https://doi.org/10.1051/jnwpu/20193720323

基于多互补特征融合的视频跟踪算法

张君昌[1,2], 张登[1], 万锦锦[2]

(1.西北工业大学 电子信息学院, 陕西 西安 710072; 2.光电控制技术重点实验室, 河南 洛阳 471000)

摘 要:为充分利用跟踪过程中样本信息的多样性,提高跟踪器的泛化能力,在 Staple 算法的基础上融入了基于轮廓特征的物体性模型预测结果,并针对该算法中对不同预测响应结果简单线性进行加权带来的不确定性,提出一种新的自适应权重系数的响应图融合方法,从而有效地提升了跟踪算法的可靠性。理论分析与实验仿真表明,所提算法在精准度和鲁棒性较经典的 Staple 算法有着较大提高,并且保持着较高的实时性。

关 键 词:相关滤波;目标跟踪;轮廓特征;自适应权重
中图分类号:TP391 **文献标志码**:A **文章编号**:1000-2758(2019)02-0323-07

近年来,基于相关滤波器的跟踪算法受到广泛关注,该类方法属于判别式跟踪,其实质是将跟踪问题转化成一个分类问题,通过在线训练的分类器将前景和背景进行分离,实现目标跟踪的任务。

文献[1]提出了一种基于相关滤波器最小输出均方误差和的视频跟踪算法。文献[2]设计了 2 个一致并且相对独立的相关滤波器, 分别实现了目标位置跟踪和尺度变换。文献[3]将检测建议框架与相关滤波跟踪器框架相结合,并采用了特征集成、鲁棒更新与建议拒绝等优化方法,保证了各部分的有效整合。通过全局搜索,文献[4]找出了利用轮廓特征标记出的建议候选样本,得到较优的正负样本和更强的检测器,来完成分类器测试和更新,有效减少了分类器搜索空间,降低虚假目标干扰。为了克服相关滤波器跟踪算法对于形变和旋转鲁棒性较差的问题,文献[5]通过融合特征响应图的形式,将相关滤波器框架和颜色模型框架进行融合,提高了跟踪算法的鲁棒性。但是在互补特征的选择以及响应图的融合方式上还有待进一步研究。

本文将物体性检测理论引入到相关滤波跟踪和颜色模型跟踪的框架中,多互补模型的分类器框架能够有效地对目标和背景进行分离,克服跟踪器易受光照、背景干扰的问题,降低跟踪虚假目标的概率。并且本文通过衡量不同模型预测响应图的置信度来推理该模型对目标和背景的分辨能力,从而确定合适的加权权重,提升跟踪性能。

1 融合多互补特征的跟踪算法

Staple[5-6]算法通过相关滤波模型和颜色模型同时对目标和背景进行分类,提高了跟踪器的泛化能力。然而,跟踪模型还是易受到环境、光照等因素的干扰。基于上述问题,本文在 Staple 算法的基础上融入了物体性检测模型[7],充分利用样本信息多样化的特点。物体性检测模型基于物体边缘轮廓特征,对光照、背景变换有良好的适应性。新算法通过结合 3 个互补模型分别对目标位置进行分类预测,最后将得到的各个模型的预测响应图进行自适应加权融合来进一步提高跟踪器鲁棒性。算法整体框架如图 1 所示。

在第 $t-1$ 帧,得到最佳目标为 $\boldsymbol{\Pi}_{t-1}$,其中心位置矢量为 $\boldsymbol{p}_{t-1}$,大小为 $w_{t-1}\times h_{t-1}$。在 t 帧,依据文献[5]提取搜索图像块 z_d,中心位置矢量为 $\boldsymbol{p}_{t-1}$,大小为 $(w_{t-1}+l)\times(h_{t-1}+l)$,$l=(w_{t-1}+h_{t-1})/2$,并通过插值法将 z_d 等纵横比缩放为尺寸 $\psi w_0\times\psi h_0$ 的标准图像块 $\bar{z}_d$,使得 $\psi w_0\times\psi h_0$ 为一固定值 C。然后通过滑动采样的方式,得到样本边界框集合 $\boldsymbol{\Omega}$。由于本

收稿日期:2018-04-18 **基金项目**:光电控制技术重点实验室航空科学基金(2016515303)资助
作者简介:张君昌(1969—),西北工业大学教授,主要从事图像信号处理研究。

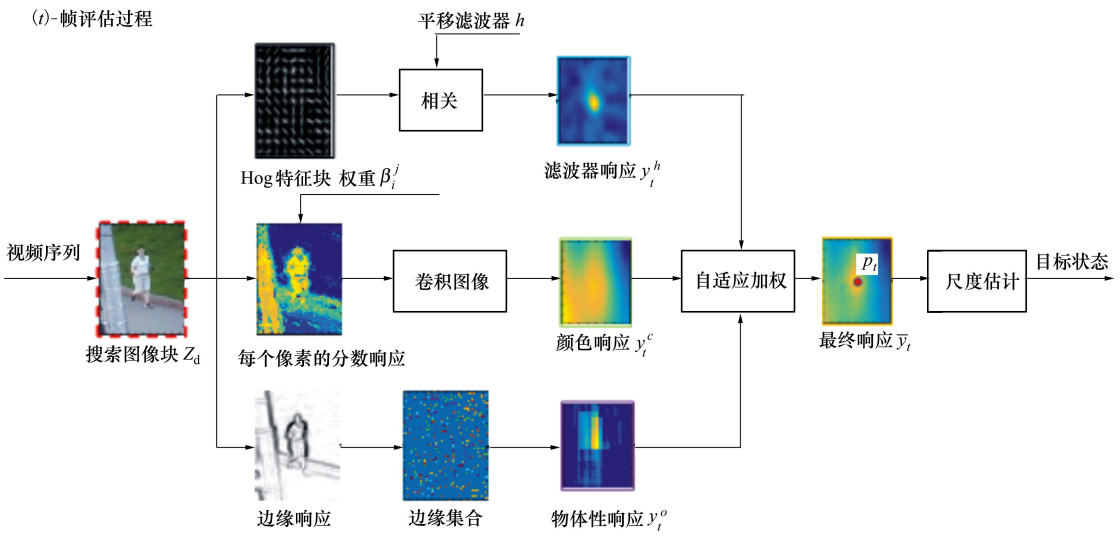


图 1 多互补特征融合跟踪器框架

文算法将目标的位移估计和尺度估计相分离,所以滑动采样中的边界框 $\boldsymbol{\Pi}$,$\boldsymbol{\Pi} \in \boldsymbol{\Omega}$,大小固定为 $b_w \times b_h = \dfrac{C}{(w_{t-1}+l)(h_{t-1}+l)}(w_{t-1} \times h_{t-1})$,其中 $l = (w_{t-1}+h_{t-1})/2$,即记尺寸为 $b_w \times b_h$。然后新算法通过滤波器模型、物体性检测模型以及颜色模型分别对 $\boldsymbol{\Omega}$ 集合中的所有边界框 $\boldsymbol{\Pi}$ 进行打分。并通过自适应权重的方法计算所有边界框的综合得分

$$\bar{S}(\boldsymbol{\Pi}) = (w_t^c)S_c(\boldsymbol{\Pi}) + (w_t^o)S_o(\boldsymbol{\Pi}) + (w_t^f)S_f(\boldsymbol{\Pi}) \tag{1}$$

式中,w_t^c,w_t^o 和 w_t^f 分别为对应模型的加权权重。最后选择得分最高的边界框的中心位置为最佳目标中心位置 $\boldsymbol{p}_t$,此过程同样等同于对模型各个响应图进行加权,后取加权后响应中响应幅值最大的响应点作为目标中心位置。

1.1 平移滤波器模型与其响应图计算

在训练过程中,算法首先提取标准图像块 $\bar{z}_d$ 及相应的 $d(d=28)$ 维方向梯度直方图特征(HOG)f^l,$l \in (1,\cdots,d)$ 描述,然后通过最小化目标函数,训练一组 d 维滤波器 h。目标函数公式为

$$\dot{o} = \left\| g - \sum_{l=1}^{d} h^l * f^l \right\|^2 + \lambda \sum_{l=1}^{d} \| h^l \|^2 \tag{2}$$

式中,$*$ 表示循环相关,h^l 为第 l 维特征对应的滤波器系数,g 为回归响应函数,本文选择高斯函数。第二项 λ 表示正则化项。通过在频域下利用帕塞瓦尔定理进行快速求解,得到滤波器的频域形式为:

$$H^l = A_t^l / B_t$$

$$A_t^l = \bar{G}F^l \quad B_t = \sum_{k=1}^{d} \bar{F}^k F^k + \lambda \quad l = 1,\cdots,d \tag{3}$$

式中,$\bar{G}$ 为高斯响应的 DFT 共轭项,A_t^l 和 B_t 分别代表 l 维滤波器系数 H^l 的分子和分母项,F^k 和 $\bar{F}^k$ 分别为 k 维特征图像块 f^k 对应的频域形式以及对应共轭项。滤波器系数的更新策略如下:

$$A_t^l = (1-\eta)A_{t-1}^l + \eta_f \bar{G}F_t^l$$

$$B_t = (1-\eta)B_{t-1} + \eta_f \sum_{k=1}^{d} \bar{F}_t^k F_t^k \tag{4}$$

式中,$l = 1,2,\cdots d$,η 为学习因子。

当获取到一帧新的图像序列后,在上一帧目标中心位置矢量 $\boldsymbol{p}_{t-1}$ 提取搜索图像块 z_d,并按上文方法归一化为分辨率为固定值 C 的标准图像块 $\bar{z}_d$,然后提取 $\bar{z}_d$ 的 d 维方向梯度直方图特征(HOG)z_t^l,$l = 1,2,\cdots d$,并得到各维度特征块的频域形式 Z_t^l,最后与训练好的相关滤波器进行卷积得到滤波器的权重分数响应 y_t^h

$$y_t^h = F^{-1}\left(\frac{\sum_{l=1}^{d} \overline{A_{t-1}^l} Z_t^l}{B_{t-1} + \lambda} \right) \tag{5}$$

式中,A_{t-1}^l 和 B_{t-1} 是上一帧训练得到的位置滤波器的分子项和分母项。

得到的滤波器响应图 y_t^h 可等效地表示在滤波器模型下,预测得到的不同位置固定尺寸 $(w_{t-1} \times h_{t-1})$ 边界框 $\boldsymbol{\Pi}$ 的权重分数 $S_f(\boldsymbol{\Pi})$。

1.2 物体性检测模型预测

基于轮廓信息的物体性检测模型可以通过图像

的轮廓信息对样本边界框集合 $\boldsymbol{\Omega}$ 中所有边界框打分,分数大小表示边界框内包含目标的概率。算法通过结构化边缘检测计算出标准检测图像块 $\bar{z}_d$ 中每一个像素点 p 对应的边缘幅值 m_p 和方向 θ_p,并将 $m_p > 0.1$ 的像素点记作边界点。将形成相干边界点组成的一条边或曲线定义为轮廓并将边界群定义为属于相邻轮廓的边界像素的集合。因此边界群内的相邻点不仅应该距离靠近,梯度方向也应该接近,而那些未连接或通过高曲率的轮廓相连的边界像素就不太会属于同一边界群。基于此依据,文献[7]定义了 2 个边界群之间的相似度。设任意 2 个边界群 s_i 和 s_j 对应的平均中心位置分别为 x_i 和 x_j,平均梯度方向为 θ_i 和 θ_j,θ_{ij} 为 θ_i 和 θ_j 之间的夹角,那么 s_i 和 s_j 的相似度 $a(s_i,s_j)$ 表示为:

$$a(s_i,s_j) = |\cos(\theta_i - \theta_{ij})\cos(\theta_j - \theta_{ij})|^{\gamma} \quad (6)$$

式中,γ 用来调节相似度对方向的敏感程度。当 2 个边界群之间的距离大于 2 个像素时,则两者的相似度就为 0。

算法首先计算标准检测图像块 $\bar{z}_d$ 中所有的边界群,对于位于 $\bar{z}_d$ 中任意位置的一个边界框样本 $\boldsymbol{\Pi}$,文献[7]中计算出每一个边界群 s_i 完全位于 $\boldsymbol{\Pi}$ 的权重分数分 ζ_i,$\zeta_i \in [0,1]$,边界框 b 的物体性权重分数 $S_o(\boldsymbol{\Pi})$ 可表示为:

$$S_o(\boldsymbol{\Pi}) = \frac{\sum_i \zeta_i m_i}{2(b_w + b_h)^k} - \frac{\sum_{r \in b_{in}} m_r}{2(b_w + b_h)^k} \quad (7)$$

式中,b_w 和 b_h 分别为边界框 $\boldsymbol{\Pi}$ 的宽和高。b_{in} 代表边界框 $\boldsymbol{\Pi}$ 的中心区域,尺寸为 $(b_w/2) \times (b_h/2)$;r 表示边界框内的任意一个边界像素,m_r 为 r 的响应幅值,m_i 为边界群 s_i 中所有边界像素幅值之和,ζ_i 为边界群 s_i 完全位于 b 的权重分数,边界像素响应幅值和 ζ_i 的计算方法可见文献[7],ζ_i 越高表示 s_i 完全位于 $\boldsymbol{\Pi}$ 内的可能性就越大;k 为对目标长、宽大小的惩罚系数。

因此,可以通过物体性模型有效地计算出边界框样本集合 $\boldsymbol{\Omega}$ 中任意位置边界框的物体性分数 $S_o(\boldsymbol{\Pi})$,由于响应图中每个点对应 $\boldsymbol{\Omega}$ 中一个样本,其响应值对应 $\boldsymbol{\Omega}$ 的得分,因此可以得到整个物体性响应图 y_t^o。

1.3　颜色模型预测

基于颜色模型的目标跟踪方法具有旋转、平移和尺度不变性的优点。文献[5]中颜色频率的计算和像素权重分数的计算都是在映射空间中完成的。训练阶段算法将训练图像块分割为目标区域和背景区域两部分,并将所有 RGB 像素映射到 bin 空间下的索引像素,并分别统计任意索引像素 j 在目标区域的频率 $\rho_t^j(O)$ 和属于背景的频率值 $\rho_t^j(B)$,且每一帧分别对频率信息进行更新。

$$\rho_t(O) = (1 - \eta_c)\rho_{t-1}(O) + \eta_c\rho_t'(O)$$
$$\rho_t(B) = (1 - \eta_c)\rho_{t-1}(B) + \eta_c\rho_t'(B) \quad (8)$$

式中,$\rho_t(\cdot)$ 是由 $\rho_t^j(\cdot)$,$j = 1,\cdots M$,组成的 M 维向量,其中 $M = 2\,048$。

在预测阶段,搜索标准图像块 $\bar{z}_d$ 中的每一个像素点 v 属于目标的权重 $\beta_t^{\varphi(v)}$ 和像素索引 $j = \varphi(v)$ 出现在目标区域的频率值 $\rho_t^{\varphi(v)}(O)$ 以及像素索引 j 出现在背景区域的频率值 $\rho_t^{\varphi(v)}(B)$ 有关,$\beta_t^{\varphi(v)}$ 的计算表示为

$$\beta_t^{\varphi(v)} = \frac{\rho_t^{\varphi(v)}(O)}{(\rho_t^{\varphi(v)}(O) + \rho_t^{\varphi(v)}(B) + \lambda)} \quad (9)$$

式中,任意的 $\rho_t^j(O)$ 和 $\rho_t^j(B)$ 已经在训练阶段计算得到,所以 $\rho_t^{\varphi(v)}(O)$ 和 $\rho_t^{\varphi(v)}(B)$ 可通过查找表直接得到。λ 用来保证分母非零。当计算出搜索图像块 $\bar{z}_d$ 中的每一个像素点 v 属于目标的权重 $\beta_t^{\varphi(v)}$ 后,就得到了与 $\bar{z}_d$ 尺寸相同的每个像素的分数响应图,再通过计算滑动窗口产生的每一个边界框 $\boldsymbol{\Pi}$ 内所有像素权重平均值,并用图像卷积进行加速计算。

$$S_c(\boldsymbol{\Pi}) = \frac{1}{|\boldsymbol{\Pi}|}\sum_{v \in \boldsymbol{\Pi}} \beta_t^{\varphi(v)} \quad (10)$$

式中,$|\boldsymbol{\Pi}|$ 表示边界框 $\boldsymbol{\Pi}$ 的像素总数。最终可以得到边界框样本集合 $\boldsymbol{\Omega}$ 中任意边界框的颜色权重分数 $S_c(\boldsymbol{\Pi})$,进而同物体性检测模型相同,可以得到颜色响应 y_t^c。

1.4　基于响应图置信度的自适应融合算法

文献[5]对于 2 个响应图的融合采用了固定权重的融合方式,然而,目标跟踪过程中各个时刻跟踪器中各个模型受干扰程度往往不尽相同,固定权重不能准确有效地利用各个模型的预测信息,导致跟踪器对目标跟踪的鲁棒性不高。对于融合多通道的跟踪器来说,若不能对受不同程度干扰的模型动态加入合适的权重系数,跟踪的稳定性难以得到保证。为此,本文通过有效地计算不同模型预测响应图的置信度,进而自适应地对模型权重进行调整,进一步提高跟踪算法的鲁棒性。

文献[6]中用不同模板对正负样本的分别能力来表示跟踪的置信度,其计算过程需要一定数量有

对应标签的目标样本。然而,目标样本标签的得到一般是通过跟踪器历时帧的跟踪结果来确定的,因此跟踪结果发生飘逸将会导致置信度的计算错误,跟踪器陷入循环的错误计算中。而文献[6]通过将上述问题转换为半监督的部分标签问题,首先计算一个可能的部分带有标签的图像块样本集合:

$$\{(x^1,z^1),\cdots(x^i,z^i)\},i=1,\cdots n。 \tag{11}$$

候选目标边界框集合的获得,需要分别计算滤波器模型响应图、颜色模型响应图以及物体性检测模型响应中响应权重较大的目标边界框集合 $\boldsymbol{v}_1$,$\boldsymbol{v}_2$ 和 $\boldsymbol{v}_3$,并取其并集 $\boldsymbol{v}=\boldsymbol{v}_1\cup\boldsymbol{v}_2\cup\boldsymbol{v}_3$,得到总的候选边界框集合 $\boldsymbol{v}$。同时,通过迭代合并相邻位置点使得到的任意 2 个边界框之间的重合度小于 0.5,保证集合中只有一个边界框为真目标边界框。最终得到的这些边界框集合是基于各个模型估计的属于目标可能性高的边界框,对于这些边界框是否属于目标的进一步的分辨能力往往决定了其跟踪目标的性能。设最终的边界框集合为 $\hat{\boldsymbol{v}}=\{x^1,x^2,\cdots x^n\}$,而对于集合 $\hat{\boldsymbol{v}}$ 中的每一个推荐目标框 x^i 都对应一个标记 $S^i=(\varphi^i,l^i)$,其中,φ^i 表示候选边界框属于目标或者背景的二进制标签 $\varphi^i\in\{-,+\}$,l^i 代表边界框的中心位置的坐标。所以可以得到边界框集合 $\hat{\boldsymbol{v}}=\{x^1,x^2,\cdots x^n\}$ 标签的所有边界框可能对应一个标签集合 $\boldsymbol{z}=\{S_1,\cdots S_n\}$,而 $\boldsymbol{S}_j=((\varphi_j^1,l_j^1),\cdots,(\varphi_j^n,l_j^n))$ 表示其中的一种标签组合,并且仅有一个边界框的标签为真,其他目标框的标签都为假,即当 $i=j$,其他 $i\neq j$ 时 $\varphi_j^j=0$,标签集 $\boldsymbol{z}=\{S_1,\cdots S_n\}$,的取值空间为 $\boldsymbol{y}=(\{-,+\}\times Z^2)^n$。得到具有标签的边界框集合后,通过公式(12)计算出各个模型响应图的置信度。如公式(12)所示,置信度 l 由 2 项组成,第一项 L 为由响应图得到的最大似然概率,第二项 H 为由响应图得到损失函数。同时 ζ 为两者的调节因子,最终置信度计算公式表示为

$$y_l(\theta)=-L(\theta;\hat{\boldsymbol{v}},z)+\zeta H(S\mid\hat{\boldsymbol{v}},z;\theta) \tag{12}$$

最大似然概率 L 和损失熵 H 的计算公式分别为:

$$L(\theta;\hat{v},z)=\max_{y\in z}\log P(y\mid\hat{v};\theta) \tag{13}$$

$$H(S\mid x,z;\theta)=\sum_{y\in Y}P(S\mid x,z;\theta)\log P(S\mid x,z;\theta) \tag{14}$$

这里假设目标所在位置概率模型 $P(l^i\mid\phi^i)$ 为高斯分布函数。并且假设 $P(l^i\mid\phi^i,x^i;\theta)=P(l^i\mid\phi^i;\theta)$ 成立,则条件概率可以表示为

$$P(S\mid\hat{\boldsymbol{v}};\theta)=\prod_i P(l^i\mid\phi^i)P(\phi^i\mid x^i;\theta) \tag{15}$$

同时令

$$P(S\mid\hat{\boldsymbol{v}},z;\theta)=\frac{\delta_z(S)P(S\mid\hat{v};\theta)}{\sum_{S'\in Y}\delta_z(S')P(S'\mid\hat{\boldsymbol{v}};\theta)} \tag{16}$$

式中,$S'\in z$ 时 $\delta_z(S')$ 取 1,否则为 0。详细过程可以参考文献[6]。得到复合特征响应图和滤波器响应图的置信度后,就可以计算各个响应图的加权权重。

在预测结果的加权权重设计中,本文通过考察不同模型对不同位置边界框是否属于目标的分辨能力强弱,并结合模型对历史帧边界框分辨能力来决定该当前帧该模型预测结果的加权比重。在跟踪过程中,当某一模型预测结果的置信度高时,则保持该模型预测响应图加权权重不变,当某一模型的预测结果置信度低时,降低其预测响应图加权权重。模型预测结果置信度高低的判断方法如下:设 y_t^* 为 t 时刻不同模型的预测响应图,E_t^* 为该模型的响应图,$\bar{E}_t^*$ 为该模型响应图前 Δ 帧的平均置信度,$\bar{E}_t^*$ 的计算公式分别为:

$$\bar{E}_t^*=\frac{1}{\Delta}\sum_{l\in[t-1-\Delta,t-1]}E_l^* \tag{17}$$

基于上述讨论,本文提出通过从以下方式对 HOG 模型在 t 时刻的加权权重 w_t^* 进行自适应调整

$$w_t^*=\begin{cases}a^* & E_t^*\geqslant\vartheta\bar{E}_t^*\\ a^*\beta & E_t^*<\vartheta\bar{E}_t^*\end{cases} \tag{18}$$

式中,a^* 表示不同模型预测响应图的初始加权权重,β 为当模型响应图置信度低时响应图加权权重的衰减系数,ϑ 为预先设定的模型在 t 时刻置信度应高出前 Δ 帧平均置信度值的对比系数。最终的响应图 $\bar{y}_t$ 的融合公式如下:

$$\bar{y}_t=w_t^f y_t^h+w_t^c y_t^c+w_t^o y_t^o \tag{19}$$

式中,w_t^f,w_t^c 和 w_t^o 分别为滤波器响应图 y_t^h、颜色响应图 y_t^c 和物体性响应图 y_t^o 的自适应加权权重。

1.5 尺度的计算和优化

当平移滤波器确定出目标中心坐标后,引用文献[5]的方法,用一维尺度滤波器计算目标当前尺度。尺度金字塔建立如下

$$\zeta^n w_{t-1}\times\zeta^n h_{t-1},n\in\{-(\rho-1)/2,\cdots,(\rho-1)/2\} \tag{20}$$

式中,w_{t-1},h_{t-1} 分别为目标在上一帧的宽和高。ζ 为尺度放缩因子,本文选择 1.02,ρ 为尺度金字塔的层数,本文选择 33。具体计算可参考文献[5]。

1.6　算法实现

本文算法可分为 4 个主要组成部分：初始化、位置估计、尺度估计和模型训练。具体步骤如下：

1）初始化

初始设置算法的初始参数，以及第一帧序列中目标的位置和尺度状态。

2）位置估计

a）在 t 帧序列中以上一帧目标中心位置矢量为 $\boldsymbol{p}_{t-1}$ 和尺度 $(w_{t-1} \times h_{t-1})$ 提取搜索图像块 z_d，大小为 $(w_{t-1}+l) \times (h_{t-1}+l)$，并插值得到尺寸归一化后的图像块 $\bar{z}_d$。

b）提取 $\bar{z}_d$ 的 28 维特征，并通过 A_t^l、B_t 和公式（5）计算出滤波器响应图 y_t^h。

c）分别计算 $\bar{z}_d$ 的边界响应和图像块中所有的边界群，并通过公式（7）计算出物体性响应图。

d）根据公式（9）计算出 $\bar{z}_d$ 的每个像素颜色响应图，并通过积分图（公式（10））计算出颜色响应分数。

e）根据公式（11）计算每个模型响应图置信度，并通过公式（18）计算出每个响应图对应的自适应加权权重。最后通过（19）计算出最终的响应图 $\bar{y}_t$，选择响应值最大的位置作为新的目标中心位置。

3）尺度估计

根据尺度金子塔对目标尺度进行估计。

4）模型训练

根据新的目标中心位置矢量 $\boldsymbol{p}_t$ 和尺度 $(w_t \times h_t)$ 提取训练图像块，通过公式（4）更新滤波器系数，利用公式（8）和（9）更新颜色模型的权重分数 β_t^j。

2　实验结果

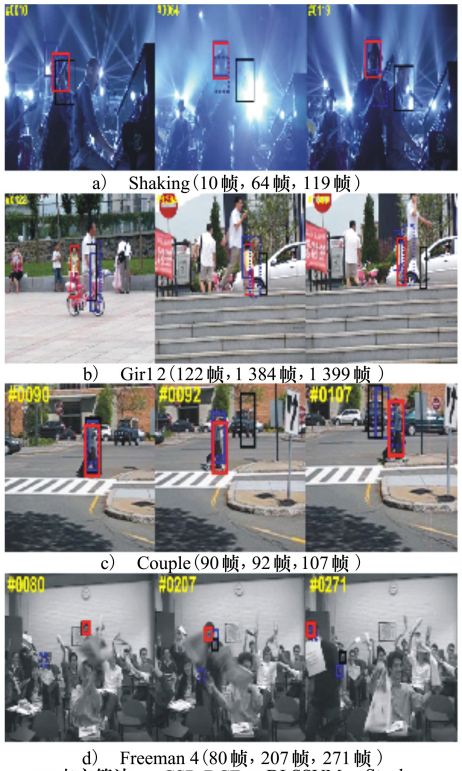
本文在计算机处理器为 i7，主频为 3.5GHz CPU，内存为 8G RAM 的硬件配置下通过 MATLAB2016b 软件对算法在 OTB2015 提供的 100 个视频序列上进行了实验评估。测试视频序列包含的干扰特征包括特征有：遮挡，光照变化，尺度变化，运动模糊，快速运动，平面外旋转，变形等。实验中的参数设置如表 1 所示。

表 1　本文算法参数设置

参数名	取值
滤波器学习速率 η_f	0.01
尺度滤波器学习速率	0.035
初始滤波器模型加权权重 a^l	0.55
初始滤波器模型加权权重 a^o	0.2
初始颜色模型加权权重 a^c	0.25
固定区域	150
颜色空间	RGB
bin 颜色空间	32×32×32
颜色模型学习速率 η_c	0.04
β	0.6
ϑ	0.9
惩罚系数 k	1.5
调节因子 γ	2
ζ	0.7
HOG 特征胞元大小	4
历史帧数 Δ	5

2.1　定性分析

如图 2 所示，本节选取了 4 个同时包含多种挑战因素的视频序列来对算法进行定性评估，并列出了包括双线性支持向量机（DLSSVM）、具有通道和空间可靠性的判别相关滤波器（CSR-DCF）以及融合相关滤波器模板和颜色特征模板跟踪算法（Staple）以及本文算法在内的 4 种跟踪算法的跟踪结果。



a）　Shaking（10 帧，64 帧，119 帧）

b）　Girl2（122 帧，1 384 帧，1 399 帧）

c）　Couple（90 帧，92 帧，107 帧）

d）　Freeman 4（80 帧，207 帧，271 帧）

本文算法　CSR-DCF　DLSSVM　Staple

图 2　各个算法在各测试序列上部分时刻的跟踪结果

Shaking 序列中存在剧烈的光照变化，并存在一

定目标形变。如图 2a)所示,由于 CSR-DCF 依赖颜色信息,当第 10 帧时光照发生强烈变化时,CSR-DCF 产生了严重的跟踪飘逸。在第 64 帧 Staple 算法已经跟踪失败, DLSSVM 也已经产生了较大的跟踪误差。而本文算法能够较好地对目标进行跟踪。

Girl 2 序列中存在着完全遮挡的干扰,本文算法当目标出现后能及时捕获到目标,同样在第 1 384 帧后目标又出现了完全遮挡的问题,在第 1 399 帧,本文算法依旧能够较准确地跟踪到目标。

Couple 序列中存在着相似背景和运动模糊的干扰,在第 92 帧和 107 帧时,Staple 算法和 CSR-DCF 相继出现跟踪失败。而 DLSSVM 算法和本文算法能有效地提取正确的目标表观信息,较好地适应目标表观变化,完成目标跟踪的任务。

在 Freeman 4 序列中,目标分辨率较低同时存在着一定的目标遮挡。如第 207 帧 DLSSVM、CSR-DCF 以及 Staple 算法都出现了跟踪飘逸,在第 271 帧 Staple 和 CSR-DCF 算法甚至已经完全跟踪失败,而本文算法能够较好地对目标跟踪进行跟踪。

2.2 定量分析

本节采用中心位置误差(CLE)和重叠率(OR)进行性能评估。一般来说,较小的平均误差和较大的重叠率意味着更准确的跟踪结果。表 2 和表 3 分别表示不同算法在各个测试序列中的中心位置平均误差和边界框重叠率平均值的比较结果。表 4 给出了各个算法在 OTB2015 中 100 序列中的平均跟踪帧率。表 2、表 3 以及表 4 中最优及次优的结果分别用 1 和 2 进行标注。

表 2 不同算法在各个视频序列的平均中心位置误差

序列	算法			
	DLSSVM	CSR-DCF	Staple	本文算法
Couple	7.916	7.741[2]	34.588	6.184[1]
Freeman 4	50.385	16.836[2]	21.212	5.099[1]
Shaking	11.201[2]	11.505	125.060	6.972[1]
Girl 2	72.887	44.720[2]	111.738	12.122[1]

表 3 不同算法在各个视频序列的平均边界框重叠率

序列	算法			
	DLSSVM	CSR-DCF	Staple	本文算法
Couple	0.589	0.643[2]	0.527	0.721[1]
Freeman4	0.143	0.389	0.405[2]	0.596[1]
Shaking	0.668	0.674[2]	0.057	0.769[1]
Girl2	0.375[2]	0.276	0.070	0.680[1]

表 4 不同算法的平均帧率 帧/秒

DLSSVM	CSR-DCF	Staple	本文算法
15	20	80[1]	60[2]

2.3 基于可视基准的评估

本文采用一次性评估实验 OPE 通过 OTB2015 提供的 100 个视频序列对 Staple、DLSSVM、CSR-DCF、基于循环特征映射的跟踪算法(LMCF)、空间正则化滤波跟踪器(SRDCF)、滤波网络跟踪器(DCFNet)、全卷积网络跟踪器(SiamFC)、卷积神经网络支持向量机跟踪器(CNN-SVM)和本文算法在内的 9 种跟踪器进行整体性能测试和对比。如图 3 所示,本文算法由于在 Staple 算法的基础上,融合了物体性检测模型,充分利用了样本信息的多样化的特性,并采用自适应权重的融合方式来调整各模型加权比重,有效提升了跟踪算法的鲁棒性。新算法在和 OTB2015 序列集下得到的精度和成功率相比较于其他跟踪算法均取得了最好的性能。其次,引入的置信度计算本身复杂度较低,所以保证新算法依然能够保持较高的跟踪实时性。通过表 4 可得,新算法和 Staple 算法相比,依然有着较高的跟踪实时性。

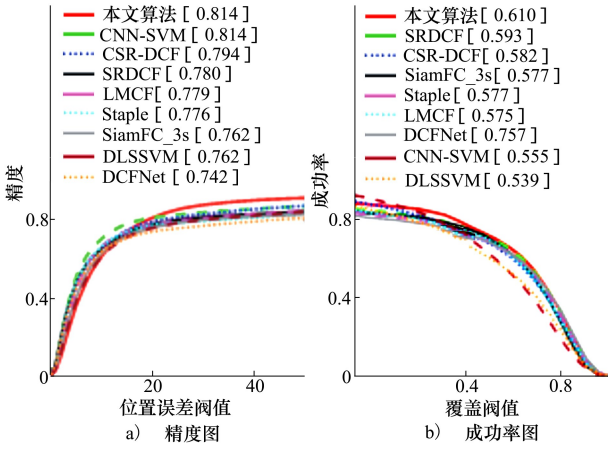


图 3 跟踪器在 OTB2015 上的测试结果

3 结 论

为了提高跟踪算法适应复杂场景的能力,本文在 Staple 算法的基础上融入了基于轮廓特征的物体性检测模型,与原有基于 HOG 特征的滤波器模型和颜色模型结合形成一种融合多互补模型的跟踪算法。并针对该算法中对不同模型预测响应结果进行

简单线性加权带来的跟踪性能不稳定的问题,提出一种以最大似然概率和损失熵作为响应图置信度的自适应预测响应图加权权重计算方法。实验结果表明,新算法在多种复杂场景中能有效地解决目标跟踪性能较低的问题,并且实时性较强。

参考文献:

[1] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual Object Tracking Using Adaptive Correlation Filters[C]//Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition. Washington, 2010:2544-2550

[2] DANELLJAN M, HäGER G, KHAN F S, et al. Accurate Scale Estimation for Robust Visual Tracking [C]//Proceedings of the British Machine Vision Conference, Durham, 2014: 65.1-65.11

[3] HUANG D. Enable Scale and Aspect Ratio Adaptability in Visual Tracking with Detection Proposals[C]//Proceedings of the British Machine Vision Conference, Durham, 2015:185

[4] ZHU G, PORIKLI F, LI H. Beyond Local Search: Tracking Objects Everywhere with Instance-Specific Proposals[C]//Proceedings of the 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, 2016: 943-951

[5] BERTINETTO L, VALMADRE J, Golodetz S, et al. Staple: Complementary Learners for Real-Time Tracking [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Washington, 2016: 1401-1409

[6] ZHANG J, MA S, SCLAROFF S. MEEM: Robust Tracking via Multiple Experts Using Entropy Minimization[C]//Proceedings of the 2014 European Conference on Computer Vision, Berlin, 2014: 188-203

[7] ZITNICK C L, DOLLÁR P. Edge Boxes: Locating Object Proposals from Edges[C]//Lecture Notesin Computer Science: 8693. Heidelberg: SpringerVerlag, 2014: 391-405

A New Video Tracking Algorithm Based on Multi-Complementary Features Fusion

ZHANG Junchang[1,2], ZHANG Deng[1], WAN Jinjin[2]

(1.School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China;
2.Science and Technology on Electro-Optic Control Laboratory, Luoyang 471000, China)

Abstract: In order to make full use of the diversity of sample information in the tracking process and improve the generalization ability of the tracker, this paper integrates the object model prediction results on the basis of the Staple algorithm, and applies weighted bands to the simple linearity of different predictive response results in the algorithm. To the uncertainties, a new adaptive response factor graph fusion method with weight coefficients is proposed, which effectively improves the reliability of the video target tracking algorithm. Theoretical analysis and experimental simulation show that the proposed algorithm is more accurate and robust than the classical Staple algorithm, and it maintains high real-time performance.

Keywords: correlation filters; object tracking; contour features; adaptive weights; feature fusion

© 2019 Journal of Northwestern Polytechnical University.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.