

文章编号 1004-924X(2019)05-1178-10

## 采用响应图置信区域自适应特征融合的相关滤波跟踪

高 贇<sup>1,2\*</sup>, 赵江珊<sup>1</sup>, 罗久桓<sup>1</sup>, 周 浩<sup>1</sup>

(1. 云南大学 信息学院, 云南 昆明 650504;

2. 昆明物理研究所, 云南 昆明 650223)

**摘要:**针对固定特征融合权重的相关滤波跟踪算法在光照变化、目标形变下跟踪失败的问题,提出了采用响应图置信区域自适应特征融合的相关滤波跟踪算法以提高算法鲁棒性。首先,将响应图中高于响应图期望值的区域作为响应图置信区域,然后,根据 HOG 特征响应图置信区域计算出 HOG 特征响应图和颜色直方图特征响应图的融合权重,实现 HOG 和颜色直方图特征的自适应融合。仿真实验采用跟踪基准数据库(OTB-2015)中的 100 组视频序列进行实验,对比了 5 种流行的相关滤波跟踪算法。实验结果表明,本文算法的综合 AUC 和精度分别为 0.609 和 0.814,尤其在光照环境下 AUC 和精度分别为 0.655 和 0.847,相比 Staple 分别提升 5.7% 和 5.6%。本文算法在光照和形变交叉环境下 AUC 达到 0.681。在光照变化、目标形变、背景混乱、尺度变化等场景下,本文算法具备更优的跟踪性能。

**关键词:**目标跟踪;特征融合;自适应;响应图;置信区域

**中图分类号:**TP391 **文献标识码:**A **doi:**10.3788/OPE.20192705.1178

## Adaptive feature fusion with the confidence region of a response map as a correlation filter tracker

GAO Yun<sup>1,2\*</sup>, ZHAO Jiang-shan<sup>1</sup>, LUO Jiu-huan<sup>1</sup>, ZHOU Hao<sup>1</sup>

(1. School of Information Science and Engineering, Yunnan University, Kunming 650504, China;

2. Kunming Institute of Physical, Kunming 650223, China)

\* Corresponding author, E-mail: gausegao@163.com

**Abstract:** To correct the failure of some correlated filter trackers using a fixed weight feature fusion under illumination variation and deformation, the correlation filter tracker with an adaptive feature fusion and a confidence region of the response map is proposed for enhancing tracking robustness. The confidence region of the response map is the region where each response value is higher than the expectation of the response map. The fusing weights of a HOG response map and color histogram response map at every frame were calculated using the confidence region of the HOG response map, realizing adaptive fusion. The simulated experiments compared the proposed tracker with five popular

**收稿日期:**2018-11-15; **修订日期:**2019-01-06.

**基金项目:**国家自然科学基金青年项目资助(No. 61802337);国家自然科学基金地区项目资助(No. 11663007);中国博士后科学基金第 62 批面上项目资助(No. 2017M623323XB);云南省应用基础研究面上项目资助(No. 2016FB103)

correlation filter trackers using a benchmark video database, OTB-2015. The experimental results show that the AUC and precision were 0.609 and 0.814, respectively, whereas under OTB-2015 values of 0.655 and 0.847, respectively, were obtained. Under illumination variation, the obtained values were 5.7% and 5.6% higher than Staple, and the AUC was 0.681 under illumination variation and deformation. With illumination variation, target deformation, background clutter and scale variation, the proposed tracker exhibited better tracking performance than the previously developed methodologies.

**Key words:** object tracking; feature fusion; adaptive; response map; confidence region

## 1 引 言

视频目标跟踪是计算机视觉领域的一个研究热点,其目的是为了在视频图像序列中估计目标的位置。该技术在视频监控、人机交互、机器人领域和无人驾驶汽车等众多应用中扮演着十分重要的角色。近几年,虽然该领域已经取得了许多进步,但是,由于复杂场景下光照变化、外观变形、局部遮挡、快速运动、运动模糊、背景相似等诸多因素的干扰,视频目标跟踪仍然是一项具有挑战性的任务。

卷积原理表明耗时的卷积运算在傅立叶变换域可以转换为元素点积运算。基于卷积原理的相关滤波技术被引入到目标跟踪中,其极高的处理速度引起了国内外研究人员的广泛关注和深入研究<sup>[1-8]</sup>。Bolme 等<sup>[1]</sup>提出的 MOSSE (Minimum Output Sum of Squared Error Filter) 算法首次将相关滤波的思想引入到目标跟踪领域,可以实现高达 669 帧的处理速度。Henriques 等<sup>[3]</sup>基于核相关滤波实现了 KCF (Kernelized Correlation Filters) 跟踪算法,其通过循环矩阵性质将相关滤波模板训练阶段的采样问题等价于特征矩阵循环移位操作,来完成训练样本密集采样,循环采样有效增加了样本的数量,进一步提高了跟踪器的鲁棒性。在尺度适应方面,Danelljan 等<sup>[5]</sup>在 MOSEE 的基础上提出了 DSST (Discriminative Scale Space Tracker) 算法,分别利用两个单独的滤波器对位置和尺度进行估计。Yang Li 等<sup>[8]</sup>提出 SAMF (Scale Adaptive Multiple Feature) 算法实现目标尺度的自适应跟踪,且融合了 HOG (Histogram of Oriented Gradient) 特征<sup>[9]</sup>和 Color Names 特征<sup>[10]</sup>提高了跟踪器的鲁棒性。Danelljan 等<sup>[11]</sup>提出了 fDSST 算法,该算法通过对位置

滤波器和尺度滤波器分别进行 PCA (Principal Component Analysis) 降维和 QR 分解,实现了 DSST 算法的加速处理。研究人员也发现相比复杂的跟踪策略和分类器而言,目标特征对跟踪性能有着至关重要的影响。大多数跟踪器进行跟踪时,通常使用单一特征或多特征融合实现特征提取。Danelljan 等<sup>[10]</sup>采用 Color Names 特征扩展了 CSK<sup>[11]</sup> (Circulant Structure Kernels) 算法,在彩色视频序列的跟踪中取得了较好的效果。文献<sup>[13-17]</sup>将深度卷积特征引入到相关滤波跟踪算法中,进一步促进了跟踪算法的鲁棒性,但付出了极大的速度代价,难以满足跟踪应用对实时性要求。在多特征融合方面,文献<sup>[15]</sup>将 HOG 特征和 Color Names 特征直接叠加。Staple 算法<sup>[17]</sup>将 HOG 特征和颜色直方图<sup>[18]</sup>特征以 0.7 和 0.3 的权重比例进行融合。以上固定权重的多特征融合方式有效促进了跟踪鲁棒性,但对特定视频场景缺乏针对性。

针对固定权重融合方式对特定视频场景缺乏针对性的问题,本文在 Staple 算法的基础上,提出了一种采用响应图置信区域自适应特征融合的相关滤波跟踪算法。我们将响应图中高于响应图期望值的区域作为响应图置信区域,根据 HOG 特征响应图置信区域可以计算出 HOG 特征响应图和颜色直方图响应图融合权重,实现了 HOG 和颜色直方图特征的自适应融合。为了更好地验证本文算法的性能,利用 OTB-2015 基准数据集<sup>[19]</sup>中的评价标准和 100 个视频序列测试本文算法,同时与 5 种主流算法进行性能比较。

## 2 Staple 跟踪算法

回顾多特征融合的核相关滤波跟踪算法,以 Staple 算法<sup>[17]</sup>为例,其采用 HOG 特征和颜色直

方图以固定权重比例进行特征融合。HOG 特征对运动模糊和光照变化不敏感,对形变敏感,而颜色直方图恰恰相反。这使得二者成为较优的互补特征。以下将对 Staple 算法原理进行介绍和分析。

### 2.1 算法原理

令  $x_t$  表示第  $t$  帧图像,  $S_t$  表示第  $t$  帧候选目标集合,在模型参数  $\theta$  的约束下可以得到一个候选目标  $p \in S_t$  的评分  $f(T(x_t, p); \theta)$ ,所有候选目标的评分结果即该帧的响应图矩阵,其最高评分即响应图最大值对应位置的候选目标即为该帧跟踪结果  $p_t \in S_t$ ,其计算方法如式(1)所示:

$$p_t = \operatorname{argmax}_{p \in S_t} f(T(x_t, p); \theta_{t-1}), \quad (1)$$

其中:  $L(\theta; X_t)$  表示损失函数,  $X_t = X\{x_i, p_i\}_{i=1}^t$  表示历史帧跟踪结果序列,  $R(\theta)$  表示正则项以防止过拟合,  $\lambda$  为正则项参数,则模型参数  $\theta_{t-1}$  的计算方法如式(2)所示:

$$\theta_{t-1} = \operatorname{argmin}_{\theta \in Q} \{L(\theta; x_{t-1}) + \lambda R(\theta)\}. \quad (2)$$

损失函数  $L(\theta; X_t)$  是所有样本损失的线性组合,如式(3)和式(4)所示:

$$L(\theta, x_t) = \sum_{i=1}^t w_i \ell(x_i, p_i, \theta), \quad (3)$$

$$\ell(x, p, \theta) = d(p, \operatorname{argmax}_{q \in S} f(T(x, q); \theta)). \quad (4)$$

在 Staple 算法中,令  $h$  和  $\beta$  分别表示 HOG 特征和颜色直方图特征的模型参数,二者通过两个独立的岭回归问题进行求解,如式(5)和式(6)所示:

$$h_\tau = \operatorname{argmin}_h \left\{ L_{\text{unpl}}(h; X_t) + \frac{1}{2} \lambda_{\text{unpl}} \| |H| \|^2 \right\}, \quad (5)$$

$$\beta_t = \operatorname{argmin}_\beta \left\{ L_{\text{hist}}(\beta; X_t) + \frac{1}{2} \lambda_{\text{hist}} \| |\beta| \|^2 \right\}. \quad (6)$$

评分函数  $f$ 、损失函数  $L$ 、正则项  $R$  等对目标位置的准确定位都非常重要。对于评分函数,Staple 算法采用 HOG 特征对应评分函数  $f_{\text{unpl}}$  和颜色直方图特征对应评分函数  $f_{\text{hist}}$  的线性组合进行计算,如式(7)所示:

$$f(x) = \gamma_{\text{unpl}} f_{\text{unpl}}(x; h) + \gamma_{\text{hist}} f_{\text{hist}}(x; \beta), \quad (7)$$


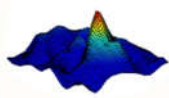
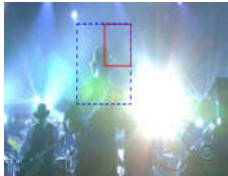
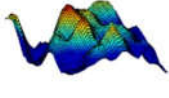

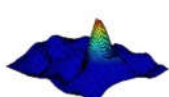
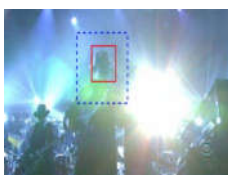
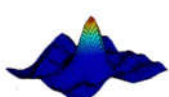
其中:  $\gamma_{\text{unpl}}$  和  $\gamma_{\text{hist}}$  是线性组合参数,Staple 算法采用 0.7 和 0.3 的固定权重比例。

### 2.2 算法分析

Staple 算法采用固定权重比例对 HOG 特征响应图和颜色直方图响应图进行线性组合。然而,在实际复杂环境下,被跟踪目标可能会受到光照变化、外观变化、快速运动、局部遮挡等各种因素的影响,采用固定权重比例进行融合的方式难以适应各种实际复杂环境。

表 1 Shaking 序列采用不同融合权重的跟踪结果

Tab.1 Tracking results of Shaking sequence using different fusing weights in

$\gamma_{\text{unpl}}$	$\gamma_{\text{hist}}$	HOG response map at 55th frame		HOG response map at 60th frame	
0.7	0.3				
0.8	0.2				

对特定视频序列采用不同固定权重比例进行融合,会产生不同的跟踪结果。如图 1 所示,对 Shaking 序列分别采用不同的固定融合权重进行

跟踪,上行对应权重分别为 0.7 和 0.3,下行对应权重分别为 0.8 和 0.2。在第 55 帧光照变化较小的情况下,两种融合权重都能准确跟踪。然而,

在第 60 帧跟踪目标遇到光照剧变的情况下,由于颜色直方图对光照剧变非常敏感,采用 0.3 的融合比例使得颜色直方图对融合后的响应图产生了较大累积影响,响应图呈现多峰状,最大响应值对应位置不再是目标的正确位置,进而导致跟踪失败。当颜色直方图的融合比例降为 0.2 时,第 60 帧光照剧变的情况下,融合后的响应图依然呈现单尖峰状,最大响应值对应位置仍然是跟踪目标的正确位置。

同一视频序列中,不同视频帧也具有不同的场景特点。当目标处于光照变化场景下,颜色直方图对光照敏感而 HOG 特征不敏感,可以增加 HOG 特征响应图的融合比例;当目标发生重度形变时,颜色直方图是全局特征,受到的影响非常小,而 HOG 特征响应图可信度会降低,此时可以增加颜色直方图的融合比例。

由此可见,在跟踪过程中,由于目标和背景都在不断发生变化,不同的视频序列难以采用统一的固定权重比例进行融合,同一视频序列的不同视频帧也应根据场景情况选取自适应的权重比例进行融合。

### 3 自适应特征融合

为了解决 Staple 算法固定权重融合方式在光照变化、目标形变及背景混乱等场景下容易跟踪失败的问题,本文提出采用响应图置信区域实现 HOG 和颜色直方图的自适应融合,以适应每个视频帧的特定场景,并提升了跟踪算法的鲁棒性和准确性。

#### 3.1 响应图置信区域

相关滤波跟踪中,理想情况下每帧响应图呈均单尖峰状,响应图最大值对应位置即为当前帧跟踪结果位置。响应图的单尖峰越尖锐表示跟踪结果越可信;反之,单尖峰越平坦表示跟踪结果越不可信。然而,由于实际场景受到复杂因素的干扰,实际的响应图可能呈多峰状,正确的目标位置可能在响应图的最高峰,也可能会在次峰或其它峰<sup>[15]</sup>。

响应图中每个响应值的大小代表其对应位置是真实目标的可能性。响应图峰状越尖锐,其最大值相比响应图期望值也越大,峰状越平坦,其最

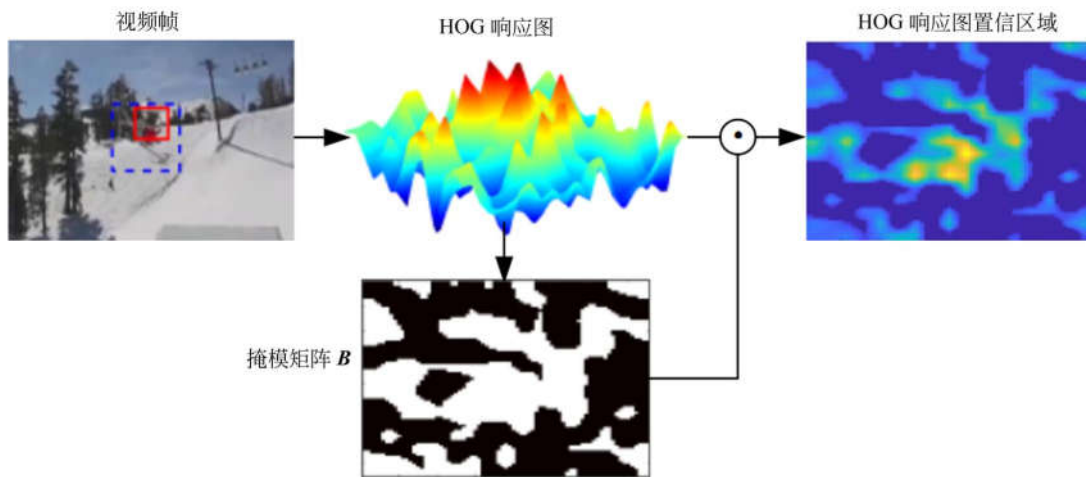


图 1 响应图置信区域示意图

Fig. 1 Schematic diagram of response map of confidence region

大值也越接近响应图期望值。响应图中大于期望值的区域被称为响应图置信区域  $F'$ ,其计算方法如式(8)所示,计算流程示意图如图 1 所示。

$$F' = F \odot B, \quad (8)$$

其中:  $B$  为掩模矩阵,其  $(i, j)$  位置元素值  $B(i, j)$  的计算方法如式(9)和式(10)所示:

$$B(i, j) = \begin{cases} 1, & F(i, j) > E_F \\ 0, & F(i, j) < E_F \end{cases}, \quad (9)$$

$$E_F = \sum_{i=1}^m \sum_{j=1}^n F(i, j) / m \times n, \quad (10)$$

其中:  $F(i, j)$  表示响应图矩阵  $(i, j)$  位置的元素值,响应图  $F$  的尺寸为  $m \times n$ ,  $E_F$  是响应图的期望。

### 3.2 自适应融合权重

HOG 特征响应图和颜色直方图响应图的自适应融合方法,如公式(11)所示:

$$F_{\text{fusion}} = \alpha F_{\text{HOG}} + (1 - \alpha) F_{\text{color}}, \quad (11)$$

其中: $\alpha$ 表示 HOG 响应图  $F_{\text{HOG}}$  自适应融合权重,  $(1 - \alpha)$  表示颜色直方图特征响应图  $F_{\text{color}}$  融合权重。 $F_{\text{color}}$  一般呈平坦状,其每个响应值代表对应位置像素是目标像素的可能性。 $F_{\text{fusion}}$  为融合后响应图。

根据响应图置信区域的计算方法,我们可以计算出当前帧 HOG 特征响应图  $F_{\text{HOG}}$  的置信区域  $F'_{\text{HOG}}$ 。融合权重  $\alpha$  的计算方法如式(12)所示:

$$\alpha = \frac{\sum_{i=1}^m \sum_{j=1}^n F'_{\text{HOG}}(i, j)}{\sum_{i=1}^m \sum_{j=1}^n F_{\text{HOG}}(i, j)}. \quad (12)$$

权重  $\alpha$  会随着 HOG 特征响应图峰状的尖锐程度而变化,峰状越尖锐则权重  $\alpha$  越大,表示 HOG 特征响应图的融合比例也越高,对应跟踪结果也更可信。也就是说,HOG 特征响应图越可信时,颜色直方图响应图的权重就会降低。反之,HOG 特征响应图不可信时,颜色直方图响应图的权重就会增加。

## 4 实验

本文算法基于 Matlab R2017b 实现,硬件实验环境为 Intel Core i7-6700 CPU、主频 3.4 GHz、内存 8 GB 配置的计算机。实验所用视频集为 OTB-2015<sup>[19]</sup> 基准视频集中的 100 个视频,其中包含 11 种复杂场景的挑战。对比算法包括 KCF<sup>[3]</sup>, Staple<sup>[17]</sup>, LMCF<sup>[15]</sup>, SRDCF<sup>[14]</sup>, SAMF<sup>[8]</sup> 等 5 种最近主流的跟踪方法。

### 4.1 定量分析

实验采用了视频跟踪领域常用的两个评价指标对相关算法进行定量估计:成功率曲线下面积 (Area Under Curve, AUC)、精度 (Precision)<sup>[20]</sup>。每一帧跟踪结果的评分为  $t = \text{area}(ROI_T \cap ROI_G) / \text{area}(ROI_T \cup ROI_G)$ , 其中  $ROI_T$  和  $ROI_G$  分别表示跟踪结果矩形框和正确位置标注框。 $t$  大于某一阈值的帧数占所有帧数的比例为该阈值对应的成功率。阈值  $[0, 1]$  对应的成功率所形成的曲线即为成功率曲线。每一帧的像素误差是跟

踪结果矩形框的中心点和正确位置标注框中心点的距离。一个视频序列中像素误差小于某个值的帧数占所有帧数的比例即为该值对应的 Precision。文中给出的 Precision 是像素误差为 20 pixel 对应的精度值。AUC 和 Precision 皆是值越大表示其算法性能越好。

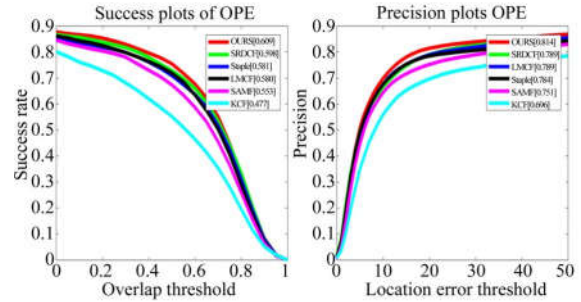


图 2 AUC 和 Precision 曲线对比图

Fig. 2 Curve comparison plot between AUC and Precision

图 2 是 6 种算法在 OTB-2015 视频集 100 个视频下综合 AUC 和 Precision 的曲线对比图。本文算法的综合 AUC 和 Precision 分别为 0.609 和 0.814, 皆优于其它 5 种对比算法,且相比 Staple 基础算法分别提升了 2.8% 和 3.0%。

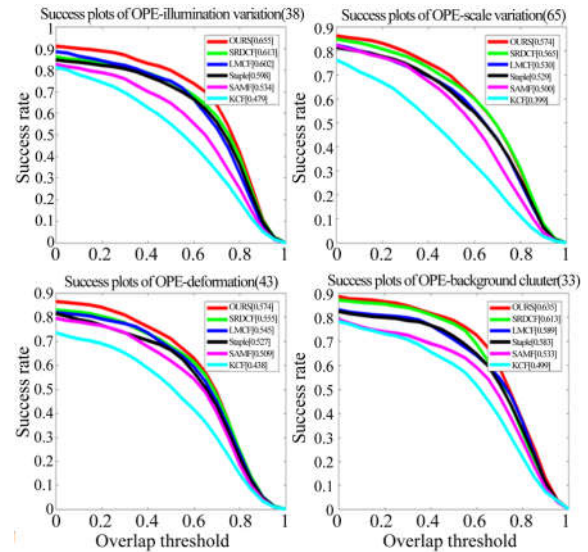


图 3 部分挑战环境下 AUC 曲线对比图

Fig. 3 Curve comparison plot of AUC in some challenging environment

OTB-2015 视频集包含了 11 种挑战性环境,分别是尺度变化(SV)、遮挡(OCC)、光照变化(IV)、运动模糊(MB)、形变(DEF)、快速运动(FM)、平面外旋转(OPR)、背景混乱(BC)、离开视野(OV)、平面内旋转(IPR)和低分辨率(LR)<sup>[19]</sup>。为了更好地分析本文算法的性能,对 11 种挑战性环境下算法的 AUC 和 Precision 进行了分析。图 3 和表 2 分别是各种挑战性环境下 AUC 值的对比曲线和对比数据(彩图见期刊电子版)。

表 2 11 种挑战性环境下 AUC 对比数据(红色表示最佳、蓝色表示次佳)

Tab.2 AUC values for 11 challenging attributes(top two results are shown in red and blue)

Challenge	OURS	KCF	Staple	LMCF	SRDCF	SAMF
IV	0.655	0.479	0.598	0.602	0.613	0.534
OPR	0.567	0.457	0.538	0.559	0.553	0.540
SV	0.574	0.399	0.529	0.530	0.565	0.500
OCC	0.568	0.443	0.548	0.556	0.559	0.540
DEF	0.574	0.438	0.555	0.527	0.545	0.509
MB	0.574	0.456	0.558	0.576	0.610	0.534
FM	0.574	0.454	0.541	0.555	0.599	0.509
IPR	0.573	0.467	0.551	0.543	0.543	0.518
OV	0.511	0.393	0.481	0.539	0.460	0.480
BC	0.635	0.499	0.583	0.613	0.589	0.533
LR	0.512	0.307	0.411	0.484	0.480	0.459

本文算法在 8 种挑战环境下优于其它算法,2 种环境次优,所有场景相比 Staple 算法都有不同程度地提升。尤其是在光照环境下,本文算法对 38 个光照视频的 AUC 达到 0.655,比 Staple 基础算法高出了 5.7%。特别地,本文对光照变化(IV)和其它 10 种挑战环境交叉情况下的跟踪结果进行了分析。表 3 是 10 种交叉挑战环境下 AUC 值的对比数据,本文算法在光照和形变(IV+DEF)环境下 AUC 达到了 0.681。图 4 和表 4 分别是各种挑战性环境下 Precision 值的对比曲

线和对比数据。本文算法在 5 种环境下优于其它算法,在 4 种环境下次优。尤其在光照环境下 Precision 达到了 0.847,比 Staple 基础算法高出了 5.6%。

表 3 10 种光照交叉环境下 AUC 对比数据(红色表示最佳、蓝色表示次佳)

Tab.3 AUC values for 10 mixtured challenging attributes with IV(top two results are shown in red and blue)

Challenge	OURS	KCF	Staple	LMCF	SRDCF	SAMF
IV+OPR	0.627	0.484	0.564	0.592	0.572	0.517
IV+SV	0.613	0.385	0.532	0.567	0.573	0.486
IV+OCC	0.590	0.470	0.579	0.570	0.580	0.527
IV+DEF	0.681	0.519	0.657	0.598	0.627	0.528
IV+MB	0.579	0.408	0.562	0.550	0.602	0.468
IV+FM	0.526	0.410	0.497	0.521	0.541	0.433
IV+IPR	0.573	0.411	0.482	0.533	0.485	0.424
IV+OV	0.417	0.403	0.422	0.555	0.430	0.483
IV+BC	0.608	0.462	0.528	0.607	0.560	0.482
IV+LR	0.354	0.171	0.307	0.429	0.333	0.355

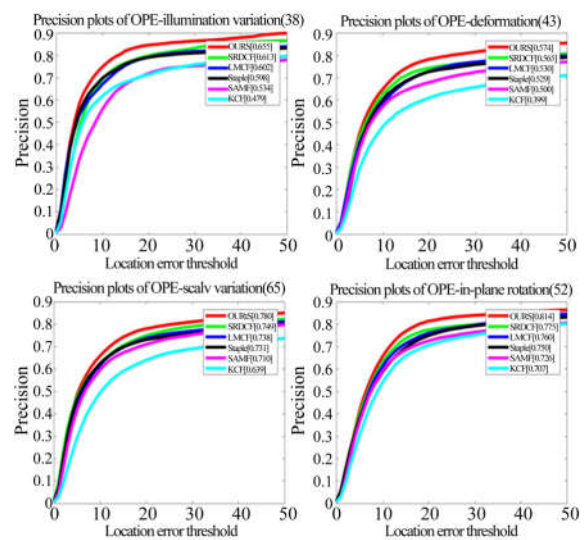


图 4 部分挑战环境下 precision 曲线对比图  
Fig.4 Curve comparison plot of precision some challenging environment

表 4 11 种挑战性环境下 precision 对比数据(红色表示最佳、蓝色表示次佳)

Tab. 4 Precision values for 11 challenging attributes(top two results are shown in red and blue)

Challenge	OURS	KCF	Staple	LMCF	SRDCF	SAMF
IV	0.847	0.719	0.791	0.798	0.792	0.715
OPR	0.782	0.682	0.742	0.771	0.746	0.743
SV	0.780	0.639	0.731	0.738	0.749	0.710
OCC	0.738	0.630	0.726	0.740	0.735	0.726
DEF	0.783	0.608	0.743	0.727	0.728	0.679
MB	0.732	0.618	0.726	0.747	0.782	0.677
FM	0.749	0.629	0.704	0.747	0.768	0.665
IPR	0.814	0.707	0.775	0.747	0.750	0.726
OV	0.662	0.501	0.661	0.693	0.597	0.628
BC	0.827	0.724	0.777	0.832	0.783	0.706
LR	0.686	0.546	0.609	0.705	0.631	0.710

## 4.2 定性分析

### 4.2.1 光照变化

图 5 的第 1~2 行列出了 Shaking 和 Kite-Surf 视频中部分视频帧跟踪结果。对于 Shaking 视频,开始阶段每个跟踪算法都能正确跟踪目标。56~60 帧背景环境出现了强光干扰,本文算法和 LMCF 算法可以正确跟踪,而 Staple, KCF, SAMF, SRDCF 逐步偏离目标。67 帧之后强光照射结束,由于目标模板的漂移,偏离目标的算法无法接续跟踪。对于 KiteSurf 视频,本文算法可以正确持续跟踪。而在 29 帧时, KCF, LMCF 开始偏离目标。之后目标在跳跃后开始低头,背景中伴随出现光照变化, Staple, LMCF 采用固定比例融合两种特征,跟踪情况非常不稳定。从结果来看,本文算法的自适应融合机制在光照环境下能够正确持续跟踪。

### 4.2.2 目标形变

图 5 的第 3~4 行列出了 Skiing 和 Panda 视频中部分视频帧跟踪结果。对于 Skiing 视频,目标快速移动过程中伴随着平面内旋转。在跟踪开

始阶段,除了本文算法能够正确跟踪, Staple, SRDCF, KCF, LMCF, SAMF 几种算法皆偏离了目标。其中, KCF 采用了 HOG 单一特征,由于特征本身的局限性无法正确跟踪; Staple, LMCF, SRDCF, SAMF 采用了多特征固定权重融合,固定权重无法自适应特定视频帧场景环境,导致跟踪失败。对于 Panda 视频,目标持续行走导致目标形变,本文算法能够保持正确持续跟踪。从开始跟踪到 657 帧的过程中,其它算法依次偏离目标直到彻底跟踪失败。因此,本文算法通过判别特征在特定视频帧场景环境的跟踪性能实现自适应融合,克服了目标形变的干扰。

### 4.2.3 背景混乱

图 5 的第 5~6 行列出了 Matrix 和 Couple 视频的部分视频帧跟踪结果。对于 Matrix 视频,背景混乱且与目标十分相似,目标也在不断改变姿势。44 帧之后,由于背景中雨滴和大量观众的干扰,除了本文算法之外 5 种算法均偏离目标, KCF, LMCF, SAMF 无法分辨跟踪目标和干扰目标,跟踪结果偏离至观众,而 Staple, SRDCF 偏离到目标衣服区域。对于 Couple 视频,只有本文算法和 SRDCF 保持了正确跟踪。由于混乱背景的干扰,18 帧时 SAMF 偏离目标,47 帧时 KCF, LMCF 偏离目标,97 帧时 Staple 偏离目标。由此可见,由于本文算法自适应融合特征保持了较为准确的目标模板,能够有效克服背景环境中混乱目标的干扰。

### 4.2.4 尺度变化

图 5 的第 7~8 行列出了 Vase 和 ClifBar 视频的部分视频帧跟踪结果。对于 Vase 视频,目标大小不断发生变化。在跟踪过程中, KCF, SAMF, SRDCF 跟踪结果的尺寸多次明显与实际目标不符。本文算法, Staple, LMCF 基本保持了正确的目标尺寸。对于 ClifBar 视频,目标尺寸不断发生变化且存在快速运动和运动模糊。227 帧时本文算法和 Staple 能够正确跟踪, LMCF, KCF, SAMF 较大程度偏离目标, SRDCF 有一定程度偏离目标。之后, LMCF 重新找到目标,而 KCF, SAMF 终偏离目标, SRDCF 跟踪结果尺寸不断缩小。因此,本文算法在目标尺度发生变化时也能够较好地适应。



图 5 部分视频帧跟踪结果

Fig. 5 Screenshots for some tracking results

#### 4.2.5 自适应融合权重

图 6 列出了以上光照变化、目标形变、背景混乱、尺度变化环境下代表性视频 Shaking, Panda, Couple, ClifBar 的自适应融合权重曲线和跟踪结果的 CLE(中心点像素距离误差)曲线对比图。本文算法根据 HOG 特征跟踪结果响应图自适应

计算特定视频帧场景下两种特征的融合权重, 自适应融合权重曲线随着特定视频帧场景而波动。自适应融合后的响应图最大值对应位置较好地逼近真实目标位置, CLE 曲线对比图中本文算法较贴近水平轴即本文算法跟踪结果与真实目标位置像素误差较小。

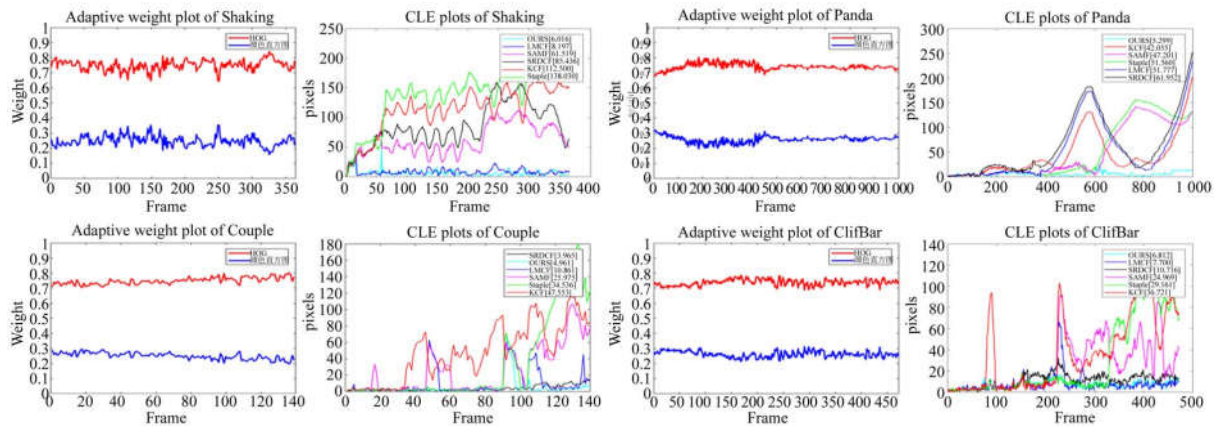


图 6 自适应权重图和 CLE 曲线对比图

Fig. 6 Adaptive weight plots and curve comparison plots of CLE

## 5 结 论

为了提高多特征对视频帧特定场景的适应性,本文采用响应图置信区域实现了多特征的自适应融合。响应图中高于响应图期望值的区域即响应图置信区域,根据 HOG 特征响应图置信区域可以计算出 HOG 特征响应图和颜色直方图特征响应图的融合权重,进而实现 HOG 和颜色直

方图特征的自适应融合。针对 OTB-2015 基准数据集的 100 个视频进行测试,本文算法的综合 AUC 和精度分别为 0.609 和 0.814,尤其在光照环境下 AUC 和精度分别达到 0.655 和 0.847,相比 Staple 分别提升 5.7% 和 5.6%。特别地,本文算法在光照和形变交叉环境下 AUC 达到 0.681。实验结果表明,在光照变化、目标形变、背景混乱、尺度变化等场景下,与近几年 5 种主流相关滤波跟踪算法相比,本文算法具备更优的跟踪性能。

## 参考文献:

- [1] BOLME D S, BEVERIDGE J R, DRAPER B A, *et al.*. Visual object tracking using adaptive correlation filters [C]. *IEEE conference Computer Vision and Pattern Recognition (CVPR)*, 2010 *IEEE Conference on*. IEEE, 2010: 2544-2550.
- [2] 刘教民, 郭剑威, 师硕, 等. 自适应模板更新和目标重定位的相关滤波器跟踪 [J]. *光学精密工程*, 2018, 26(8):2100-2111.  
LIU J M, GUO J W, SHI SH, *et al.*. Correlation filter tracking based on adaptive learning rate and location refiner [J]. *Opt. precision Eng.*, 2018, 26(8):2100-2111. (in Chinese)
- [3] 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.
- [4] 葛宝义, 左宪章, 胡永江, 等. 基于双步相关滤波的目标跟踪算法 [J]. *红外与激光工程*, 2018, 47(12):388-397.  
GE B Y, ZUO X Z, HU Y J, *et al.*. Object tracking algorithm based two-step correlation filter [J] *Infrared and laser engineering*, 2018, 47(12): 388-397. (in Chinese)
- [5] DANELLJAN M, HAGER AND G, KHAN F S, *et al.*. Accurate scale estimation for robust visual tracking [C]. *British Machine Vision Conference*, 2014: 1-5.
- [6] 杨德东, 毛宁, 杨福才, 等. 利用最佳伙伴相似性的改进空间正则化判别相关滤波目标跟踪 [J]. *光学精密工程*, 2018, 26(2):492-502.  
YANG D D, MAO N, YANG F C, *et al.*. Improved SRDCF object tracking via the Best-Buddies similarity [J]. *Opt. precision Eng.*, 2018, 26(2): 492-502. (in Chinese)
- [7] 张博, 江沸菠, 刘刚. 利用视觉显著性和扰动模型的上下文感知跟踪 [J]. *光学精密工程*, 2018, 26(8):2112-2121.  
ZHANG B, JIANG F B, LIU G, *et al.*. Context-aware tracking based on a visual saliency and perturbation model [J]. *Opt. precision Eng.*, 2018, 26

- (8):2112-2121. (in Chinese)
- [8] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration [C]. *European Conference on Computer Vision*, Springer, Cham, 2015:254-265.
- [9] BERNARDINI C, SILVERSTON T, FESTOR O. MPC: Popularity-based caching strategy for content centric networks[C]. *IEEE International Conference on Communications*, ., 2014:3619-3623.
- [10] 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 (CVPR)*. IEEE, 2014:1090-1097.
- [11] DANELLJAN M, HÄGER G, KHAN F S, *et al.*. Discriminative scale space tracking [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(8):1561-1575.
- [12] HENRIQUES J F, RUI C, MARTINS P, *et al.*. Exploiting the circulant structure of tracking-by-detection with kernels[C]. *Computer Vision - EC-CV 2012*. Springer Berlin Heidelberg, 2012: 702-715.
- [13] MA C, HUANG J B, YANG X, *et al.*. Hierarchical convolutional features for visual tracking [C]. *IEEE International Conference on Computer Vision*. IEEE Computer Society, 2015:3074-3082.
- [14] DANELLJAN, MARTIN. Learning spatially regularized correlation filters for visual tracking [J]. *IEEE International Conference on Computer Vision IEEE Computer Society*, 2015:4310-4318.
- [15] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps [C]. *IEEE conference Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [16] DANELLJAN M, BHAT G, KHAN F S, *et al.*. ECO: efficient convolution operators for tracking [J]. *arXiv: 1611.09224v1*, 2016:6931-6939.
- [17] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al.*. Staple: complementary learners for real-time tracking [C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.
- [18] POSSEGGER H, MAUTHNER T, BISCHOF H. In defense of color-based model-free tracking [C]. *IEEE conference Computer Vision and Pattern Recognition (CVPR)*, 2015:2113-2120.
- [19] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, 37(9):1834-1848.
- [20] EVERINGHAM M, GOOL L V, WILLIAMS C K I, *et al.*. *The pascal, visual object classes (VOC) Challenge* [J]. *International Journal of Computer Vision*, 2010, 88(2):303-338.

#### 作者简介:



高 贇(1980—),女,山西侯马人,博士,副教授,硕士生导师,2007年于西安电子科技大学获得硕士学位,2014年于云南大学获得博士学位,主要研究方向为视频运动目标跟踪。E-mail: gausegao@163.com



赵江珊(1995—),女,云南普洱人,硕士研究生,2013年于楚雄师范学院获得学士学位,主要研究方向为视频运动目标跟踪。E-mail:missandai@163.com