

结合纹理和形状特征的在线混合随机朴素贝叶斯视觉跟踪器

郭鹏宇^{1,2} 苏昂^{1,2} 张红良^{1,2} 张小虎^{1,2*} 于起峰^{1,2}

¹国防科学技术大学航天科学与工程学院, 湖南长沙 410073

²国防科学技术大学图像测量与视觉导航湖南省重点实验室, 湖南长沙 410073

摘要 基于机器学习的思想并充分利用外观信息,提出一种在线选择纹理和形状特征的混合随机朴素贝叶斯视觉跟踪器。构造归一化空间金字塔,通过强度二值特征和金字塔梯度方向直方图二值特征,描述全局与局部区域的纹理和形状;并根据特征描述的二值性和多模性,设计并实现了在线混合朴素贝叶斯分类器。分类器预测类别后验概率生成信任图,跟踪器通过分析信任图实现目标跟踪,并利用极大似然估计和交叉验证实现外观学习和特征选择。选用基准测试集比较同类方法,从性能和复杂度两方面评估了跟踪器。实验结果表明跟踪器对光照变化,部分遮挡等情况具有一定的适应能力,且执行速度较快,存储空间较小。

关键词 机器视觉;机器学习;视觉跟踪器;纹理和形状特征;混合随机朴素贝叶斯

中图分类号 TP187

文献标识码 A

doi: 10.3788201535.0315002

Online Mixture of Random Naïve Bayes Tracker Combined Texture with Shape Feature

Guo Pengyu^{1,2} Su Ang^{1,2} Zhang Hongliang^{1,2} Zhang Xiaohu^{1,2} Yu Qifeng^{1,2}

¹College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, Hunan 410073, China

²Hunan Key Laboratory of Videometrics and Vision Navigation, National University of Defense Technology, Changsha, Hunan 410073, China

Abstract Based on the idea of machine learning and the sufficient appearance, a mixture random Naïve Bayes visual tracker with online texture and shape feature selection is proposed. The texture and shape of global and local region is described with binary feature of intensity and pyramid histogram of oriented gradients using normalized spatial pyramid. An online mixture of Naïve Bayes classifier is designed and realized according to binary and multi-model description. The classifier predicts the class posterior probability to generate the confidence map, then the tracker analyzes the confidence map to track the object, learns the appearance with maximum likelihood estimation, and selects the feature with cross validation. Compared with homogeneous methods, the tracker is evaluated with performance and complexity based on benchmarks. The experimental results show that the tracker has certain adaption to illumination change and partial occlusion, and fast execution speed as well as little memory space.

Key words machine vision; machine learning; visual tracker; texture and shape feature; mixture random Naïve Bayes

OCIS codes 150.4065; 100.4999; 330.4150

收稿日期: 2014-09-10; 收到修改稿日期: 2014-11-05

基金项目: 国家自然科学基金(11272347)、国家973计划(2013CB733100)

作者简介: 郭鹏宇(1985—),男,博士研究生,主要从事视觉跟踪与运动分析等方面的研究。E-mail: pengyu.guo@nudt.edu.cn

导师简介: 于起峰(1958—),男,教授,中国科学院院士,主要从事空天图像测量与视觉导航等方面的研究。

E-mail: yuqifeng@vip.163.com

*通信联系人。E-mail: zhx1302@hotmail.com

本文电子版彩色效果请详见中国光学期刊网 www.opticsjournal.net

1 引 言

近年来,随着视觉传感器技术的不断发展,视频监控已经应用到了公共场所监视和无人机侦查等民用和军用领域,通过变化检测或人工标注出感兴趣目标,对目标进行跟踪可以给出目标在图像上的运动信息,从而对目标进行增强显示或运动分析。由于事先不具备目标的先验信息,只提供首帧的目标信息,当目标在图像上的外观发生变化时,为了稳定跟踪,可以通过机器学习的方法在线学习目标的外观。

基于机器学习的视觉跟踪方法,可以为空间和时间范围上的分类问题建模,从而通过分类/检测实现目标跟踪。早期的研究集中在离线训练分类器,在线实现目标跟踪,如支持矢量跟踪器(SVT)^[1]。这类方法属于有监督学习,需要大量标记样本一次性训练分类器,由于较大的时空开销且在很多应用中样本是依时间顺序得到的,因此近来的研究主要集中在半监督学习^[2],将新样本加入到已有分类器,在线跟踪并学习的方法,如在线集体学习跟踪器(oab)^[3],在线随机森林跟踪器(orf)^[4],在线随机朴素贝叶斯跟踪器(ornb)^[5],核结构输出支持矢量跟踪器(struck)^[6]等。但现有的基于检测的跟踪方法仅关注在线分类器的设计,并没有充分利用视频提供的外观信息,文献[7]指出分类器针对不同目标时应该选择不同的特征,主要包括形状和纹理两类特征,例如飞机目标选用形状特征,老虎目标选择纹理特征。因此为了获得更好的跟踪性能,本文设计并实现了在线选择纹理和形状特征的混合随机朴素贝叶斯跟踪器,该跟踪器通过纹理和形状两类二值特征建模外观,再根据类别后验概率生成的信任图定位目标,并由极大似然估计在线更新概率分布,同时依据当前场景,利用交叉验证选择具有判别力的特征。

2 相关工作

基于检测的跟踪方法是当前视觉跟踪领域的主要研究内容,其中一部分工作通过分类器实现检测。这类方法需要解决特征表达、分类器训练和目标定位三个环节,其难点在于存在光照变化、目标位姿变化、轻微视点变化、噪声干扰、复杂背景干扰和遮挡等因素,从而影响目标在图像上呈现的外观发生变化时的稳定和准确跟踪。

视觉跟踪可以依赖丰富的外观信息,外观可由纹理和形状特征描述。局部二值模式(LBP)^[8],二进制健壮独立元素特征(BRIEF)^[9]等纹理二值特征近年来广泛流行;而描述形状特征多采用形状上下文^[10]、边界^[11]、梯度方向直方图(HOG)^[12]和金字塔梯度方向直方图(PHOG)^[13]等。旨在通过联合纹理和形状特征,充分利用外观信息,在BRIEF和PHOG的基础上,借助归一化,空间金字塔和二值描述的优势,提出了归一化BRIEF(NP-BRIEF)和二进制金字塔梯度方向直方图(B-PHOG)特征,对部分遮挡和光照变化具有一定的适应性。

为了适应外观的逐渐变化,分类器必须具备在线学习的能力。分类器主要包括离线和在线两种学习模式。由于在线学习方法针对序列获得的样本可以减少重复学习的时空开销,从而近年来得到广泛的关注。根据分类器种类的不同,现有方法包括,在线Boosting学习方法^[3,14]、在线子空间学习方法^[15]、在线随机森林学习方法^[4]、在线随机朴素贝叶斯学习方法^[5]、在线最近邻学习方法^[16]、在线支持矢量机学习方法^[6]等。由于随机树分类器具有容易联合多特征、训练和测试速度较快且适合并行、无需构造多个二值分类器即可实现多类判别^[5]等优点,在视觉问题中得到广泛应用。而随机朴素贝叶斯分类器依据朴素贝叶斯的特征独立假设,简化了随机森林的结构,可以产生与其相似的性能^[5]。综合考虑效率和性能,配合两类特征的二值描述方式,设计了一种在线混合随机朴素贝叶斯分类器用于视觉跟踪。

已知分类器在图像中搜索目标的过程为目标定位。常用的方法包括滑动窗口^[17-18]、广义霍夫变换^[19]、广义距离变换^[20]等。文献[18]指出,滑动窗口定位方法,参数简单,对后续识别有足够的表达力,且更容易真值标记。因此目标由矩形框近似表达,选取简单的滑动窗口方法实现目标定位。

结合纹理和形状特征的分类器在目标分类检测领域得到了广泛研究,如文献[7]通过离线学习融合金字塔词袋直方图(PHOW)和PHOG特征的随机森林或随机朴素贝叶斯分类器实现图像分类;文献[21]通过增量Boosting方法学习包含纹理和形状的视觉字母表实现目标检测。自适应整合特征的跟踪方法,多以粒子滤波为框架,如文献[22-23],但在基于检测的视觉跟踪方法中,尚未有在线选择纹理和形状特征的相关研究。混合随机朴素贝叶斯分类器融合纹理和形状两类特征,并通过分类器误差调整混合系数,实现分类器的在线学习,而文献[5]只考虑了纹理信息,以直方图作为弱学习器,并通过遗忘因子实现弱学习器的在线更新。

3 在线混合随机朴素贝叶斯视觉跟踪器

跟踪器在纹理和形状两类二值特征描述下,通过混合随机朴素贝叶斯分类器,由滑动窗口方法根据分类器给出的最大类别后验概率预测目标位置,并通过在线参数估计学习目标外观。

3.1 特征描述

由 NP-BRIEF 特征和 B-PHOG 特征描述目标的纹理和形状,分别记为 f_T 和 f_S 。由于特征通过构造空间金字塔同时考虑整体信息和局部信息,且由相对大小关系建模特征,描述方法对部分遮挡和光照变化具有一定的适应能力。

3.1.1 NP-BRIEF

NP-BRIEF 特征是在归一化空间金字塔上定义的 BRIEF 特征^[9],归一化描述有利于多尺度目标的特征共享。对归一化尺度空间 $s_j=(w_j, h_j) \in [0, 1]$ 构建 L 级空间金字塔,在第 l 级 ($0 \leq l < L$) 空间金字塔中随机生成 G 组,每组 D 个坐标点对 (x_{ij}^l, y_{ij}^l) ($0 \leq i \leq G, 0 \leq j \leq D$),映射到图像尺度空间 $s_l=(w_l, h_l)$,得到点对中各采样点的像素强度 I ,通过比较大小得到第 l 级金字塔第 i 组第 j 个特征 f_{ij}^l 为

$$f_{ij}^l(p; x_{ij}^l, y_{ij}^l) = \begin{cases} 1, & \text{iff } I(p; s_l \cdot x_{ij}^l) \geq I(p; s_l \cdot y_{ij}^l) + \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

式中 p 为图像块, ε 为小量, \cdot 运算标识对应元素相乘。在归一化尺度空间,由各向同性高斯核 $(x^l, y^l) \sim N(c_l, 1/25 \times [\max(s_j)/2]^2)$ 生成采样点对,其中 c_l 为各级空间金字塔包含元组的中心。规定组间特征不相容,组内特征独立,据此生成二值编码,则每组对应 2^D 个可能的编码形式,每个二值编码对应一个十进制数(范围为 $0 \sim 2^{D-1}$),将 L 级空间金字塔,每级 G 组特征组合,即可得到 NP-BRIEF 特征,其中 \oplus 为级联操作。

$$f_T := \bigoplus_{0 \leq l \leq L-1} \{ \bigoplus_{1 \leq i \leq G} [\sum_{1 \leq j \leq D} f_{ij}^l(p; x_{ij}^l, y_{ij}^l) \times 2^{j-1}] \}. \quad (2)$$

图 1(a)~(d)以 $G=2, D=5, L=2$ 为例说明纹理特征的描述过程。对图像块构建 $l=0$ 级空间金字塔[图 1(b)],随机生成 2 组,每组 5 个坐标点对[图 1(a)],根据(1)式将点对映射到图像块,并比较采样点像素大小关系,得到两组二值特征 $f_1^0=11010$ 与 $f_2^0=10100$,分别对应十进制数 $f_1^0=26$ 和 $f_2^0=20$;同理,根据构建的 $l=1$ 级空间金字塔[图 1(d)]和随机生成的坐标点对[图 1(c)],得到两组二值特征 $f_1^1=01100$ 与 $f_2^1=00011$,分别对应十进制数 $f_1^1=12$ 和 $f_2^1=3$,则根据(2)式 NP-BRIEF 特征为 $f_T=\{(26, 20), (12, 3)\}$ 。

3.1.2 B-PHOG

为了配合随机朴素贝叶斯分类器,对 PHOG^[13]进行二值描述得到 B-PHOG 特征。记第 l 级金字塔的 PHOG 为 h^l ,随机生成 G 组,每组 D 个等级对 (x, y) ,映射到 PHOG 得到各等级的方向密度值,通过比较大小得到第 l 级金字塔第 i 组第 j 个特征 f_{ij}^l ,

$$f_{ij}^l(h^l; x_{ij}^l, y_{ij}^l) = \begin{cases} 1, & \text{iff } h^l(x_{ij}^l) \geq h^l(y_{ij}^l) + \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

等级对采样服从均匀分布,若每个元组包含 K 个等级,则第 l 级金字塔第 c 个元组的采样公式为 $x^l \sim U[(c-1) \times K + 1, c \times K], y^l \sim U[(c-1) \times K + 1, C \times K]$ 。与 NP-BRIEF 类似, L 级空间金字塔的 B-PHOG 特征可数值化,其中 \oplus 为级联操作。

$$f_S := \bigoplus_{0 \leq l \leq L-1} \{ \bigoplus_{1 \leq i \leq G} [\sum_{1 \leq j \leq D} f_{ij}^l(h^l; x_{ij}^l, y_{ij}^l) \times 2^{j-1}] \}. \quad (4)$$

图 1(e)~(h)以 $K=20, G=2, D=2, L=2$ 为例说明形状特征的描述过程。对图像块构建 $l=0$ 级空间金字塔[图 1(f)],随机生成 2 组,每组 2 个等级对[图 1(e)],根据(3)式将等级对映射到 PHOG,并比较方向密度值的大小关系,得到两组二值特征 $f_1^0=01$ 与 $f_2^0=11$,分别对应十进制数 $f_1^0=1$ 和 $f_2^0=3$;同理,根据构建的 $l=1$ 级空间金字塔[图 1(h)]和随机生成的等级对[图 1(g)],得到两组二值特征 $f_1^1=00$ 与 $f_2^1=11$,分别对应十进制数 $f_1^1=0$ 和 $f_2^1=3$,则根据(4)式 B-PHOG 特征为 $f_S=[(1, 3), (0, 3)]$ 。

3.2 分类器训练

分类器训练,主要解决样本选择与在线学习两个环节。定义样本 $x(f, c)$,其中 $f \in f_T \cup f_S$ 是样本的 NP-BRIEF

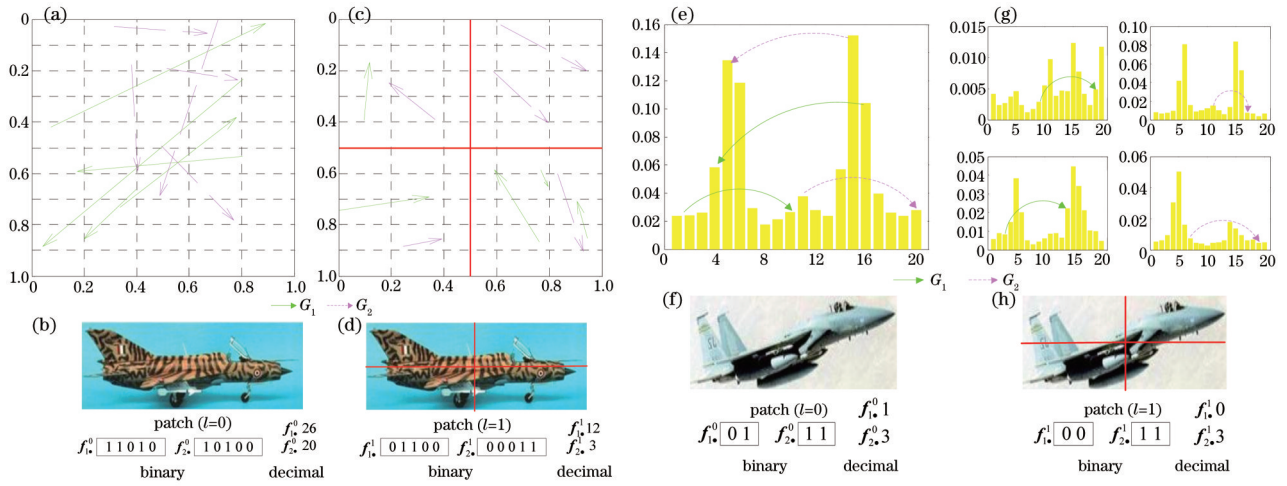


图 1 纹理和形状特征描述。(a)~(d) NP-BRIEF 纹理描述; (e)~(h) B-PHOG 形状描述

Fig.1 Description of texture and shape feature. (a)~(d) NP-BRIEF texture descriptor; (e)~(h) B-PHOG shape descriptor and B-PHOG 特征, $c \in (-1, +i) (i \in \mathbb{N}^+)$ 是样本类别标记, -1 表示负样本 (背景), $+i$ 表示第 i 类正样本 (目标)。

3.2.1 样本选择

与离线学习相比,在线学习能够获得的样本数量有限,面临小样本学习问题;且每次增量学习只能提供给分类器单帧图像和目标区域 w , 正样本数量远小于负样本数量,同时面临不平衡样本学习问题。对此,常根据目标图像块变换出正样本并在非目标区域选择负样本,例如文献[24]根据位移量和目标尺度,在当前图像中生成不同尺度的扫描框 s , 计算 s 与 w 的 Jaccard 距离 J , 选取 $J \geq \sigma_1$ 的 s 为正样本集合 P_s , $J \leq \sigma_2$ 的 s 为负样本集合 N_s 。显然,这种采样方式没有考虑介于 P_s 和 N_s 之间的样本,它们是最容易干扰分类器的信息,定义 $\sigma_2 < J < \sigma_1$ 的 s 为中间样本集合 M_s , 其中 σ_1, σ_2 反映了跟踪结果的精度,由于跟踪存在误差,根据 σ_1 和 σ_2 可以在已有跟踪结果的基础上,尽可能保证跟踪器能够学习到无误差目标。

由于跟踪过程中目标外观具有连续性,如文献[24]假设正样本之间存在相似或仿射变换,同时考虑图像噪声,运动模糊和光照的影响,对变换 P_s 生成 n 个变换样本 P'_s ,

$$P'_s = G(\delta) \otimes (H \times P_s) + V, \quad (5)$$

式中 $G(\delta)$ 为高斯卷积核, H 为变换关系, 高斯白噪声 $V \sim \mathcal{N}(0, \Sigma)$, 考虑到 NP-BRIEF 和 B-PHOG 不敏感光照变化,因此变换模型未考虑光照变换。 P_s 和 P'_s 组成目标样本,而背景样本来自全部中间样本 M_s 和部分负样本 N_s 。

以 Boy 测试集^[25]为例,已知目标区域 w , 首帧选择的正样本 P_s , 变换样本 P'_s , 部分中间样本 M_s 和负样本 N_s 如图 2 所示。采样参数 $\sigma_1=0.95, \sigma_2=0.2$, 得到 3 个如图 2(b) 红框标注的 P_s , 1772 个 M_s 和随机采样的 1000 个 N_s ; 选取仿射模型 $H = R_\theta R_\phi \text{diag}(\lambda_1, \lambda_2) R_{-\phi}$, θ, ϕ 在 $[-1/18\pi, 1/18\pi]$ 区间上均匀采样, $\lambda_1 = \lambda_2 = 1$, 卷积核带宽随

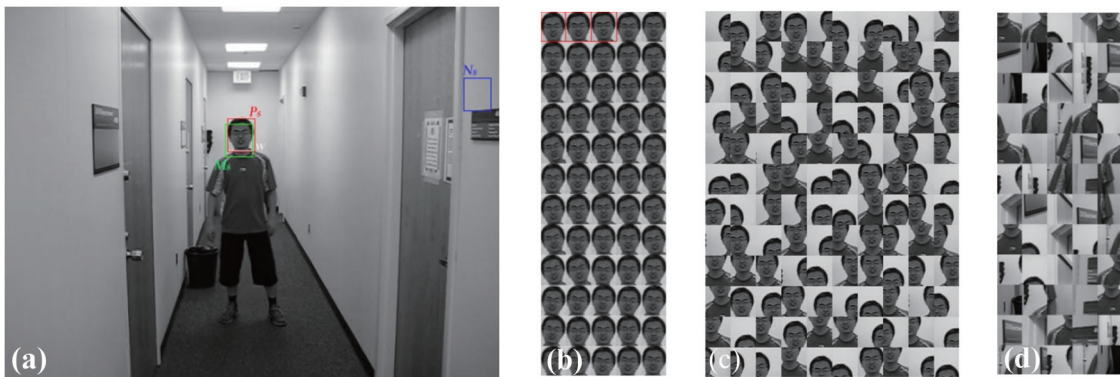


图 2 样本采样示例。(a) 已知 w 生成样本; (b) P_s 和 P'_s ; (c) M_s ; (d) N_s 。

Fig.2 Examples of samples. (a) Generate samples with known w ; (b) P_s and P'_s ; (c) M_s ; (d) N_s .

机选择,噪声水平 $\Sigma=25$,变换个数 $n=20$,则共有60个 P'_s 。

3.2.2 在线学习

由于3.1节定义的NP-BRIEF特征 f_T 和B-PHOG特征 f_s 属于多模态表达,包含各特征组间的多模态,混合系数为常数1;空间金字塔的多模态,混合系数为 τ_l 和 γ_l ;特征空间的多模态,混合系数为 α 和 β ,则特征属于第 i 类目标的后验概率可由混合朴素贝叶斯公式估计。

$$P(c=if) = \alpha \sum_{l=0}^{L-1} \tau_l \sum_{g=1}^G P_i^{l,g}(c=if_T) + \beta \sum_{l=0}^{L-1} \gamma_l \sum_{g=1}^G P_i^{l,g}(c=if_s), \quad (6)$$

式中待估计参数包括混合系数 (α, β) , (τ_l, γ_l) 以及第 l 层金字塔第 g 组纹理特征属于第 i 类目标的后验概率 $P_i^{l,g}(c=if_T)$ 和第 l 层金字塔第 g 组形状特征属于第 i 类目标的后验概率 $P_i^{l,g}(c=if_s)$ 。由于组内特征独立, $P_i^{l,g}(c=if_T)$ 和 $P_i^{l,g}(c=if_s)$ 可由朴素贝叶斯公式估计,它是在贝叶斯公式的基础上,假设特征 f 的 D 个分量 f_d 独立,从而估计类别的后验概率,

$$P_i^{l,g}(c=if) = \frac{P_i^{l,g}(f|c=i)P_i^{l,g}(c=i)}{P_i^{l,g}(f)} = \frac{\prod_{d=1}^D P_i^{l,g}(f_d|c=i)P_i^{l,g}(c=i)}{P_i^{l,g}(f)}, \quad (7)$$

式中 $f=f_T$ 或 f_s ,先验 $P_i^{l,g}(c)$ 通常假设为均匀分布,且总概率 $P_i^{l,g}(f)$ 与 c 无关。由于 D 个 f_d 可以编码为 2^D 个特征 F ,从而通过考虑拉普拉斯平滑的极大似然估计直接求解 $k \in F$ 的 $P_i^{l,g}$,其中 $N_{k,i}$ 为第 i 类样本 $F=k$ 的样本个数, N_i 为第 i 类样本的个数, N 为样本类别总数。

$$P_i^{l,g}(c=if) \propto \prod_{d=1}^D P_i^{l,g}(f_d|c=i), \quad (8)$$

$$\prod_{d=1}^D P_i^{l,g}(f_d|c=i) \propto P_i^{l,g}(F=k|c=i) = \frac{N_{k,i} + 1}{N_i + N}. \quad (9)$$

表1给出了跟踪器的在线学习算法,主要思路是通过极大似然估计在线更新第 l 层金字塔第 g 组特征的类别后验概率 $P_i^{l,g}$,通过交叉验证学习混合系数,从而更新特征属于第 i 类目标的后验概率 $P(c=if)$ 。具体步骤包括:1)先将负样本集合 N_s 划分为不相交的训练负样本集合 N_s^T 和验证负样本集合 N_s^V ,由正样本 P_s ,变换正样本 P'_s ,中间样本 M_s ,训练负样本 N_s^T 组成训练样本集合 S^T ,由验证负样本 N_s^V 组成验证样本集合 S^V (如步骤1);2)根据 S^T 中样本的类别标记 c ,和由(6)式定义的measure操作估计的类别后验概率 P ,更新分类器:若样本为背景,则直接更新;若样本为第 i 类目标,则当后验概率 P 小于阈值 T_{th} 时更新分类器,更新操作update依据样本权值 w 由 $(N_{k,i} + w_{k,i} + 1)/(N_i + w_{k,i} + w_{k,-1} + N)$ 在线更新 $P_i^{l,g}$ (如步骤3~10),其中 $w_{k,i}$ 为特征 $F=k$ 属于第 i 类目标的权值, $w_{k,-1}$ 为特征 $F=k$ 属于背景的权值, $N_{k,i}$, N_i , N 与(9)式定义相同;3)通过交叉验证,估计 S^V 中样本第 l 层金字塔纹理特征属于目标的后验概率 P_i^l ,与纹理特征空间属于目标的后验概率 P_i^l ,并根据 T_{th} 确定各级金字塔的分类错误次数 λ_i^l 和特征空间的分类错误次数 λ_i (如步骤12~23),以及形状特征的 P_i^l , P_s 和 λ_i^l, λ_s (与步骤12~23的计算方法相同),从而估计各特征中每级金字塔的错误率 e_i^l 和 e_s^l ,各特征的错误率 e_i 和 e_s ,与对应的分类器投票权重 $\Delta_i^l, \Delta_i, \Delta_s^l, \Delta_s$ (如步骤24~25),即可根据学习率参数 θ 调整混合系数 (α, β) 和 (τ_l, γ_l) (如步骤27~28)。

3.3 目标定位

根据已有的分类器,在 t 时刻跟踪第 i 类目标,首先需要根据 $t-1$ 时刻的目标区域 \tilde{w}_{t-1} ,定义 t 时刻的目标搜索区域 R_t ,并根据目标可能的尺度建立扫描窗口 w_t ,通过 $f(w)$ 计算窗口NP-BRIEF特征 f_T 和B-PHOG特征 f_s ,则满足(10)式的 \tilde{w}_t 即为 t 时刻的目标区域。

$$\tilde{w}_t = \arg \max_{w_t \in R_t} P[c=if(w_t)]. \quad (10)$$

令 \tilde{w}_{t-1} 和 w_t 的中心分别为 (u_{t-1}, v_{t-1}) , (u_t, v_t) ,则定义矩形扫描区域 $R_t = \{(u, v) | |u - u_{t-1}| < r_w, |v - v_{t-1}| < r_h\}$,其中 r_w, r_h 分别为矩形宽高扫描半径。图3给出目标定位过程,目标由矩形区域描述,矩形搜索区域的扫描半径 $r_w=46, r_h=50$ 。图3(a)标记出 $t-1$ 时刻的目标区域 \tilde{w}_{t-1} 与 t 时刻的搜索区域 R_t 和部分扫描窗口 w_t ;图3(b)是 t 时刻 R_t

中各 w 的目标信任图与其分布图;图 3(c)是通过非极大值抑制在信任图中搜索到的 t 时刻目标位置 \tilde{w}_t 。

表 1 混合朴素贝叶斯视觉跟踪器的在线学习算法

Table 1 Online learning algorithm of mixture Naïve Bayes visual tracker

Algorithm 1 Online learning of tracker
Input: Training samples $\mathbf{x}(f,c)$, class label $c \in \{-1, +i\} (i \in N^+)$
Input: mixture factor $(\alpha, \beta) = (0.5, 0.5)$, $(\tau_l, \gamma_l) = (1.0/L, 1.0/L)$, threshold $T_{th} = 0.5$, learning rate $\theta = 0.8$
Output: Tracker T
1: Training samples $S^T = P_s \cup P'_s \cup M_s \cup N_s^T$, validation samples $S^V = N_s^V$ with capacity M
2: //update
3: for each \mathbf{x} in S^T do
4: $(P, \sim) = \text{measure}(T, \mathbf{x}, f)$
5: if $x.c == -1$ then
6: $w_{k,i} = 0$ $w_{k,-1} = 1$ $T = \text{update}(T, \mathbf{x}, f, w_{k,i}, w_{k,-1})$
7: else if $x.c == i$ && $P < T_{th}$ then
8: $w_{k,i} = 1$ $w_{k,-1} = 0$ $T = \text{update}(T, \mathbf{x}, f, w_{k,i}, w_{k,-1})$
9: end if
10: end for
11: //validate
12: for each x in S^V do
13: for each l in L do
14: $(\sim, P_l^i) = \text{measure}(T, \mathbf{x}, f_l)$
15: if $P_l^i > T_{th}$ then
16: $\lambda_l^i = \lambda_l^i + 1$
17: end if
18: end for
19: $P_i = \sum_{l=0}^{L-1} P_l^i / L$
20: if $P_i > T_{th}$ then
21: $\lambda_i = \lambda_i + 1$
22: end if
23: end for
24: $e_i^l = \lambda_i^l / M$, $e_i = \lambda_i / M$, $\Delta_i^l = \ln[(1 - e_i^l) / (e_i^l + \varepsilon)]$, $\Delta_i = \ln[(1 - e_i) / (e_i + \varepsilon)]$
25: $e_i^l = \lambda_i^l / M$, $e_i = \lambda_i / M$, $\Delta_i^l = \ln[(1 - e_i^l) / (e_i^l + \varepsilon)]$, $\Delta_i = \ln[(1 - e_i) / (e_i + \varepsilon)]$
26: //adjust
27: $\tau_i = \theta \times \tau_i + (1 - \theta) \times (\Delta_i^l / \sum_{l=0}^{L-1} \Delta_i^l)$ $\gamma_i = \theta \times \gamma_i + (1 - \theta) \times (\Delta_i^l / \sum_{l=0}^{L-1} \Delta_i^l)$
28: $\alpha = \theta \times \alpha + (1 - \theta) \times [\Delta_i / (\Delta_i + \Delta_i)]$ $\beta = \theta \times \beta + (1 - \theta) \times [\Delta_i / (\Delta_i + \Delta_i)]$

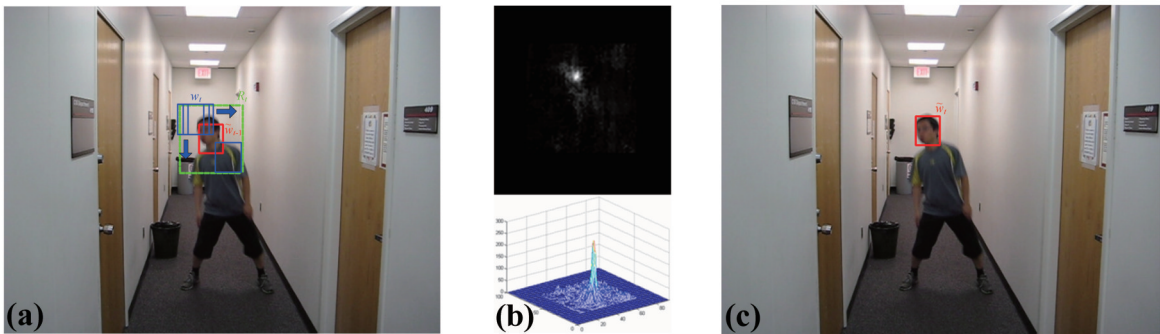


图 3 目标定位过程。(a) \tilde{w}_{i-1} , w 和 R ; (b) 信任图; (c) \tilde{w}_t

Fig.3 Procedure of object localization. (a) \tilde{w}_{i-1} , w and R ; (b) confidence map; (c) \tilde{w}_t

当由(10)式得到 \tilde{w}_t 对应的后验概率 $P < 0.5$ 时, 可以认为跟踪器还不能很好地识别外观, 需要在线学习, 从而根据 \tilde{w}_t 按照 3.2.1 节选择样本, 再由 3.2.2 节在线学习跟踪器。

4 实验与结果分析

实验硬件平台选择 Windows X86 32 位 PC 机,配置 Intel Core i5 3470 CPU(3.2 GHz)和 3.0 G 内存,算法通过 VS2008 开发环境,由 VC++ 和 OPENCV2.4.4 实现,针对单目标跟踪应用场景,选取已标注真值的 8 组基准测试集^[25],从性能,时空开销和混合系数学习三个方面验证了当目标存在尺度变化(Boy, David),非刚性变化(David, Tiger),面内旋转(Boy, Coke, David, Deer, Face, Girl, Suv, Tiger),离面旋转(Boy, Coke, David, Face, Girl, Tiger),快速运动(Boy, Coke, Tiger),运动模糊(Boy, David, Deer, Tiger),以及光照变化(Coke, David, Face, Tiger),遮挡(Coke, David, Face, Girl, Suv, Tiger)等情况时跟踪器的能力。

4.1 性能

记第 t 帧目标真实中心为 $c_g(t)$,真实矩形区域为 $r_g(t)$,跟踪器预测的中心位置为 $c_i(t)$,预测的矩形区域为 $r_i(t)$ 。本文方法 rnb 所需参数金字塔层数 $L=3$,特征组数 $G=10$,组内特征数 $D=13$,样本采样参数 $\sigma_1=0.95, \sigma_2=0.2$,负样本采样个数 $N_s=1000$,变换样本数 $n=20$,特征混合系数 (α, β) 初值 $(0.5, 0.5)$,各特征空间金字塔混合系数 (τ_i, γ_i) 初值 $(1.0/L, 1.0/L)$,判别阈值 $T_{th}=0.5$,学习率 $\theta=0.8$,在全部测试集中固定;且以比例 0.6,扩充 $t-1$ 时刻目标区域的宽高作为当前帧的搜索区域。由于随机性的存在,rnb 将单个序列执行五次,取中值作为结果;对照算法 oab^[3],orf^[4],struck^[6]的源码由文献提供并选择默认参数。记 $r_g(t)$ 与 $r_i(t)$ 之间的 Jaccard 距离为第 t 帧的重叠率,记 $c_g(t)$ 与 $c_i(t)$ 之间的欧式距离为第 t 帧的中心误差,则各测试集的平均重叠率和中心误差如表 2 所示。由表 2 可知,rnb 针对大多数测试集可以给出较高的重叠率和较小的中心误差。其中 David 测试集的 orf 和 struck 结果,没有文献[4]和文献[6]给出的结果好,是因为文献中忽略了前段全局光照变化较大的序列,全局光照变化严重影响了 orf 和 struck 的性能;Deer 测试集由于 orf 在第 5 帧发生了漂移,因此重叠率较小且中心误差较大;Suv 测试集第 29 帧至 59 帧目标进出视场,由于四个方法都没有考虑目标出视场的情况,因此从第 60 帧开始跟踪。

表 2 各测试集的平均重叠率和平均中心误差

Table 2 Average overlap ratio and centroid error in each benchmark

Sequence	Frames	Resolution /pixel	Average overlap ratio				Average centroid error /pixel			
			oab	orf	struck	rnb	oab	orf	struck	rnb
Boy	602	640×480	0.78	0.75	0.76	0.80	3.3	3.7	3.7	2.8
Coke	291	640×480	0.65	0.54	0.60	0.70	11.6	19.2	12.8	9.8
David	761	320×240	0.28	0.16	0.12	0.77	30.5	49.7	44.3	6.5
Deer	71	704×400	0.74	0.04	0.74	0.74	5.0	232.6	5.5	5.9
Face	812	320×240	0.75	0.79	0.78	0.79	8.3	5.6	6.5	5.0
Girl	500	128×96	0.72	0.63	0.74	0.73	3.5	7.6	2.9	2.9
Suv	886	320×240	0.53	0.50	0.60	0.84	29.4	48.8	28.3	4.2
Tiger	354	640×480	0.24	0.32	0.58	0.72	88.5	58.6	20.0	10.8

图 4 给出各测试序列在某一帧的跟踪结果,可见在部分遮挡的情况下[图 4(b)],struck 和 rnb 都给出了较好的跟踪结果,但由于多次遮挡[图 4(g)]导致除 rnb 以外的跟踪器发生漂移;在光照全局变化的情况下[图 4(c)],由于 rnb 选用的二值特征,具有对全局光照变化的不敏感性,只有 rnb 给出了正确的跟踪结果;而在局部光照变化的情况下[图 4(h)],struck,orf,rnb 可以给出正确的跟踪结果,但 orf 的中心定位误差较大;在面内旋转的情况下[图 4(e)],四个方法都能正确跟踪,但 oab 有较大的误差;在运动模糊的情况下[图 4(d)],由于四个方法选择了统计特征,因此可以适应运动模糊;在离面旋转的情况下[图 4(f)],四个方法都能正确跟踪目标,但 rnb 和 orf 有较大的误差;在尺度变化的情况下[图 4(a)],由于四个方法都没有考虑尺度搜索,虽然都能给出正确结果,但有很大的定位误差。

文献[25]定义成功图和精度图评价算法的性能,它们展示的是由重叠率或中心误差统计的精度,随重叠率阈值 t_o 或距离阈值 t_c 的变化关系。对跟踪算法的评估,仅考虑精度 P 是不够的,查全率 R 和综合标准 $F^{[24]}$ 同样重要,因此定义由重叠率统计的 P_o, R_o, F_o 随 t_o 的变化关系 [$t_o \in (0, 1)$],以及由中心误差统计的 P_c, R_c, F_c 随 t_c 的变化关系 [$t_c \in (0, 50)$],用于评估全部测试集的跟踪性能。如图 5 所示,由重叠率统计的结果 $P_o-t_o, R_o-t_o, F_o-t_o$ 表

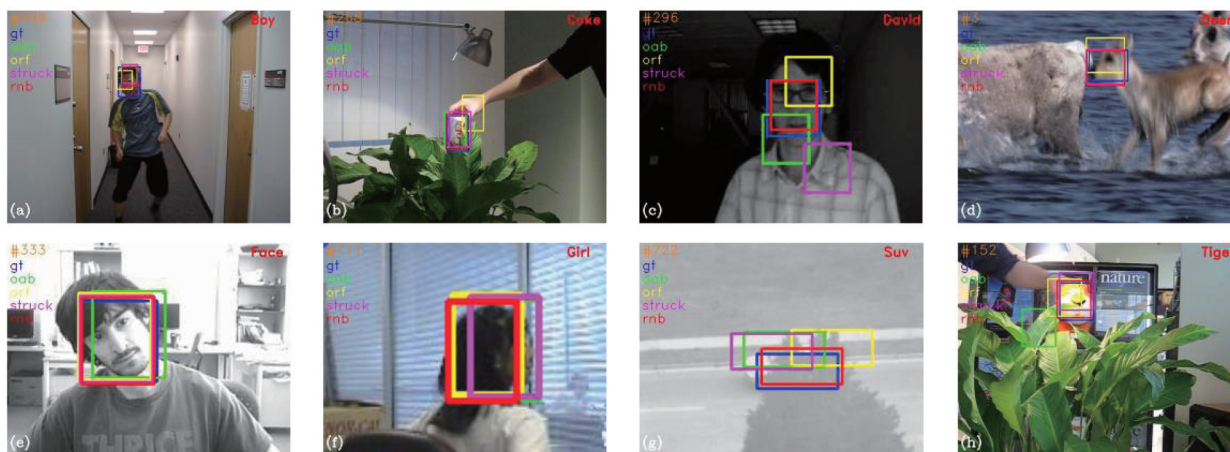


图 4 各测试集的跟踪结果

Fig.4 Tracking results in each benchmark

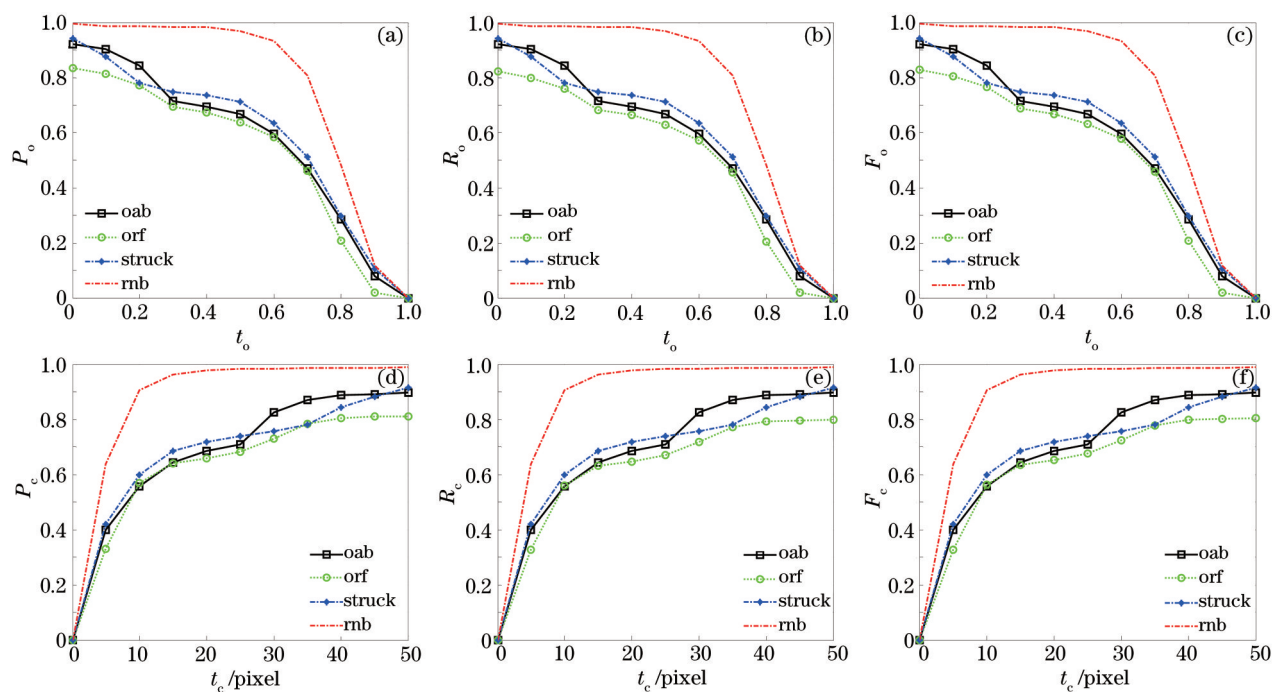


图 5 全测试集的成功图和精度图

Fig.5 Success plot and precision plot of all benchmarks

明, rnb 在 $0.2 < t_o < 0.7$ 时, 明显优于其他算法; 当 $t_o < 0.2$ 时, P_o, R_o, F_o 趋近 1。由中心误差统计的结果 $P_c-t_c, R_c-t_c, F_c-t_c$ 表明 rnb 在 $5 < t_c < 30$ 时, 明显优于其他算法; 当 $t_c > 40$ 时, rnb 的 P_c, R_c, F_c 趋近 1。

4.2 复杂度

图 6 给出对比算法的单帧执行时间, 其中横轴为帧序号, 纵轴为执行时间, 图例中给出各方法的平均执行时间。由图中的结果可知, orf 执行速度最快, oab 次之, 而除 Deer 和 Face 数据集外, rnb 要快于 struck。由于 Deer 和 Face 的目标分辨率较大, rnb 对应较多的搜索位置, 因此需要更多的执行时间。由 Suv 测试集执行时间可知, 当算法无需在线学习仅执行跟踪时, rnb 的速度要明显快于其他方法。

学习分辨率为 $r \times c$ 图像的外观, 最基本的方法是, 记录这个图像块, 若有 M 个目标, 以 8 位灰度图像记录, 则需要记录 $M \times r \times c \times 8$ 位的存储空间, 且随样本个数的增多存储空间不断增大。而基于分类器的跟踪方法, 通过寻找不变特征和在线学习, 在保证识别性能的前提下, 压缩特征空间。若 G 组特征, 每组包含 2^D 个可能的特征编码, 以 32 位浮点数存储后验概率, 如果采用基于内容的特征建模方法^[24], 与目标的尺度量级相关, 则 S 个尺度空间需要存储空间为 $M \times S \times G \times 2^D \times 32$ 位; 而采用归一化尺度空间, 需要的存储空间为 $M \times G \times 2^D \times 32$

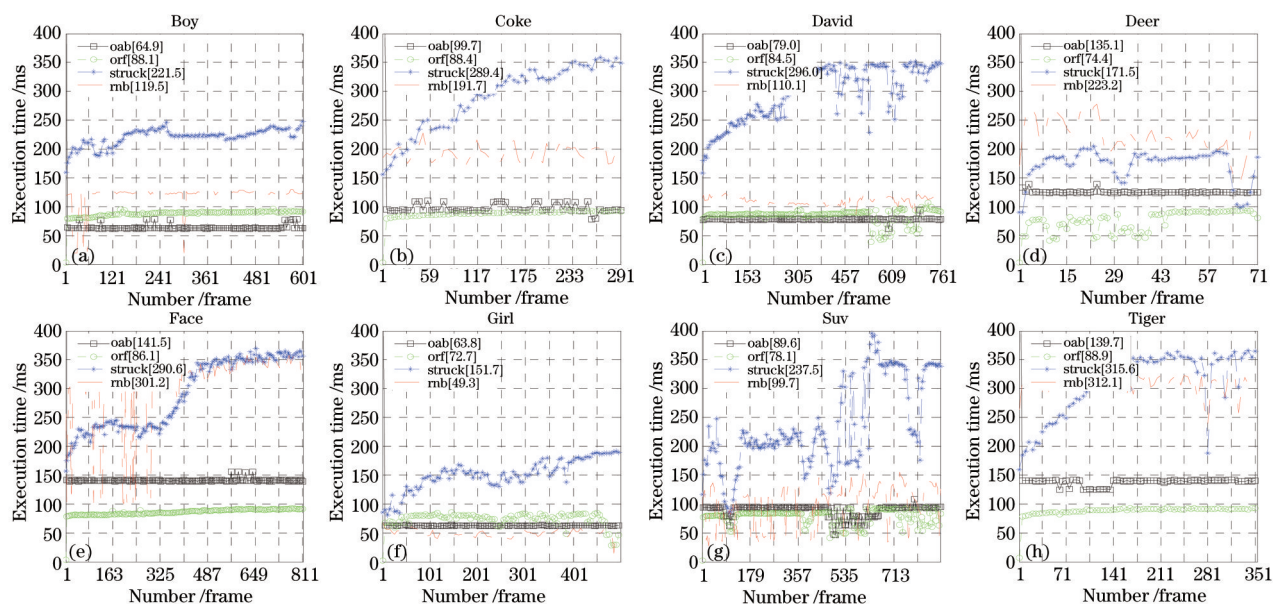


图 6 单帧执行时间

Fig.6 Execution time in one frame

位,且整个实验过程中,所需存储空间是不发生改变的。

4.3 混合系数学习

图 7 给出 Suv 数据集混合系数随时间的变化情况,其中横轴为帧序号,纵轴为混合系数的值。混合系数 α, β 是对纹理和形状特征加权,较大的权值反映当前时刻某一特征更具判别力。由于目标在未遮挡时形状占优,而在遮挡时纹理占优,图 7(a)中 α, β 的变化准确地反映了这一过程,其中矩形框标注出第 731 帧 (α, β) = (0.4959, 0.5041),此时目标正脱离遮挡。而纹理和形状特征的空间金字塔混合系数 τ 与 γ 是对各自特征的空间信息加权,较大的权值反映当前时刻某一空间更具判别力;图 7(b)为第 603 帧目标出现局部遮挡,此时纹理特征的第 2 级金字塔混合系数 τ_1 占优;图 7(c)为第 480 帧目标出现大面积遮挡,此时形状特征的第 3 级金字塔混合系数 γ_2 占优。可见混合系数 $\alpha, \beta, \tau, \gamma$ 结合学习率参数 θ 可以根据当前场景与历史场景,选择具有判

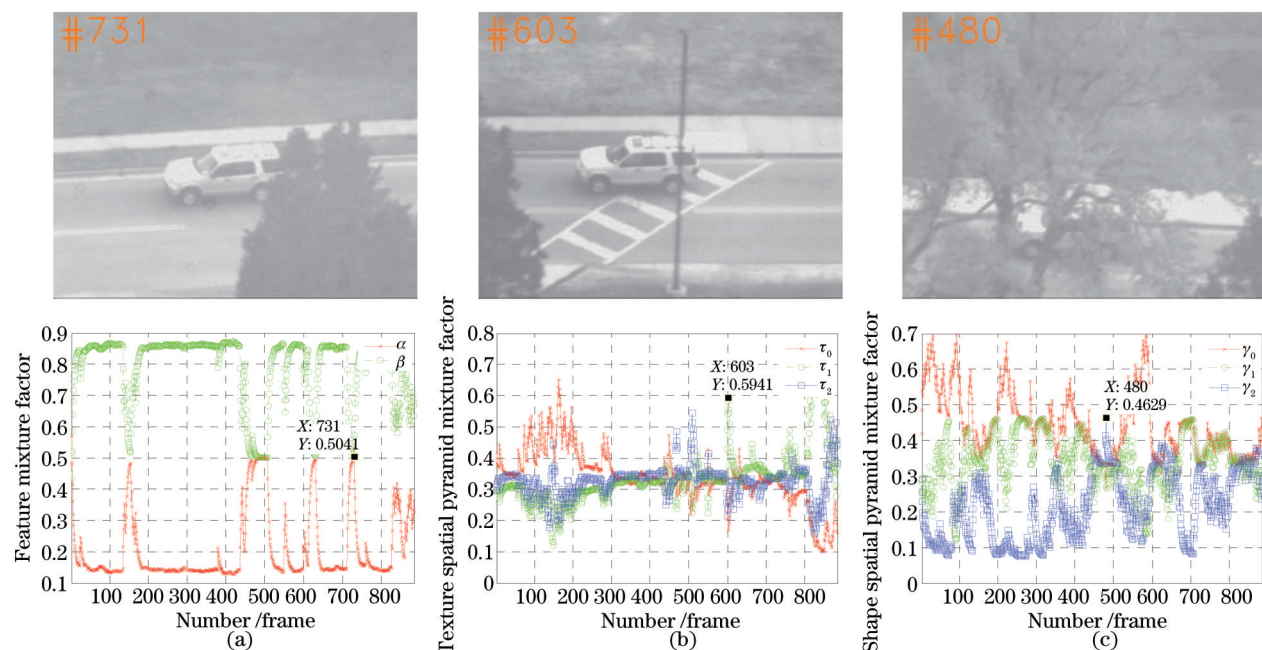


图 7 Suv 数据集混合系数学习示例。(a) 特征混合系数 α, β ; (b) 纹理空间金字塔混合系数 τ ; (c) 形状空间金字塔混合系数 γ

Fig.7 Paradigm of mixture coefficient learning in Suv dataset. (a) Feature mixture factor α, β ; (b) texture spatial pyramid mixture factor τ ; (c) shape spatial pyramid mixture factor γ

别力的特征和空间。

5 结 论

为了充分利用视频图像提供的丰富外观信息,定义了纹理和形状的二值描述方式,在此基础上提出了混合朴素贝叶斯跟踪器,通过在线学习选择判别特征并更新目标外观,从而实现目标外观和场景变化时的稳定跟踪。实验表明该方法具有以下优点:1) 由于考虑了空间金字塔,在目标部分遮挡时仍能正确跟踪;2) 相对大小特征描述方式使得跟踪结果不受光照变化的影响;3) 在线特征选择和外观学习提高了跟踪器对目标和场景变化的适应能力。不足的是,特征描述的主要是二维信息,存在离面旋转时,跟踪误差较大。

参 考 文 献

- 1 S Avidan. Support vector tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064–1072.
- 2 O Chapelle, B Schölkopf, A Zien. Semi-Supervised Learning [M]. London: The MIT Press, 2006.
- 3 H Grabner, H Bischof. On-line boosting and vision [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2006, 1: 260–267
- 4 A Saffari, C Leistner, J Santner, *et al.*. On-line random forests [C]. IEEE International Conference on Computer Vision Workshop, 2009: 1393–1400.
- 5 M Godec, C Leistner, A Saffari, *et al.*. On-line random naive bayes for tracking [C]. IEEE International Conference on Pattern Recognition, 2010: 3545–3548.
- 6 S Hare, A Saffari, P H Torr. Struck: Structured output tracking with kernels [C]. IEEE International Conference on Computer Vision, 2011: 263–270.
- 7 A Bosch, A Zisserman, X Muoz. Image classification using random forests and ferns [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1–8.
- 8 M Heikkila, M Pietikainen. A texture-based method for modeling the background and detecting moving objects [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 657–662.
- 9 M Calonder, V Lepetit, C Strecha, *et al.*. Brief: Binary robust independent elementary features [C]. European Conference on Computer Vision, 2010: 778–792.
- 10 S Belongie, J Malik, J Puzicha. Shape matching and object recognition using shape contexts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509–522.
- 11 B Wu, R Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors [J]. International Journal of Computer Vision, 2007, 75(2): 247–266.
- 12 N Dalal, B Triggs. Histograms of oriented gradients for human detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2005, 1: 886–893.
- 13 A Bosch, A Zisserman, X Munoz. Representing shape with a spatial pyramid kernel [C]. ACM International Conference on Image and Video Retrieval, 2007: 401–408.
- 14 S Avidan. Ensemble tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261–271.
- 15 D A Ross, J Lim, R S Lin, *et al.*. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision, 2008, 77(1): 125–141.
- 16 S Gu, Y Zheng, C Tomasi. Efficient visual object tracking with online nearest neighbor classifier [C]. Asian Conference on Computer Vision, 2010: 271–282.
- 17 P Viola, M Jones. Rapid object detection using a boosted cascade of simple features [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2001, 1: 1–511–1–518.
- 18 C H Lampert, M B Blaschko, T Hofmann. Efficient subwindow search: a branch and bound framewrok for object localization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 31(12): 2129–2142.
- 19 B Leibe, A Leonardis, B Schiele. Robust object detection with interleaved categorization and segmentation [J]. International Journal of Computer Vision, 2008, 77(1–3): 259–289.
- 20 P F Felzenszwalb, D P Huttenlocher. Pictorial structures for object recognition [J]. International Journal of Computer Vision, 2005, 61(1): 55–79.

- 21 A Opelt, A Pinz, A Zisserman. Learning an alphabet of shape and appearance for multi-class object detection [J]. International Journal of Computer Vision, 2008, 80(1): 16-44.
- 22 Zhang Hui, Zhao Baojun, Tang Linbo, *et al.*. Infrared object tracking based on adaptive multi-features integration [J]. Acta Optica Sinica, 2010, 30(5): 1291-1296.
张 辉, 赵保军, 唐林波, 等. 基于自适应多特征整合的红外目标跟踪[J]. 光学学报, 2010, 30(5): 1291-1296.
- 23 Anping Li, Zhongliang Jiang, Shiqiang Hu. Particle filter based visual tracking with multi-cue adaptive fusion [J]. Chin Opt Lett, 2005, 3(6): 326-329.
- 24 Z Kalal, K Mikolajczyk, J Matas. Tracking-learning-detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 34(7): 1409-1422.
- 25 Y Wu, M H Yang, J Lim. Online object tracking: A benchmark [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.

栏目编辑: 苏 岑