了跟踪的精度,在目标形变、运动模糊等干扰下具有良好的跟踪效果。

然而,上述方法只基于单个特征对目标进行外观描述,当目标外观变化时,模型误差易积累,从而影响跟踪器的稳健性。因此,多个特征融合描述目标的方法被引入到相关滤波器中,Li 等<sup>[9]</sup>提出尺度自适应多特征跟踪器(SAMF),串联灰度特征、CN 和 HOG,提升了跟踪器的整体性能;Bertinetto 等<sup>[10]</sup>提出了结合模板和像素特征的分类器(Staple),分别训练颜色直方图和 HOG 分类器,并在响应阶段按固定比例进行融合,实现了实时稳健的跟踪;另外,大量基于深度学习的算法相继提出,Ma 等<sup>[11]</sup>利用分层卷积神经网络(CNN)特征建立目标外观模型,对每一层卷积层分别训练相关滤波器,线性加权融合各分层响应,实现目标位置的准确定位;Danelljan 等<sup>[12]</sup>提出了连续卷积相关滤波跟踪算法(CCOT),结合 CNN、HOG、CN 进行目标跟踪,跟踪精度得到了显著提升。上述研究表明,图像特征的选择对跟踪效果具有很大影响。基于相关滤波的算法尽管能够实时跟踪目标,但由于特征描述能力有限而影响跟踪器的性能。基于深度学习的算法尽管跟踪效果很好,但其复杂特征的提取增加了运算量,算法的实时性受到了影响。

上述方法在准确性和稳健性方面都获得了很好

的跟踪效果,但是,目标在出现局部遮挡、旋转运动以及背景干扰等复杂场景的情况下,仍会发生跟踪错误或跟丢的问题<sup>[13]</sup>,此外,基于简单线性插值的模型更新机制容易导致模型退化,从而导致跟踪器发生漂移<sup>[14]</sup>。为保证跟踪实时性,同时考虑到算法性能,本文在核相关滤波跟踪算法的框架上,提出一种基于自适应特征融合和模型更新的相关滤波跟踪算法,对不同特征的响应图采用平均峰值-相关能量(APCE)方法<sup>[15]</sup>进行自适应融合,解决了单一特征稳健性差的问题;采用自适应模型更新方法确定相关滤波模型的更新率,使滤波器能够适应目标外观的变化,解决了目标遮挡造成的目标跟踪失败的问题。

## 2 本文算法

### 2.1 核相关滤波器

核相关滤波跟踪算法中,在前一帧中目标中心位置上,选取高度为  $H$ 、宽度为  $W$  的矩形区域图像块  $\mathbf{x}$  训练线性分类器模型  $f(\mathbf{x}) = \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle$ ,分类器将  $\mathbf{x}$  的循环移位图像块作为训练样本  $\mathbf{x}_i$  ( $i$  为样本序号),提取特征后,对应的标签  $y_i$  为高斯函数与  $\mathbf{x}_i$  对应的期望输出。采用正则化最小二乘分类器并引入核函数,得到目标函数模型为

$$\mathbf{w} = \arg \min_{\mathbf{w}} \sum_i [\langle \mathbf{w}, \varphi(\mathbf{x}_i) \rangle - y_i]^2 + \lambda \|\mathbf{w}\|^2, \quad (1)$$

式中:  $\mathbf{w}$  为分类器权重系数;  $\langle \cdot, \cdot \rangle$  表示内积;  $\varphi(\mathbf{x})$  表示内核  $k$  从原始空间到 Hilbert 特征空间的映射;  $\lambda$  为控制过拟合的正则化项。

因此,目标解为  $\mathbf{w} = \sum_i \alpha_i \varphi(\mathbf{x}_i)$ , 引入核函数  $k(\mathbf{x}, \mathbf{x}') = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle$  求解 ( $\mathbf{x}'$  为图像块)。利用离散傅里叶变换和循环矩阵的性质,得到核化正则化最小二乘最优解为

$$\boldsymbol{\alpha} = \mathcal{F}^{(-1)} \left[ \frac{\mathcal{F}(\mathbf{y})}{\mathcal{F}(\mathbf{k}^{xx}) + \lambda} \right], \quad (2)$$

式中:  $\mathcal{F}$  和  $\mathcal{F}^{(-1)}$  分别表示离散傅里叶变换及其逆变换;  $\mathbf{k}^{xx} = \mathbf{k}(\mathbf{x}, \mathbf{x})$  表示训练样本的自相关核。选用高斯核作为核函数,定义为

$$\mathbf{k}^{xx'} = \exp \left\{ -\frac{1}{\sigma^2} \{ \|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2\mathcal{F}^{(-1)}[\mathcal{F}(\mathbf{x}) \odot \mathcal{F}^*(\mathbf{x}')] \} \right\}, \quad (3)$$

式中:  $\odot$  表示元素点积;  $\sigma$  为高斯核函数带宽参数;  $*$  表示共轭。

在新一帧中,获取与前一帧大小相同的窗口图像块  $z$  预测目标的位置,分类器的输出响应为

$$\hat{y} = \mathcal{F}^{-1}[\mathcal{F}(k^{z\hat{x}}) \odot \mathcal{F}(\hat{\alpha})], \quad (4)$$

式中:  $\hat{y}$  为输出响应值;  $\hat{x}$  和  $\hat{\alpha}$  分别为学习得到的目

标外观模型和滤波器参数;  $k^{z\hat{x}}$  表示候选区域  $z$  与目标外观模型  $\hat{x}$  之间的核相关。

本文跟踪算法为核相关滤波器的改进,主要差别体现在自适应特征融合模块与模型更新模块两部分。图 1 为跟踪算法框架。

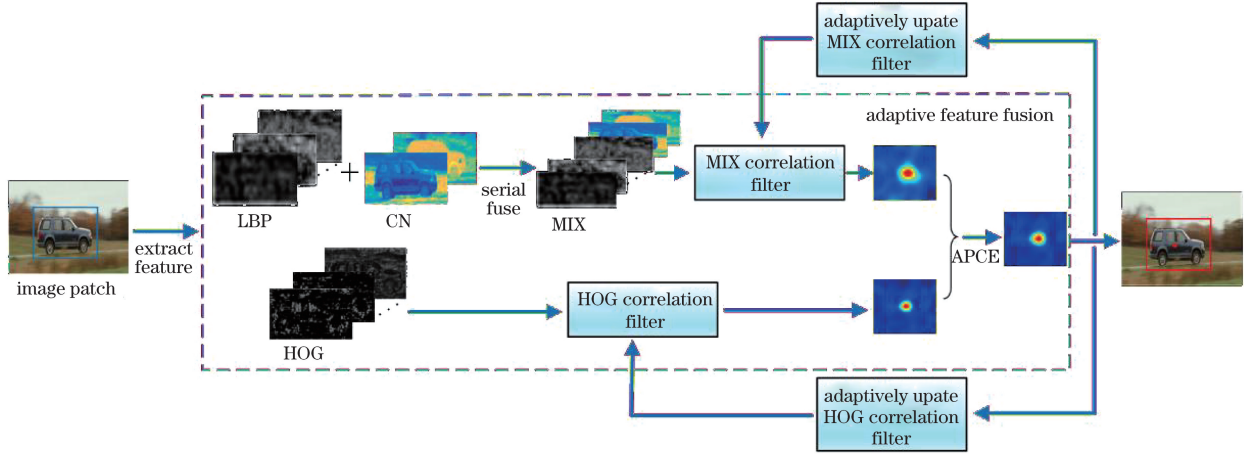


图 1 本文跟踪算法框架

Fig. 1 Framework of proposed tracking algorithm

首先,提取 3 种特征,分别为纹理特征<sup>[16]</sup> (LBP)、CN 和 HOG 特征,HOG 能够较好地适应目标的快速运动和形变,采用 31 通道 HOG 描述目标<sup>[17]</sup>,将两个互补特征 LBP 和 CN 串行融合得到融合特征(MIX,  $M_{IX}$ );然后,基于 MIX 和 HOG 分别训练相关滤波器,将两个相关滤波器得到的响应图通过 APCE 方法合并,得到最终响应图,进而确定目标的位置;此外,采用自适应模型更新方法确定跟踪器的模型更新率,即根据响应图计算自适应权重作为置信度,进一步确定模型的更新率以提升跟踪性能。

## 2.2 自适应特征融合

为了改善跟踪结果,将两个互补特征 LBP 和 CN 融合之后得到 MIX,使用 MIX 和 HOG 分别训练相关滤波器,并构建两个独立的外观模型,采用 APCE 方法自适应地融合两个相关滤波器产生的响应图,最后在最终响应图中搜索峰值,确定目标位置。

### 2.2.1 特征融合

核相关滤波器仅需计算傅里叶域中特征的元素点积和向量范数,因此,可以提取图像的多通道特征,将其作为输入<sup>[9]</sup>,则(3)式可表示为

$$k^{x x'} = \exp\left\{-\frac{1}{\sigma^2}[\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2\mathcal{F}^{-1}(D)]\right\}, \quad (5)$$

式中:  $D = \mathcal{F}(\mathbf{x}) \odot \mathcal{F}^*(\mathbf{x}')$ ,  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_d]$  表示图像的  $d$  通道特征。特征的选择是目标跟踪中非常关键的一步,选择合适的特征可以显著提高跟踪性能。根据(5)式,可以在 LBP 的基础上引入 CN。

LBP 指局部二值模式,LBP 算子定义在  $3 \times 3$  的邻域内,以中心像素点的灰度值为阈值,将相邻的 8 个像素值与中心像素值进行比较,若大于中心像素值,则将该像素点的值记为 1,否则为 0。因此,该邻域内可产生 8 位二进制数,将其按特定顺序排列形成一组 8 位二进制数,这组二进制数(0~255)则表示中心像素点的 LBP 值。LBP 值描述像素点周围区域的纹理信息,该特征计算速度快,具有灰度不变性和旋转不变性<sup>[18]</sup>。CN 颜色特征是将 RGB 空间的 3 维颜色特征映射到黑、蓝、棕、灰、绿、橙、粉、紫、红、白和黄的 11 维颜色空间特征<sup>[19]</sup>。CN 可以对不同颜色物体进行分离,对颜色区别显著且纹理形状相似的目标和背景具有很好的区分度。当目标与背景颜色相近或存在背景噪声时,CN 无法对目标进行稳定描述,可以利用 LBP 对目标进行补充描述以弥补 CN 的缺点。

采用 LBP 谱的统计直方图作为特征向量,通过 VLFeat 工具箱中的 LBP 等价模式进行 LBP 提取,采用 8 邻域像素的 LBP 算子,其特征维数为 58 维。CN 采用自适应颜色属性算法将 RGB 空间映射到

区分度明显的 11 维颜色空间中,得到 11 维颜色特征向量,再采用主成分分析将其降为 2 维颜色特征<sup>[10]</sup>。因此,将 LBP 和 CN 串行组合成  $\mathbf{M}_{\text{IX}}$ ,假设 LBP 和 CN 向量分别为  $\mathbf{L}_m$  ( $m = 1, 2, \dots, 58$ ) 和  $\mathbf{C}_n$  ( $n = 1, 2$ ),  $\mathbf{L}_m$  和  $\mathbf{C}_n$  分别表示图像第  $m$  通道的 LBP 和第  $n$  通道的 CN,则  $\mathbf{M}_{\text{IX}} = [\mathbf{L}_1 \ \mathbf{L}_2 \ \dots \ \mathbf{L}_{58} \ \mathbf{C}_1 \ \mathbf{C}_2]$ ,即将训练图像块提取的 58 通道的 LBP 和 2 通道的 CN 进行串行融合,得到 60 通道的 MIX,将其作为(5)式的输入,对相关滤波器进行训练,构建具有稳健性的目标外观,从而实现优势互补并提高滤波器的性能。

### 2.2.2 自适应融合方法

首先,使用两个由 MIX 和 HOG 训练的相关滤波器检测所有候选图像块中的最相关位置:

$$f_{\text{HOG}}(\mathbf{z}) = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{k}_{\text{HOG}}^{\text{rx}}) \odot \mathcal{F}(\mathbf{a}_{\text{HOG}})], \quad (6)$$

$$f_{\text{MIX}}(\mathbf{z}) = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{k}_{\text{MIX}}^{\text{rx}}) \odot \mathcal{F}(\mathbf{a}_{\text{MIX}})], \quad (7)$$

式中:  $\mathbf{a}_{\text{HOG}}$  和  $\mathbf{a}_{\text{MIX}}$  分别表示由 HOG 和 MIX 训练的相关滤波器;  $\mathbf{k}_{\text{HOG}}^{\text{rx}}$  和  $\mathbf{k}_{\text{MIX}}^{\text{rx}}$  分别表示 HOG 和 MIX 的训练样本  $\mathbf{x}$  与候选区域  $\mathbf{z}$  之间的核相关;  $\odot$  表示点乘;  $\mathcal{F}$  和  $\mathcal{F}^{-1}$  分别代表离散傅里叶变换及其逆变换;  $f_{\text{HOG}}(\mathbf{z})$  和  $f_{\text{MIX}}(\mathbf{z})$  分别表示由 HOG 和 MIX 的滤波器产生的响应图,线性组合两个响应图即可得到最终响应图  $f(\mathbf{z})$ :

$$f(\mathbf{z}) = \rho_{\text{HOG}} f_{\text{HOG}}(\mathbf{z}) + \rho_{\text{MIX}} f_{\text{MIX}}(\mathbf{z}), \quad (8)$$

式中:  $\rho_{\text{HOG}}$  和  $\rho_{\text{MIX}}$  为融合权重值。本文将  $A_{\text{PCE}}$  作为每个输出响应图的置信度指标确定  $\rho_{\text{HOG}}$  和  $\rho_{\text{MIX}}$ ,该指标有效反映了响应图的波动程度和可靠性。定义为

$$A_{\text{PCE}} = \frac{(\omega_{\text{max}} - \omega_{\text{min}})^2}{\text{mean} \left[ \sum_{s,j} (\omega_{s,j} - \omega_{\text{min}})^2 \right]}, \quad (9)$$

式中:  $\omega_{\text{max}}$ 、 $\omega_{\text{min}}$  和  $\omega_{s,j}$  分别表示响应图中的最大值、最小值和坐标位置为  $(s, j)$  的响应值。进一步对其进行归一化,即将当前  $A_{\text{PCE}}$  值与之前的  $A_{\text{PCE}}$  值中的最大值进行归一化,得到

$$U_{\text{APCE}}(t) = \frac{A_{\text{PCE}}(t)}{\max\{A_{\text{PCE}}(1), A_{\text{PCE}}(2), \dots, A_{\text{PCE}}(t-1)\}}, \quad (10)$$

式中:  $U_{\text{APCE}}(t)$  表示归一化的  $A_{\text{PCE}}(t)$ ,  $t$  为帧的索引。最后, HOG 和 MIX 的权重分别为

$$\rho_{\text{HOG}} = \frac{U_{\text{APCE,HOG}}(t)}{U_{\text{APCE,HOG}}(t) + U_{\text{APCE,MIX}}(t)}, \quad (11)$$

$$\rho_{\text{MIX}} = 1 - \rho_{\text{HOG}}, \quad (12)$$

式中:  $U_{\text{APCE,HOG}}(t)$ 、 $U_{\text{APCE,MIX}}(t)$  分别表示 HOG 和 MIX 的  $U_{\text{APCE}}(t)$ 。

## 2.3 自适应模型更新策略

在整个跟踪过程中,目标的外观模型应随时更新,这样才能使跟踪器对目标的外观变化具有稳健性<sup>[20]</sup>。本文提出一种自适应更新方法以确定跟踪器的模型更新率,即定义一个自适应权重  $\beta$ ,  $\beta$  定义为当前帧融合响应图的最大响应与之前帧融合响应图的最大响应之比可表示为

$$\beta = \begin{cases} \frac{\max[\mathbf{p}(t)]}{\max\{\mathbf{p}(3), \mathbf{p}(4), \dots, \mathbf{p}(t-1)\}}, & t \geq 4 \\ 1, & t = 1, 2, 3 \end{cases}, \quad (13)$$

式中:  $\mathbf{p}(t)$  表示当前帧  $t$  的融合响应图,第 1~3 帧为初始帧,第 1 帧中目标位置给定,所以第 1 帧中滤波器无输出响应;第 2 帧中滤波器根据第 1 帧给定的真实位置截取图像块,并计算输出响应,其不同于后续每一帧中滤波器根据前一帧预测位置截取图像块并计算输出响应,因为预测位置与真实位置存在一定的距离误差,且第 2 帧响应图的峰值较高,因此,为了使  $\beta$  能够更有效地根据目标外观模型而自适应变化,从第 4 帧开始计算。

相关滤波器的跟踪模型由滤波器参数  $\hat{\mathbf{a}}$  和目标外观模型  $\hat{\mathbf{x}}$  组成,在第  $t$  帧中, HOG 和 MIX 分别训练的相关滤波器跟踪模型  $(\hat{\mathbf{a}}_t, \hat{\mathbf{x}}_t)$  的更新方式为

$$\hat{\mathbf{a}}_t = (1 - \beta L_{\text{pre}}) \hat{\mathbf{a}}_{t-1} + \beta L_{\text{pre}} \hat{\mathbf{a}}, \quad (14)$$

$$\hat{\mathbf{x}}_t = (1 - \beta L_{\text{pre}}) \hat{\mathbf{x}}_{t-1} + \beta L_{\text{pre}} \hat{\mathbf{x}}, \quad (15)$$

式中:  $L_{\text{pre}}$  为初始学习率,取值为 0.015。

## 3 实验分析

### 3.1 算法分析

以标准数据集 OTB-50 中 singer2 与 David3 视频序列为例,对提出的自适应融合方法与自适应模型更新策略进行实验分析。

#### 3.1.1 特征自适应融合

对于视频序列 singer2,以 HOG 和 MIX 训练的相关滤波器对每一帧图像产生的响应图为基础,采用 APCE 作为置信度指标,计算 HOG 权重值和 MIX 权重值,图 2 所示为 HOG 和 MIX 在每帧的自适应权重值。

由图 2 可知,特征权重在跟踪过程中不断变化, MIX 的权重值整体大于 HOG 的权重值,这是由于在 singer2 视频序列中,目标运动过程中存在光照剧烈变化和背景复杂多变的干扰,而以上的情况有利于 LBP 描述,同时整个过程中目标与背景的颜色

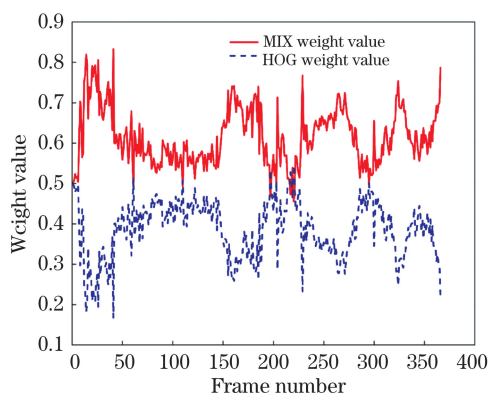


图 2 HOG 和 MIX 自适应权重值

Fig. 2 Adaptive weight values for HOG and MIX

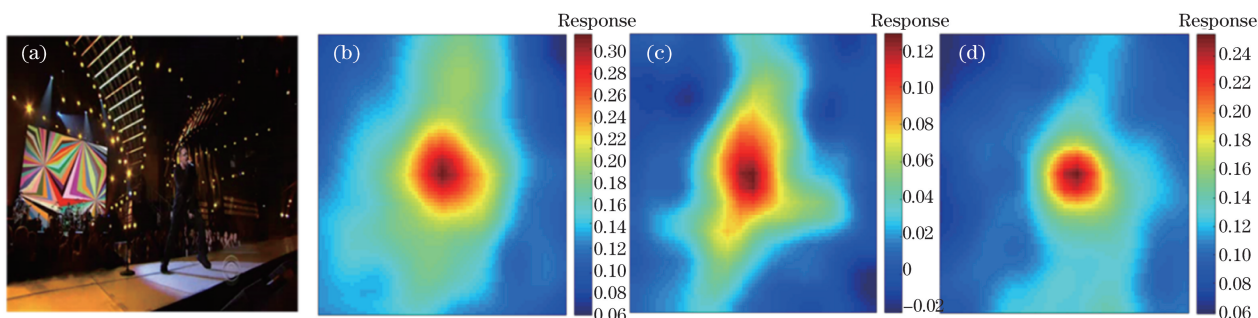


图 3 singer2 视频序列第 179 帧与对应的响应图。(a)第 179 帧；(b) MIX；(c) HOG；(d) 融合

Fig. 3 179th frame and corresponding response maps of singer2 video. (a) 179th frame; (b) MIX; (c) HOG; (d) fusion

### 3.1.2 模型自适应更新

针对目标被遮挡的跟踪问题,对于相关滤波跟踪,其滤波响应图能够有效反映跟踪质量,当目标处于遮挡情况时,响应图的峰值会急剧下降<sup>[6]</sup>。图 4 所示为视频序列 David3 中最大响应值与自适应权重以及部分帧。

由图 4 可知,在 David3 中的第 24, 83, 187 帧中,由于广告牌和大树的遮挡,跟踪下的行人几乎消失,此时响应图的最大响应值急剧减小,如果模型继续按原始学习率更新,那么模型将会逐渐弱化,因此,此处的更新率应降低。当响应图的最大响应较小时,(13)式中的自适应权重  $\beta$  也会随之减小,从而使得模型更新率减小,这样可以有效处理目标的遮挡情况,从而避免模型漂移,提高跟踪器的稳健性。

### 3.2 实验环境及设置

实验硬件环境为 Intel Core i5-3230M,主频 2.6 GHz,内存 4 GB 配置的计算机,实验平台为 Matlab R2016b 软件。实验测试序列来源于标准跟踪数据集 OTB-50<sup>[21]</sup>中的 36 组彩色序列,视频序列中包含目标跟踪中常见的复杂场景,包括快速运动、背景干扰、运动模糊、尺度变化、光照变化、目标旋转、低分辨率、目标遮挡等。将本文方法与 CCOT、

区分明显,有利于 CN 实现不同颜色物体的分离,而目标在这些场景下不利于 HOG 描述。因此,采用 APCE 方法后,两者融合后 MIX 的权重值整体上大于 HOG 值。图 3 所示为 singer2 视频序列中第 179 帧及其对应的响应图。

由图 3 可以看出,目标存在背景干扰,此时目标与背景颜色能够区分,MIX 响应图的峰值较高且集中突出,此情况不利于 HOG 描述,从而导致该帧的 HOG 响应图分布情况较分散,响应值低。因此,采用 APCE 方法可以为 HOG 分配较小权重值,减小 HOG 的影响,而 MIX 自然就分配较大权重,融合后的响应图能够准确确定目标。

SAMF、Staple、KCF、DSST<sup>[22]</sup>、CN、CSK 跟踪算法进行定量和定性比较。

### 3.3 定量评估

根据 OTB 数据集的评估方法,采用一次通过评估(OPE)模式,选择常用的评估指标评估跟踪器性能,即跟踪距离精度(DP)、跟踪重叠精度(OP)和算法时间复杂度。跟踪距离精度是指计算中心位置误差(CLE)低于某个阈值的帧数占总帧数的百分比,其中 CLE 定义为检测值与真实目标中心位置之间的平均欧氏距离,CLE 数值越小表示跟踪精度越高;跟踪重叠精度即为跟踪成功率,是指跟踪重叠率大于某个阈值的帧数占视频总帧数的百分比,其中重叠率指目标跟踪框与真实目标框的重叠面积与两者并集的总面积之比,重叠率越大表示跟踪成功率越高。根据 PASCAL 评价指标<sup>[23]</sup>,设定距离精度阈值为 20 pixel 和重叠率阈值为 0.5。算法时间复杂度采用跟踪算法每秒处理的图像帧数(FPS)衡量,FPS 定义为目标跟踪的视频总帧数与跟踪算法的总耗时之比,FPS 越高表示算法的时间复杂度越低<sup>[24]</sup>。表 1 给出了本文算法与其他 7 种算法在 36 组序列上的平均 DP、OP、FPS,图 5 所示为 8 种跟踪算法相应的距离精度曲线和成功率曲线。

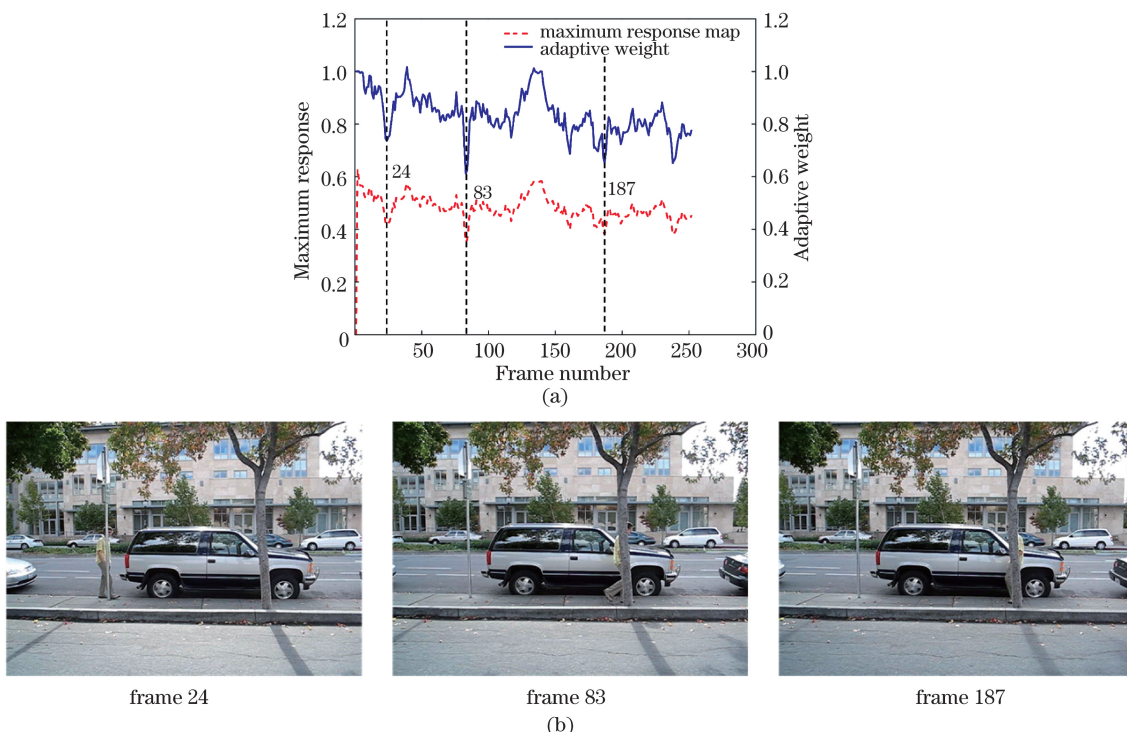


图 4 David3 中最大响应值与自适应权重以及部分帧。(a)响应图的最大响应与自适应权重；(b)部分帧

Fig. 4 Maximum response value, adaptive weight, and partial frames in David3.

(a) Maximum response map and adaptive weight; (b) partial frames

表 1 8 种跟踪算法的平均跟踪性能指标(36 组彩色序列)

Table 1 Average tracking performance indexes of eight tracking algorithms on 36 color sequences

Algorithm	Proposed	CCOT	SAMF	Staple	KCF	DSST	CN	CSK
Mean DP / %	<b>85.4</b>	83.2	79.0	78.2	77.0	76.7	67.8	55.2
Mean OP / %	<b>79.2</b>	78.0	72.0	75.2	63.8	68.9	52.9	41.8
Mean FPS / (frame·s <sup>-1</sup> )	28.39	31.42	17.53	29.26	149.60	28.12	63.23	264.33

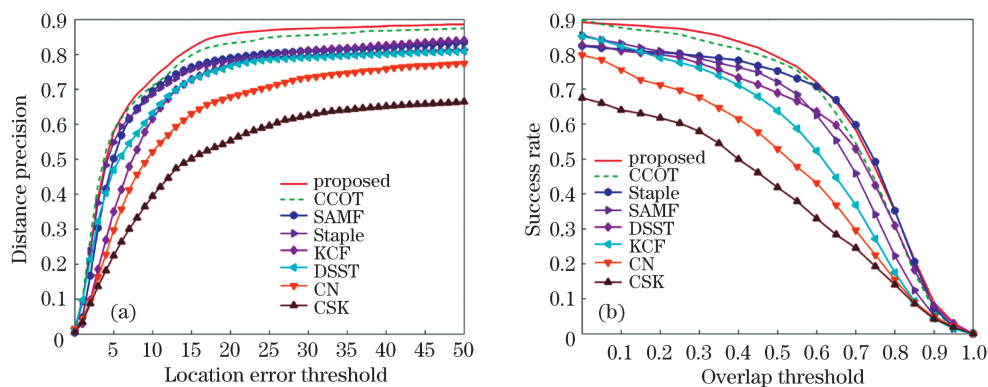


图 5 8 种跟踪算法的距离精度曲线和成功率曲线比较(36 组彩色序列)。(a)距离精度曲线；(b)成功率曲线

Fig. 5 Comparison of distance precision plots and success rate plots of eight tracking algorithms on 36 color sequences.

(a) Distance precision plots; (b) success rate plots

由表 1 可知,在 36 组彩色视频序列上,本文方法的平均 DP、平均 OP 分别为 85.4%和 79.2%,获得了最佳的结果,比采用融合 HOG 和 CN 的算法

CCOT 提高了 2.64%和 1.54%,跟踪性能具有一定程度的提升。在时间复杂度方面,跟踪速度是 KCF 和 CSK 算法的最大优势,分别达到 149.60 frame/s

和 264.33 frame/s,但其跟踪精度表现不佳,因此其综合性能仍然不足。本文算法在基于核相关滤波跟踪算法的框架上融合了多个特征,继承了相关滤波跟踪的速度优势,在提高跟踪精度的前提下,其平均FPS为 28.39 frame/s,能够应用于实时跟踪场合中。由图 5 可知,在所比较的跟踪算法中,本文算法

所得结果最优,其次为 CCOT 算法,两者均明显优于 KCF 和 DSST 算法,这说明本文算法具备较好的跟踪性能。

### 3.4 定性评估

图 6 所示为 6 种算法的跟踪结果(左上角数字为帧序号)。



图 6 6 种算法的跟踪结果。(a) Couple; (b) Lemming; (c) Liquor; (d) Girl; (e) Motor Rolling

Fig. 6 Tracking results of six tracking algorithms. (a) Couple; (b) Lemming; (c) Liquor; (d) Girl; (e) Motor Rolling

背景干扰: Couple 序列中,目标在行走过程中姿态发生变化,同时受到周围背景中来往车辆的干扰,本文算法和 CCOT 算法可以持续准确跟踪目标,而其他算法都已经发生严重漂移。第 91 帧处背景存在运动车辆和相近颜色的干扰, SAMF、Staple 和 KCF 算法跟踪失败,在后续帧中只有本文算法和 CCOT 算法能够继续保持准确跟踪。

局部遮挡: Lemming 序列中,只有本文算法和 SAMF 算法能够长期准确跟踪目标,而其他跟踪算

法都未能持续跟踪目标。第 454 帧小熊在运动过程中被游标卡尺局部遮挡,本文算法能够处理遮挡情况,准确跟踪到了目标。

类目标干扰: Liquor 和 Girl 序列中,目标在运动过程中背景复杂且多变,背景中存在相似目标的干扰。Liquor 第 1519 帧中,在移动另一个酒瓶过程中,酒瓶对目标产生了干扰,除了本文算法、CCOT、Staple 和 KCF 算法,其他算法均跟踪失败; Girl 中第 443 帧和第 471 帧处出现相似人脸的干扰

后,只有本文算法和 CCOT 算法可以持续准确跟踪目标。这是由于本文算法结合了多种特征,从而实现了目标的准确跟踪,说明本文算法具有很好的抗干扰能力和稳健性。

旋转变形:Motor Rolling 序列中,目标在快速运动过程中伴随姿态发生旋转变形和光照剧烈变化,只有本文算法能够始终准确稳定跟踪第 89 帧目标的旋转运动,其他跟踪算法都发生了漂移,本文算法能够持续准确跟踪目标。

## 4 结 论

在核相关滤波跟踪算法的框架上,提出了一种基于特征自适应融合和模型自适应更新的相关滤波跟踪算法。提取了 3 种特征——LBP、CN、HOG,将互补特征 LBP 和 CN 进行串行融合得到 MIX,对 MIX 和 HOG 的响应图采用 APCE 方法进行合并,得到最终的响应图,从而确定了目标位置,实现了复杂场景下目标的稳定跟踪。采用自适应模型更新方法确定相关滤波模型的更新率,优化了模型更新方式,滤波器能够适应目标外观的变化。实验结果表明,本文算法提高了目标跟踪的稳健性和准确性,进一步研究重点为算法优化问题,以及如何提高跟踪方法的实时性。

## 参 考 文 献

- [1] Yin H P, Chen B, Chai Y, *et al.* Vision-based object detection and tracking: a review[J]. *Acta Automatica Sinica*, 2016, 42(10): 1466-1489.  
尹宏鹏, 陈波, 柴毅, 等. 基于视觉的目标检测与跟踪综述[J]. *自动化学报*, 2016, 42(10): 1466-1489.
- [2] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift[C] // 2000 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 15, 2000, Hilton Head Island, SC, USA. New York: IEEE, 2000: 142-149.
- [3] Nummiaro K, Koller-Meier E, van Gool L. An adaptive color-based particle filter[J]. *Image and Vision Computing*, 2003, 21(1): 99-110.
- [4] Hare S, Golodetz S, Saffari A, *et al.* Struck: structured output tracking with kernels[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10): 2096-2109.
- [5] Henriques J F, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[M] // Fitzgibbon A, Lazebnik S, Perona P, *et al.* *Computer vision-ECCV 2012*. Berlin, Heidelberg: Springer, 2012: 702-715.
- [6] Bolme D, Beveridge J R, Draper B A, *et al.* Visual object tracking using adaptive correlation filters[C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 2544-2550.
- [7] 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.
- [8] Danelljan M, Khan F S, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1090-1097.
- [9] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[M] // Agapito L, Bronstein M M, Rother C. *Computer vision-ECCV 2014 workshops*. Cham: Springer, 2015: 254-265.
- [10] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: complementary learners for real-time tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1401-1409.
- [11] Ma C, Huang J B, Yang X K, *et al.* Hierarchical convolutional features for visual tracking[C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 3074-3082.
- [12] Danelljan M, Robinson A, Khan F S, *et al.* Beyond correlation filters: learning continuous convolution operators for visual tracking[M] // Leibe B, Matas J, Sebe N, *et al.* *Computer vision-ECCV 2016*. Cham: Springer, 2016: 472-488.
- [13] Liu W, Zhao W J, Li C. Long-term visual tracking based on spatio-temporal context [J]. *Acta Optica Sinica*, 2016, 36(1): 0115001.  
刘威, 赵文杰, 李成. 时空上下文学习长时目标跟踪 [J]. *光学学报*, 2016, 36(1): 0115001.
- [14] 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.  
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.
- [15] Wang M M, Liu Y, Huang Z Y. Large margin object



- tracking with circulant feature maps[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 4021-4029.
- [16] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, 24(7): 971-987.
- [17] Felzenszwalb P F, Girshick R B, McAllester D, *et al.* Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627-1645.
- [18] 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): 0515001.  
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): 0515001.
- [19] Possegger H, Mauthner T, Bischof H. In defense of color-based model-free tracking[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 2113-2120.
- [20] 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.  
王隳, 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.
- [21] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 2411-2418.
- [22] Danelljan M, Häger G, Khan F S, *et al.* Accurate scale estimation for robust visual tracking[C]// The British Machine Vision Conference 2014, September, 2014, Nottingham, UK. Guildford: BMVA Press, 2014: 1-11.
- [23] Everingham M, van Gool L, Williams C K I, *et al.* The Pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [24] Ge B Y, Zuo X Z, Hu Y J. Long-term object tracking based on feature fusion [J]. *Acta Optica Sinica*, 2018, 38(11): 1115002.  
葛宝义, 左宪章, 胡永江. 基于特征融合的长时目标跟踪算法[J]. *光学学报*, 2018, 38(11): 1115002.