

# 基于概率模型的自适应融合互补学习跟踪算法

董秋杰<sup>1,2</sup>, 何雪东<sup>1,2\*\*\*</sup>, 葛海燕<sup>3\*\*</sup>, 周盛宗<sup>1\*</sup>

<sup>1</sup>中国科学院福建物质结构研究所, 福建 福州 350002;

<sup>2</sup>中北大学大数据学院, 山西 太原 030051;

<sup>3</sup>山东理工大学电气与电子工程学院, 山东 淄博 255049

**摘要** 在互补学习实时跟踪算法(Staple)中,方向梯度直方图(HOG)特征与颜色直方图采用的融合系数均为固定值 0.3,在不同特征下相融时易造成目标丢失的问题。基于此,提出一种基于目标概率模型的自适应融合互补学习实时跟踪算法(amStaple),该算法使用分段函数得出自适应融合系数。分别在 OTB-2013 与 OTB-100 基准视频集上对所提算法进行实验测试,最终的实验结果显示,本文算法极大地提升了跟踪器性能,与 Staple 相比,在两个基准数据集上其精度分别高出 6.52% 与 3.32%,成功率分别高出 4.89% 与 3.11%。本文算法较为简单,且在与同时期优秀算法的定性定量比较中表现较优。为解决本文提出的算法在基准视频部分属性上表现欠佳的问题,在本文算法的基础上增加判定条件,提出了 amStaple1 算法。

**关键词** 机器视觉; 视觉跟踪; 概率模型; 融合系数; 分段函数

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.161505

## Adaptive Merging Complementary Learners for Visual Tracking Based on Probabilistic Model

Dong Qiujie<sup>1,2</sup>, He Xuedong<sup>1,2\*\*\*</sup>, Ge Haiyan<sup>3\*\*</sup>, Zhou Shengzong<sup>1\*</sup>

<sup>1</sup>Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian 350002, China;

<sup>2</sup>School of Data Science, North University of China, Taiyuan, Shanxi 030051, China;

<sup>3</sup>College of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, Shandong 255049, China

**Abstract** In complementary learners for real-time tracking known as Staple, the merging coefficients of histogram of oriented gradient feature and color histogram have both a fixed value of 0.3, which can easily cause the problem of losing target when they are merged under different features. To solve this problem, this study proposes an adaptive merging algorithm of complementary learners for real-time visual tracking based on an object probabilistic model known as amStaple, which uses a piecewise function to obtain the adaptive merging coefficient. Experiments on popular object tracker benchmarks including OTB-2013 and OTB-100 verify the effectiveness of the proposed algorithm. Results show that amStaple has better performance than Staple. Compared with Staple in terms of OTB-2013 and OTB-100, amStaple has 6.52% and 3.32% higher precision and 4.89% and 3.11% higher success rates, respectively. Although the proposed algorithm is relatively less innovative, its performance has been obviously improved in various aspects compared with that of a state-of-the-art algorithm from the same period. However, amStaple performs poorly on partial sequence attributes of object tracker benchmarks. To solve this problem, a decision condition is added based on amStaple, which is called amStaple1.

**Key words** machine vision; visual tracking; probabilistic model; merging coefficient; piecewise function

**OCIS codes** 150.1135; 100.4999; 330.7326

收稿日期: 2019-03-02; 修回日期: 2019-03-14; 录用日期: 2019-03-25

基金项目: 福建省中科院 STS 计划配套项目(2019T31020008, 2019T31020010)、福建省科学技术厅技术项目(cyzx201805063)

\* E-mail: zhousz@fjirsm.ac.cn; \*\* E-mail: 1178163538@qq.com; \*\*\* E-mail: 849459832@qq.com

# 1 引言

视觉跟踪技术是计算机视觉研究领域中的热点<sup>[1-7]</sup>与难点,已广泛应用于人机交互、公共安全、医学影像等领域。视觉跟踪从图像处理角度可以视为将从初始图像帧中得到的目标位置作为已知信息,在后续所有图像帧中预测目标位置,从而实现持续的视觉跟踪。

在 21 世纪初之前,视觉跟踪技术主要是使用生成式模型<sup>[8]</sup>,其性能较低。随后由于计算机视觉技术的发展,尤其是视觉检测技术的不断完善,人们发现将传统的视觉检测算法与视觉跟踪问题相结合可以得到更佳性能,所以基于“tracking by detection”的判别式模型开始取代基于统计迭代的生成式模型。

2010 年 Bolme 等<sup>[9]</sup>将相关滤波技术引入视觉跟踪领域,提出了最小化输出平方误差总和(MOSSE)算法,由此开启了视觉跟踪的相关滤波时代。2012 年 Henriques 等<sup>[10]</sup>提出了基于密集采样技术的核相关滤波(CSK)算法,该算法同时保证了跟踪器的速度与精度。2014 年 Danelljan 等<sup>[11]</sup>在 CSK 使用单一灰度特征的基础上增加使用了 11 维的颜色属性(CN)特征,同时使用主成分分析(PCA)对高维的 CN 特征进行降维。Henriques 等<sup>[12]</sup>进一步完善 CSK 算法,同时使用方向梯度直方图(HOG)特征替代灰度特征,提出了核相关滤波(KCF)算法。为解决目标尺度变化问题,Danelljan 等<sup>[13]</sup>提出了使用尺度金字塔的精确尺度估计(DSST)算法。为解决 DSST 速度问题,Danelljan 等<sup>[14]</sup>于 2017 年提出了使用 PCA 降维的快速精确尺度估计(fDSST)算法。

上述算法虽然在跟踪性能上表现较优,但是其都使用单一目标特征。大部分判别器模型依赖于目标的空间信息,当目标发生快速变化时表现较差。一个最简单的方法是寻找一种对目标变形不敏感的表现方法,而颜色属性在此方面具有天然的表现优势,因为目标发生快速运动后虽然空间信息发生了较大的变化,但是其颜色信息未发生改变。当目标与背景颜色相近时,颜色特征表现较差,而此时 HOG 特征可以从空间信息上解决此问题。因此,2015 年 Li 等<sup>[15]</sup>将 HOG 与 CN 进行集成,提出了多特征的尺度自适应(SAMF)算法。2015 年 Danelljan 等<sup>[16]</sup>提出了空间正则化相关滤波(SRDCF)算法,该算法使用了 HOG 与 CN 特征。2016 年 Bertinetto 等<sup>[17]</sup>基于 HOG 与颜色直方图

特征,提出了互补性实时跟踪(Staple)算法。

SAMF、SRDCF 以及 Staple 虽然均使用 HOG 与颜色特征来增强目标的外观特征,但是在多特征融合方面的处理效果并不理想。以 Staple 为例,其在 HOG 与颜色直方图特征响应融合时使用了固定融合系数,作者在经过实验后选定 0.3 作为融合系数,但是在不同视频帧中采用固定融合系数并不是最优方案,严重时会造成目标丢失。针对此,本文提出了一种使用目标概率模型的自适应融合互补学习实时跟踪算法,本算法通过实验设定分段函数自变量,也就是基于颜色直方图的目标概率的阈值。在 OTB-2013<sup>[18]</sup>与 OTB-100<sup>[19]</sup>基准视频集上对本文算法进行实验测试,效果显著。

## 2 互补学习实时跟踪算法

Staple 是在 DSST 基础上融合了多特征的相关滤波视觉跟踪算法,使用了 HOG 特征与颜色直方图特征,这样既解决了当目标空间信息发生快速变化时,单独使用 HOG 特征会导致判别器更新不及时,从而造成目标丢失的问题,又解决了当目标与背景颜色相近时,单独使用颜色特征会导致判别器信息提取错误,造成跟踪失败的问题。

### 2.1 基于 HOG 特征的核相关滤波判别器

采用相关滤波判别器训练模板  $\omega$ ,以找寻基于 HOG 特征的判别器函数  $f_{cf}(x) = \omega^T x$ ,使其对样本的 HOG 特征矩阵  $x_i$  及其期望输出  $y_i$  满足损失函数  $L[f(x_i), y_i]$  最小,即

$$\min_{\omega} \sum_i \{L[f(x_i), y_i] + \lambda \|\omega\|^2\}, \quad (1)$$

式中:  $\lambda$  为正则化系数,它可防止发生过拟合;  $\|\cdot\|$  为 L2 范式;  $i$  为第  $i$  帧图像。

对于损失函数  $L[f(x_i), y_i]$ ,使用岭回归方法求解,这是一种改良的最小二乘方法,其性能与支持向量机(SVM)一致,并且具有简单闭环解<sup>[20]</sup>,则

$$L[f(x_i), y_i] = [f(x_i) - y_i]^2. \quad (2)$$

将(2)式代入(1)式中得到

$$\min_{\omega} \sum_i \{L[f(x_i), y_i]^2 + \lambda \|\omega\|^2\}. \quad (3)$$

根据文献[20],(3)式的闭环解为

$$\omega = (x^T X + \lambda I)^{-1} x^T y, \quad (4)$$

式中:  $X$  为由样本 HOG 特征矩阵  $x_i$  组成的循环矩阵;  $y$  为由期望输出  $y_i$  组成的向量;  $I$  为与  $x^T X$  同维度的单位矩阵。

将  $\omega$  映射到非线性空间,将优化变量由  $\omega$  变为  $\alpha$ ,并根据文献[20]提出的核技巧,将(4)式转化为

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}, \quad (5)$$

式中:  $\boldsymbol{\alpha}$  为  $\boldsymbol{\omega}_i$  映射到非线性空间中的对应分量  $\boldsymbol{\alpha}_i$  组成的向量;  $\mathbf{K}$  为核矩阵。

由文献[12]可知,  $\mathbf{K}$  为对称矩阵, 可进行对角化。则(5)式转换到离散傅里叶域(DFT)内求得模板为

$$\hat{\boldsymbol{\alpha}} = \frac{\mathbf{y}}{\hat{\mathbf{K}} + \lambda \mathbf{I}}, \quad (6)$$

式中: “ $\hat{\cdot}$ ”为相关矩阵的对应 DFT 矢量。

对于核矩阵  $\mathbf{K}$ , 有点乘核与径向基函数(RBF)核这两种主流形式, 二者的性能几乎一致, 但是点乘核的参数多于 RBF 核, 因此大部分研究人员使用 RBF 核。而 RBF 核中使用最为广泛的为高斯核, 其形式为

$$\mathbf{K}_{xz} =$$

$$\exp\left\{-\frac{1}{\sigma^2} [\|\mathbf{x}\|^2 + \|\mathbf{z}\|^2 - 2\mathcal{F}^{-1}(\hat{\mathbf{x}} \odot \hat{\mathbf{z}}^*)]\right\}, \quad (7)$$

式中:  $\mathbf{z}$  为预测帧图像 HOG 特征矩阵;  $\sigma$  为高斯带宽;  $\mathcal{F}^{-1}$  为傅里叶逆变换;  $\hat{\mathbf{z}}^*$  为  $\hat{\mathbf{z}}$  的共轭;  $\odot$  为 Hadamard 积。

在实际使用过程中, 为了增加跟踪算法的稳健性, 使  $\mathbf{K}_{xz}$  的输出更平滑, 一般要进行归一化处理, (7)式变为

$$\mathbf{K}_{xz} =$$

$$\exp\left\{-\frac{1}{\sigma^2} [\|\mathbf{x}\|^2 + \|\mathbf{z}\|^2 - 2\mathcal{F}^{-1}(\hat{\mathbf{x}} \odot \hat{\mathbf{z}}^*)]/n\right\}, \quad (8)$$

式中:  $n$  为  $\mathbf{x}$  中元素的数目。

使用在线更新策略更新模板, 同时保留之前帧的模板参数, 即

$$\hat{\boldsymbol{\alpha}}(i)_{\text{cf}} = \begin{cases} \hat{\boldsymbol{\alpha}}, & i = 1 \\ (1 - \eta_{\text{cf}}) \hat{\boldsymbol{\alpha}}_{\text{cf}}^{(i-1)} + \eta_{\text{cf}} \hat{\boldsymbol{\alpha}}, & i > 1 \end{cases}, \quad (9)$$

式中:  $\boldsymbol{\alpha}_{\text{cf}}^{(i)}$  为当前模板;  $\boldsymbol{\alpha}_{\text{cf}}^{(i-1)}$  为之前模板;  $\eta_{\text{cf}}$  为学习率。

从而得到基于 HOG 特征的核相关滤波判别器在当前帧的时域响应矩阵为

$$\mathbf{r}_{\text{cf}} = \mathcal{F}^{-1}(\hat{\boldsymbol{\alpha}}(i)_{\text{cf}} \odot \hat{\mathbf{K}}_{xz}). \quad (10)$$

## 2.2 基于颜色直方图特征的判别器

采用颜色直方图判别器训练模板  $\boldsymbol{\beta}$ , 以找寻基于颜色直方图特征的判别器函数  $f_{\text{ch}}(\mathbf{x}; \boldsymbol{\beta}) = g(\psi_{\mathbf{x}}; \boldsymbol{\beta})$ , 使得样本  $\mathbf{x}$  及其期望输出  $\mathbf{y}$  找到跟踪搜索框  $s$ , 即

$$\min_s \left[ \boldsymbol{\beta}^T \left( \sum_u \psi_{\mathbf{x}}[u] \right) - \mathbf{y} \right]^2, \quad (11)$$

式中:  $u$  为样本  $\mathbf{x}$  的像素点;  $\psi_{\mathbf{x}}[u]$  为将样本  $\mathbf{x}$  转换成颜色直方图特征的函数, 其表现形式为

$$\psi_{\mathbf{x}}[u] = \mathbf{e}_{k[u]}, \quad (12)$$

式中:  $\mathbf{e}_{k[u]}$  是索引为  $k[u]$  的向量; 其中  $k[u]$  表现形式为

$$k[u] = \text{floor}\left(\frac{u \cdot n_{\text{hist}}}{256}\right) + 1, \quad (13)$$

式中:  $n_{\text{hist}}$  为颜色直方图数目;  $\text{floor}(\cdot)$  为向下取整函数。

为保持实时性, 独立地对图像中的每个像素进行回归计算, 对于前景区域  $\mathbf{A} \subset \mathbf{Z}^2$  与背景区域  $\mathbf{B} \subset \mathbf{Z}^2$ , (11)式的具体形式为

$$\frac{1}{|\mathbf{A}|} \sum_{u \in \mathbf{A}} (\boldsymbol{\beta}^T \psi[u] - 1)^2 + \frac{1}{|\mathbf{B}|} \sum_{u \in \mathbf{B}} (\boldsymbol{\beta}^T \psi[u])^2. \quad (14)$$

对于每一个特征维度, 将目标分解为独立的项, 则(14)式变为

$$\sum_{j=1}^M \left[ \frac{\Omega_j(\mathbf{A})}{|\mathbf{A}|} \cdot (\boldsymbol{\beta}_j - 1)^2 + \frac{\Omega_j(\mathbf{B})}{|\mathbf{B}|} \cdot (\boldsymbol{\beta}_j)^2 \right], \quad (15)$$

式中:  $M$  为特征维度数目;  $\Omega_j(\mathbf{o})$  为在区域  $\mathbf{o}$  内, 特征  $j$  索引非零的像素数。

对(15)式所示的相关岭回归问题进行求解, 得

$$\boldsymbol{\beta}_j = \frac{\Phi_j(\mathbf{A})}{\Phi_j(\mathbf{A}) + \Phi_j(\mathbf{B}) + \boldsymbol{\delta}}, \quad (16)$$

式中:  $\Phi_j(\mathbf{o})$  为目标图像在区域  $\mathbf{o}$  内的匹配值;  $\boldsymbol{\delta}$  为非零数矩阵。

预测帧图像中像素点  $u$  是所要跟踪目标中心位置的概率值为

$$p_u = \beta_u \psi[u], \quad (17)$$

其中所有  $p_u$  组成稀疏矩阵  $\mathbf{p}$ 。在矩阵  $\mathbf{p}$  上进行积分运算, 可以得到原始图像中左上角与当前点所围成的矩形区域内所有概率值  $p_u$  之和  $Q_u$ , 并可以更加突出样本图像中间区域的像素值, 即

$$Q_u = \int p_u du. \quad (18)$$

其中所有  $Q_u$  组成矩阵  $\mathbf{Q}$ 。在积分图像上截取出与跟踪搜索框  $s$  (宽度  $w$ , 高度  $h$ ) 大小一致的区域  $\mathbf{S}$ , 求得该区域内所有概率值  $p_u$  之和, 即为该区域为跟踪目标区域的概率值, 即

$$\begin{aligned} Q_{\mathbf{S}} &= \mathbf{Q}(m, n) + \mathbf{Q}(m + w, n + h) - \\ &\quad \mathbf{Q}(m + w, n) - \mathbf{Q}(m, n + h), \end{aligned} \quad (19)$$

式中:  $(m, n)$  为截取  $\mathbf{S}$  的左上角坐标。

将(19)式进行归一化处理,即为基于颜色直方图特征的判别器在当前帧的时域响应矩阵

$$\mathbf{r}_{\text{hist}} = \mathbf{Q}_s / (\omega \times h). \quad (20)$$

### 2.3 判别器响应融合

采用线性加权方式将上述两个判别器响应进行融合,得整体算法响应为

$$\mathbf{r} = (1 - \gamma)\mathbf{r}_{\text{cf}} + \gamma\mathbf{r}_{\text{hist}}, \quad (21)$$

式中: $\gamma$ 为判别器响应融合系数,在大多数算法中此系数没有根据具体特征响应进行自适应融合,采用固定值。

## 3 自适应融合

由2.2节(17)式可知,矩阵 $\mathbf{p}$ 是由 $p_u$ 组成的概率值矩阵。当颜色直方图判别器认为像素点 $p_u$ 为所要跟踪的目标中心位置时, $p_u$ 的值会增大,从而使矩阵 $\mathbf{p}$ 中所有像素点的概率总和增大。以矩阵 $\mathbf{p}$ 作为本文算法的目标概率模型,则判别器自适应融合系数 $\gamma$ 是关于 $\mathbf{p}$ 的函数,即

$$\gamma = \varphi(\mathbf{p}). \quad (22)$$

### 3.1 均值化自适应融合系数

$p_u$ 为预测帧图像中像素点 $u$ 是所要跟踪目标中心位置的概率值,因此 $p_u$ 越大,说明基于颜色直方图判别器认为该像素点为所要跟踪目标中心位置的概率越大,此时应该增加对 $\mathbf{r}_{\text{hist}}$ 的关注,因此对于(22)式使用最简单的均值化函数,即

$$\gamma = \varphi(\mathbf{p}) = \frac{1}{n_u} \sum_{u=1}^{n_u} p_u, \quad (23)$$

式中: $n_u$ 为在目标区域内的像素数

### 3.2 指数自适应融合系数

由于颜色直方图特征只关注像素的颜色特征,而忽略了像素的位置特征,因此,当目标颜色特征与背景颜色相近甚至一致时, $p_u$ 的值同样会增大,但此时应该尽量减小对于 $\mathbf{r}_{\text{hist}}$ 的关注,以避免将背景信息学习到模板中,造成模板污染。所以此时(22)式使用指数函数,即

$$\gamma = \varphi(\mathbf{p}) = \exp\left(-\frac{1}{n_u} \sum_{u=1}^{n_u} p_u\right). \quad (24)$$

由于颜色直方图忽略了像素的位置特征,因此HOG特征的表现能力优于颜色直方图特征,因此整体算法响应注重关注响应矩阵 $\mathbf{r}_{\text{cf}}$ , $\gamma$ 的值应小于0.5。所以(24)式变为

$$\gamma = \varphi(\mathbf{p}) = \exp\left(-\frac{1}{n_u} \sum_{u=1}^{n_u} p_u\right) / \xi, \quad (25)$$

式中: $\xi$ 为一个经验值,其保证得到的融合系数 $\gamma$ 的值小于0.5。

### 3.3 自适应融合系数 1

当颜色直方图判别器认为像素点 $p_u$ 为所要跟踪目标的中心位置时, $p_u$ 的值会增大,从而使矩阵 $\mathbf{p}$ 中所有像素点的概率总和增大,则相应的矩阵 $\mathbf{p}$ 的平均值也会增大,令

$$\kappa = \frac{1}{n_u} \sum_{u=1}^{n_u} p_u, \quad (26)$$

式中: $\kappa$ 为目标区域内像素点为目标的概率的均值。

由3.1节与3.2节可知,必须寻找阈值 $\epsilon$ ,当 $\kappa < \epsilon$ 时选择均值化自适应融合系数,当 $\kappa \geq \epsilon$ 时,选择指数自适应融合系数,即

$$\gamma = \varphi(\mathbf{p}) = \begin{cases} \kappa, \kappa < \epsilon \\ \exp(-\kappa) / \xi, \kappa \geq \epsilon \end{cases}. \quad (27)$$

### 3.4 自适应融合系数 2

3.3节所得到的自适应融合系数在整体效果上是比较优的,但是在目标出现快速运动、运动模糊以及跟踪图像分辨率较低时,跟踪效果欠佳,因此本文在(27)式的基础上增加判定条件,设置阈值 $\theta$ ,即

$$\gamma = \varphi(\mathbf{p}) = \begin{cases} \kappa, \text{others} \\ \exp(-\kappa) / \xi, \kappa \geq \epsilon, \zeta \geq \theta \end{cases}, \quad (28)$$

式中: $\theta$ 为一个经验值; $\zeta$ 为响应 $\mathbf{r}_{\text{hist}}$ 的均值。

图1为amStaple算法的整体框架图,amStaple算法分别从traing图像与testing图像中采集HOG特征与颜色直方图特征,然后分别进行相关性匹配,得到两者的特征响应图;依据(27)式或(28)式得到自适应融合系数;使用获得的自适应融合系数将HOG特征响应图与颜色直方图特征响应图进行融合,最后得到输出响应图。

## 4 实验与结果分析

本算法是以Staple算法为基础提出的,因此本算法主要与Staple算法进行对比。但为验证本文amStaple算法的性能,将其与同时期优秀算法进行定性定量比较。

### 4.1 实验参数与评估标准

本算法基本参数设置如表1所示。本实验的软件平台为MATLAB R2018b,硬件平台为Intel(R) Core(TM) i7-4790CPU@3.60 GHz,12 GB内存台式机。

本实验数据集为OTB-2013基准视频集与OTB-100基准视频集。其中:OTB-2013中有51个基准视频,包含11种基本视频属性;OTB-100中有100个基准视频,视频属性较OTB-2013更加复杂,

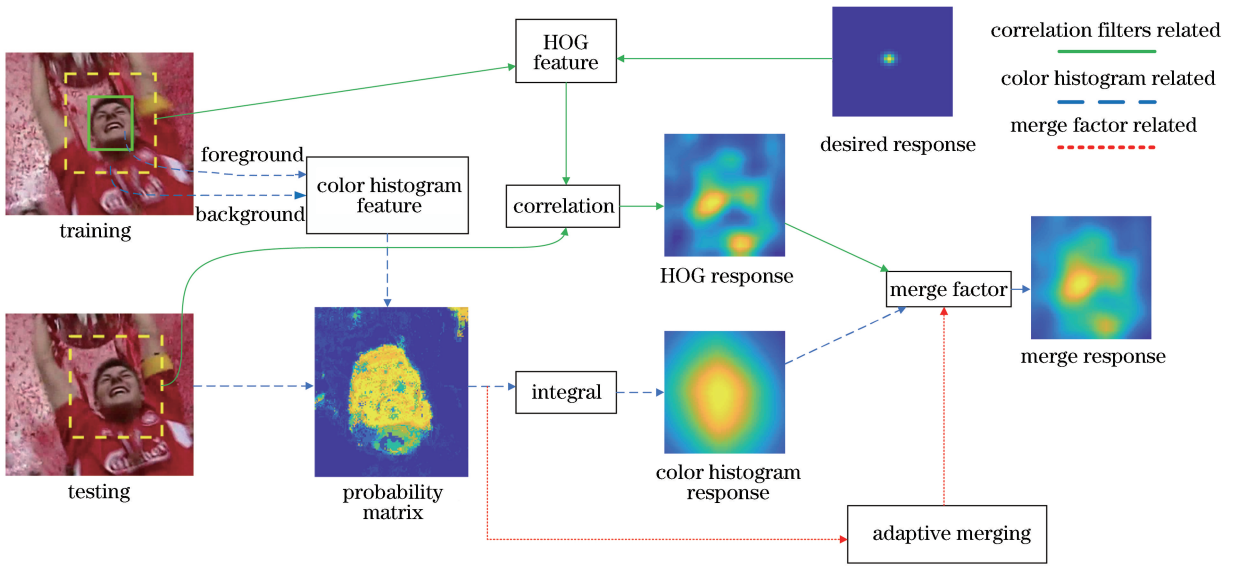


图1 amStaple 算法框架图

Fig. 1 Framework of amStaple algorithm

表 1 基本实验参数设置

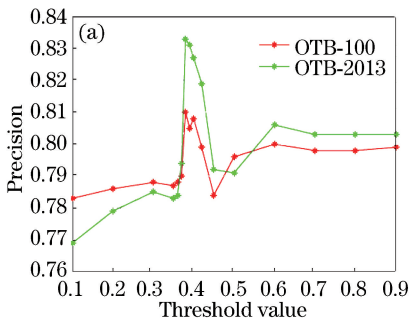
Table 1 Basic experimental parameter settings

Parameter	Value
HOG cell size	$4 \times 4$
HOG orientations	9
Learning rate (correlation filtering) $\eta_{cf}$	0.01
# bins colour histograms $n_{hist}$	$2^5 \times 2^5 \times 2^5$
Hyperparameter $\delta$	$2.22 \times 10^{-16}$
Hyperparameter $\xi$	2
Threshold $\theta$	0.5

跟踪难度更大。

#### 4.2 阈值 $\epsilon$ 的取值

3.3 节中提出寻找阈值  $\epsilon$  可实现对均值化自适应



应融合系数和指数自适应融合系数的有效选择。通过实验由粗到细地寻找使算法具有最佳性能的阈值  $\epsilon$ 。

首先在范围(0,1)内,以 0.1 步长寻找具有最佳性能的阈值  $\epsilon$ ,由实验得到最佳性能的阈值  $\epsilon = 0.4$ ;然后以 0.4 为中心设置阈值池(0.35, 0.38, 0.40, 0.42, 0.45),从阈值池中选取阈值  $\epsilon$ ,通过实验得到最佳性能的阈值  $\epsilon = 0.38$ ;最后更新阈值池为(0.36, 0.37, 0.38, 0.39),由实验得到最佳性能的阈值  $\epsilon = 0.38$ 。图 2 给出了所有阈值  $\epsilon$  在 OTB-2013 与 OTB-100 上的精度比较曲线与成功率比较曲线。

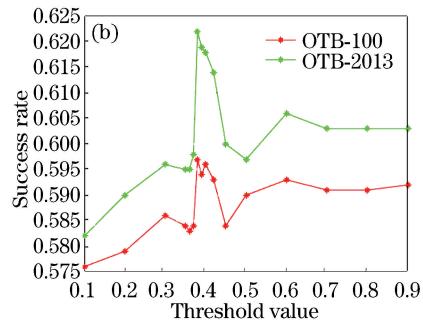


图 2 所有阈值  $\epsilon$  在不同视频集上的跟踪结果。(a)精度曲线;(b)成功率曲线

Fig. 2 Tracking results of all thresholds  $\epsilon$  on different benchmarks. (a) Precision curves; (b) success rate curves

#### 4.3 amStaple 与 Staple 性能对比

本文算法 amStaple 是基于 Staple 算法进行改进的,所以本文对 amStaple 与 Staple 进行定量、定性比较分析。本节主要对 amStaple 与 Staple 进行定量比较分析。实验结果如表 2、表 3 所示。

由表 2 可知,本文所提算法 amStaple 在 OTB-2013 与 OTB-100 上精度与成功率均超过 Staple 算法;在 OTB-2013 上精度与成功率分别提高 6.52% 与 4.89%,在 OTB100 上精度与成功率分别提高 3.32% 与 3.11%。另外,在 OTB-2013 上二者每秒

表2 amStaple与Staple在不同视频集上的精度、成功率及amStaple较Staple提高百分比  
Table 2 Precision and success rates of amStaple and Staple on different benchmarks and percentage increases of amStaple compared to Staple

Benchmark	Criterion	Staple	amStaple	Increase ratio /%
OTB-2013	Precision	0.782	0.833	6.52
	Success rate	0.593	0.622	4.89
OTB-100	Precision	0.784	0.810	3.32
	Success rate	0.579	0.597	3.11

表3 amStaple与Staple在OTB-100视频集上不同属性视频的成功率比较

Table 3 Comparison of success rates of amStaple and Staple on different attributes of OTB-100 benchmark

Attribute	Success of Staple	Success of amStaple	Increase ratio /%
Illumination variation	0.595	0.628	5.55
Out-of-plane rotation	0.534	0.555	3.93
Scale variation	0.520	0.546	5.00
Occlusion	0.543	0.563	3.68
Deformation	0.550	0.559	1.64
Motion blur	0.540	0.558	3.33
Fast motion	0.541	0.557	2.96
In-plane rotation	0.549	0.573	4.37
Out-of-view	0.476	0.507	6.51
Background clutter	0.561	0.601	7.13
Low resolution	0.399	0.405	1.50

传输帧数(FPS)均为82.3,在OTB-100上二者FPS均为78.8。这也就是说明本文所提算法amStaple在没有损伤跟踪速度的情况下提高了精度与成功率,效果显著。

由表3可知,amStaple在OTB-100的11个视频属性上均超过Staple,提高百分比最低的是低分辨率情况下,为1.50%,提高百分比最高的是背景杂乱情况下,为7.13%。

#### 4.4 amStaple与同期优秀算法的性能对比

为进一步验证amStaple算法的性能,选取8个同期的优秀算法与其进行对比,所选算法分别为CSK、CN、DSST、SAMF、KCF、Staple、SRDCF、fDSST。图3给出了9个算法在OTB-2013与OTB-100上的精度曲线与成功率曲线。

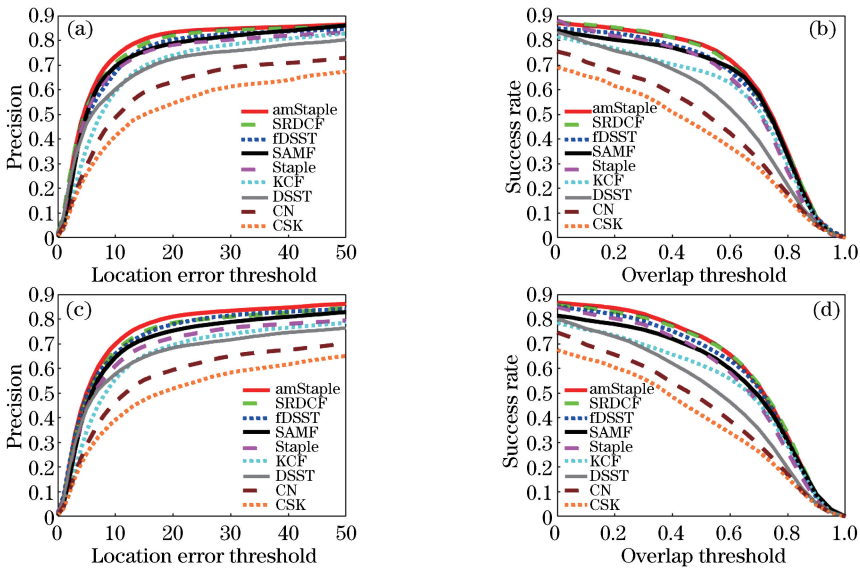


图3 9种算法在不同视频集上的跟踪结果。(a) OTB-2013视频集上的精度;(b) OTB-2013视频集上的成功率;(c) OTB-100视频集上的精度;(d) OTB-100视频集上的成功率

Fig. 3 Tracking results of nine algorithms on different benchmarks. (a) Precision for OTB-2013 benchmark; (b) success rates for OTB-2013 benchmark; (c) precision for OTB-100 benchmark; (d) success rates for OTB-100 benchmark

如图3可知,amStaple与8个优秀算法相比,整体性能较优。amStaple只在OTB-2013上性能略低于SRDCF,但是在OTB-100上其整体性能超越SRDCF,其中精度高出2.79%,成功率二者持平。

但是SRDCF并不满足实时性要求,amStaple的FPS是SRDCF的12.92倍。

图4给出了在OTB-100上amStaple性能最佳的7个视频属性,分别为光照变化、平面外旋转、遮

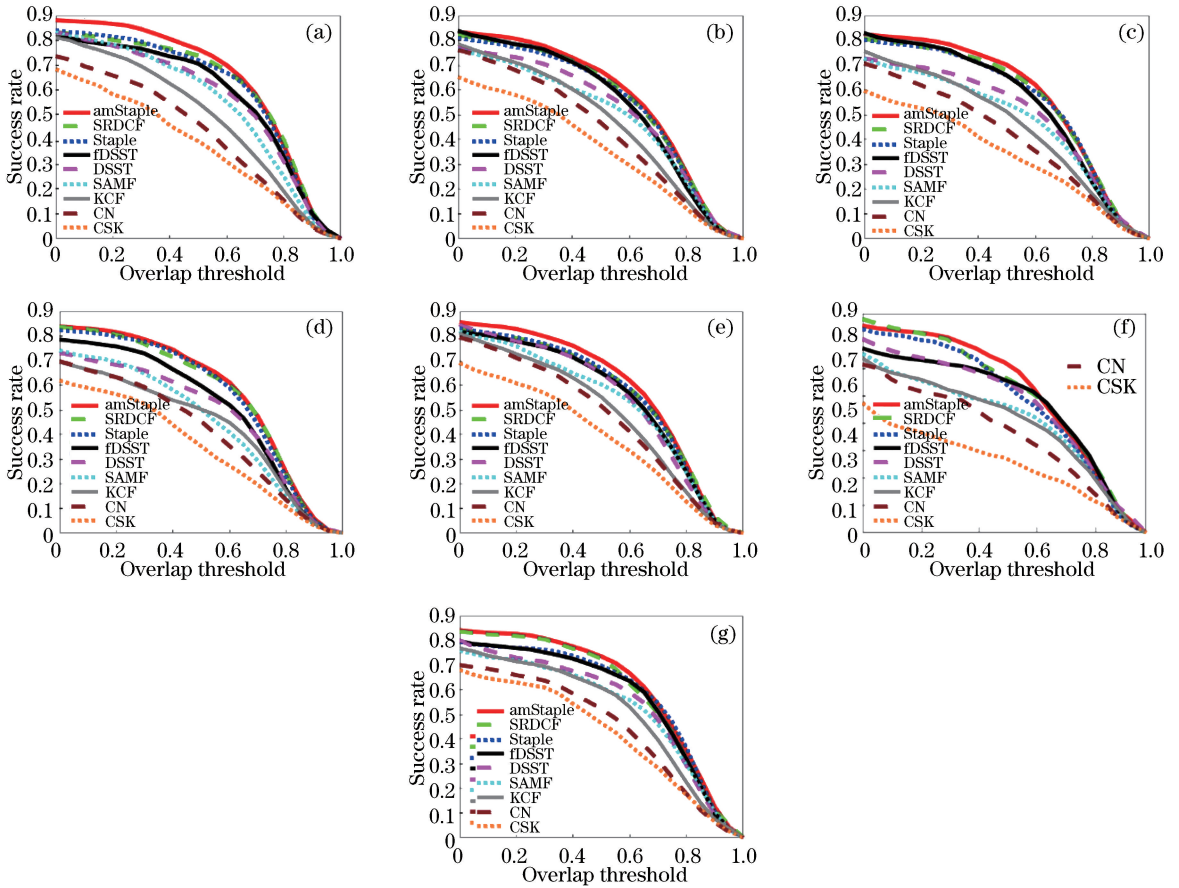


图 4 amStaple 在 OTB-100 上性能最佳的 7 个视频属性。(a)光照变化;(b)平面外旋转;(c)遮挡;(d)变形;  
(e)平面内旋转;(f)出视野;(g)背景杂乱

Fig. 4 Seven attributes in OTB-100 benchmark sequence where amStaple has best performance. (a) Illumination variation; (b) out-of-plane rotation; (c) occlusion; (d) deformation; (e) in-plane rotation; (f) out-of-view; (g) background clutter

遮挡、变形、平面内旋转、出视野、背景杂乱。

由表 3 可知,在 OTB-100 中共有 11 个视频属性,而 amStaple 在 7 个视频属性上保持性能最佳,这从另一方面说明 amStaple 的跟踪效果较优。

由图 3 可知,9 种算法中排名前三的是 amStaple、SRDCF、Staple,所以对这 3 个算法在视频集上进行定性比较分析,实验结果如图 5 所示。

由图 5 可知,本文所提算法 amStaple 较 Staple 改进效果显著,与 SRDCF 相比也不逊色,并且在一些视频属性上,amStaple 的表现是最佳的。

#### 4.5 amStaple 与 amStaple1 性能对比

在 3.4 节中提到,amStaple 在整体效果上是较优的,但是在目标出现快速运动、运动模糊以及跟踪图像分辨率较低时,跟踪效果欠佳。因此增加判定条件,提出算法 amStaple1。

图 6 给出了 amStaple1 与 amStaple 相比在 OTB-2013 与 OTB-100 上表现较佳的视频属性成功率曲线图,图 7 给出了在视频集 Deer 上 amStaple

与 amStaple1 的跟踪效果图。

Deer 视频集包含快速运动、运动模糊、低分辨率三大视频属性,由图 6 和图 7 可知,算法 amStaple1 在解决快速运动、运动模糊、低分辨率三大视频属性上性能优于 amStaple。

## 5 结 论

为解决视觉跟踪算法中多特征响应融合使用固定融合系数,降低跟踪器性能的问题,创新性地使用分段函数来处理不同目标概率下的特征融合问题,提出了一种新颖的基于目标概率模型的自适应融合互补学习实时跟踪算法。当目标概率小于阈值时使用均值化自适应融合系数,否则使用指数自适应融合系数。在 OTB-2013 与 OTB-100 基准视频集上进行实验测试,结果表明本文提出的算法与 Staple 相比在精度上分别高出 6.52% 与 3.32%,在成功率上分别高出 4.89% 与 3.11%,不仅在精度与成功率上超过 Staple,而且在基准视频所有属性上均超过



图 5 前三算法在 10 个视频上的定性比较。(a) Bird1;(b) bolt;(c) ClifBar;(d) coke;(e) dog;(f) dragon baby;  
(g) shaking;(h) girl;(i) soccer;(j) trellis

Fig. 5 Qualitative comparison of top three algorithms on ten videos. (a) Bird1; (b) Bolt; (c) ClifBar; (d) coke; (e) dog;  
(f) dragon baby; (g) shaking; (h) girl; (i) soccer; (j) trellis

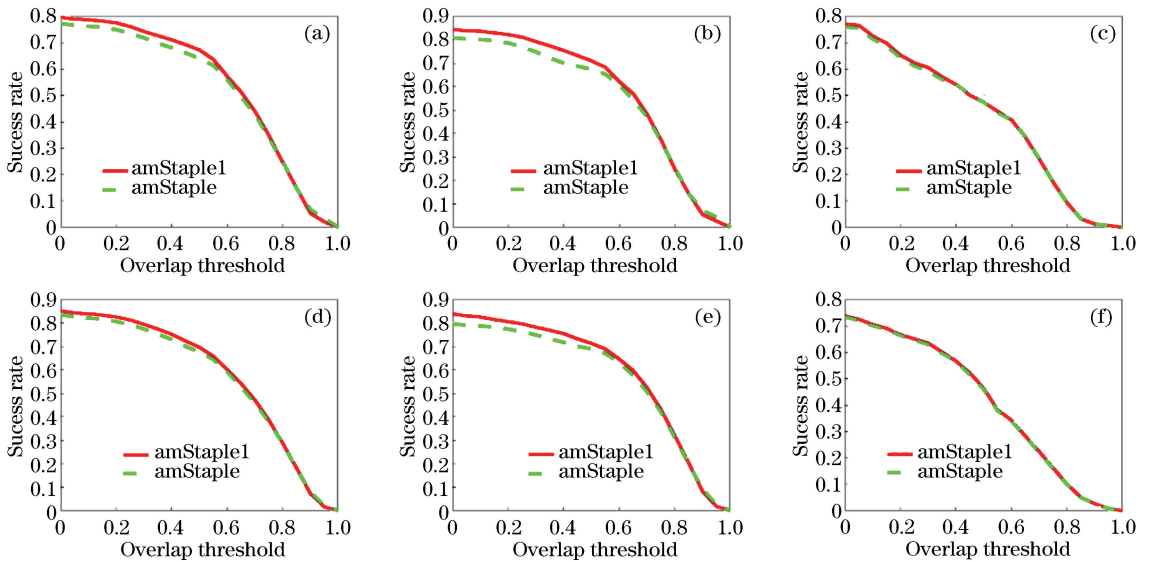


图 6 amStaple1 较 amStaple 在不同视频集上表现较佳的视频属性成功率曲线。OTB-2013 视频集上的(a)快速运动、  
(b)运动模糊和(c)低分辨率;OTB-100 视频集上的(d)快速运动、(e)运动模糊和(f)低分辨率

Fig. 6 Video attribute success rate curves where amStaple1 performs better than amStaple on different benchmarks.  
(a) Fast motion for OTB-2013 benchmark; (b) motion blur for OTB-2013 benchmark; (c) low resolution for OTB-  
2013 benchmark; (d) fast motion for OTB-100 benchmark; (e) motion blur for OTB-100 benchmark; (f) low  
resolution for OTB-100 benchmark



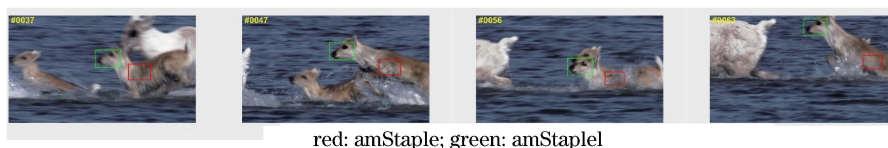


图7 amStaple1 与 amStaple 在 Deer 视频集跟踪效果图

Fig. 7 Tracking effects of amStaple1 and amStaple in Deer benchmark sequence. Red is amStaple and green is amStaple1

Staple,并且在 OTB-100 上精度高出 SRDCF 算法 2.79%,成功率与其持平。但本文所提出算法并没有有效地解决基准视频中的遮挡问题,后续工作会继续探索有效解决遮挡问题的算法,以进一步提升算法性能。

### 参 考 文 献

- [1] Zuo W M, Wu X H, Lin L, *et al.* Learning support correlation filters for visual tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(5): 1158-1172.
- [2] Zhu G, Porikli F, Li H D. Beyond local search: tracking objects everywhere with instance-specific proposals[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 943-951.
- [3] Zhu G B, Wang J Q, Wu Y, *et al.* Collaborative correlation tracking [C] // British Machine Vision Conference 2015, September 7-10, 2015, Swansea, Wales, UK. Durham, England, UK: BMVA Press, 2015: 184-184.
- [4] He X D, Zhou S Z. Fast scale adaptive kernel correlation filtering algorithm for target tracking [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121501.  
何雪东,周盛宗.快速尺度自适应核相关滤波目标跟踪算法.激光与光电子学进展,2018,55(12):121501.
- [5] Gao M F, Zhang X X. Scale adaptive kernel correlation filtering for target tracking [J]. Laser & Optoelectronics Progress, 2018, 55(4): 041501.  
高美凤,张晓玄.尺度自适应核相关滤波目标跟踪[J].激光与光电子学进展,2018,55(4):041501.
- [6] 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.
- [7] Wang X, Hou Z Q, Yu W S, *et al.* Target scale adaptive robust tracking based on fusion of multilayer convolutional features[J]. Acta Optica Sinica, 2017, 37(11): 1115005.  
王鑫,侯志强,余旺盛,等.基于多层卷积特征融合的目标尺度自适应稳健跟踪[J].光学学报,2017,37(11):1115005.
- [8] Vojir T, Neskova J, Matas J. Robust scale-adaptive mean-shift for tracking [J]. Pattern Recognition Letters, 2014, 49: 250-258.
- [9] 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.
- [10] 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. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2012, 7575: 702-715.
- [11] 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.
- [12] 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.
- [13] Danelljan M, Häger G, Khan F, *et al.* Accurate scale estimation for robust visual tracking [C] // British Machine Vision Conference 2014, September 1-5, 2014, Nottingham. Durham, England, UK: BMVA Press, 2014: 65.
- [14] Danelljan M, Häger G, Khan F S, *et al.* Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [15] Li Y, Zhu J K. A scale adaptive kernel correlation

- filter tracker with feature integration [M] // Agapito L, Bronstein M, Rother C, *et al.* Computer vision-ECCV 2014 workshops. Lecture notes in computer science. Cham: Springer, 2015, 8926: 254-265.
- [16] Danelljan M, Häger G, Khan F S, *et al.* Learning spatially regularized correlation filters for visual tracking [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 4310-4318.
- [17] 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.
- [18] 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.
- [19] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [20] Scholkopf B, Smola A J. Learning with kernels [M]. Cambridge: MIT Press, 2005: 645.