

具有尺度和旋转适应性的长时间目标跟踪

熊丹¹ 卢惠民¹ 肖军浩¹ 郑志强¹

摘要 目标发生尺度和旋转变化会给长时间目标跟踪带来很大的挑战, 针对该问题, 本文提出了具有尺度和旋转适应性的鲁棒目标跟踪算法. 首先针对跟踪过程中目标存在的尺度变化和旋转运动, 提出一种基于傅里叶-梅林变换和核相关滤波的目标尺度和旋转参数估计方法. 该方法能够实现连续空间的目标尺度和旋转参数估计, 采用核相关滤波提高了估计的鲁棒性和准确性. 然后针对长时间目标跟踪过程中, 有时不可避免地会出现跟踪失败的情况 (例如由于长时间半遮挡或全遮挡等), 提出一种基于直方图和方差加权的搜索方法. 当目标丢失时, 通过提出的搜索方法能够快速从图像中确定目标可能存在的区域, 使得跟踪算法具有从失败中恢复的能力. 本文还训练了两个核相关滤波器用于估计跟踪结果的置信度和目标平移, 通过专门的核相关滤波器能够使估计的跟踪结果置信度更加准确和鲁棒, 置信度的估计结果可用于激活基于直方图和方差加权的搜索模块, 并判断搜索窗口中是否包含目标. 本文在目标跟踪标准数据集 (Online object tracking benchmark, OTB) 上对提出的算法和目前主流的目标跟踪算法进行对比实验, 验证了本文提出算法的有效性和优越性.

关键词 尺度和旋转, 傅里叶-梅林变换, 核相关滤波, 直方图和方差

引用格式 熊丹, 卢惠民, 肖军浩, 郑志强. 具有尺度和旋转适应性的长时间目标跟踪. 自动化学报, 2019, 45(2): 289-304

DOI 10.16383/j.aas.2018.c170359

Robust Long-term Object Tracking With Adaptive Scale and Rotation Estimation

XIONG Dan¹ LU Hui-Min¹ XIAO Jun-Hao¹ ZHENG Zhi-Qiang¹

Abstract A robust long-term object tracking algorithm with adaptive scale and rotation estimation is proposed to deal with the challenges of scale and rotation changes during long-term object tracking. Firstly, a robust scale and rotation estimation method is proposed based on the Fourier-Mellin transform and the kernelized correlation filter for scale changes and rotation of the object. Then a weighted object searching method based on histogram and variance is proposed to deal with possible tracker's failures in the long-term tracking (due to long-term semi-occlusion or full occlusion, etc.). When the tracked object is lost, the object can be re-located in the image using the proposed searching method, so the tracker can be recovered from failures. Moreover, this paper also trains two other kernelized correlation filters to estimate the object's translation and the confidence of tracking results. The customized kernelized correlation filter can make the estimated confidence more accurate and robust. The confidence is used to activate a weighted object searching module and determine whether the searching windows contain objects or not. The proposed algorithm is compared with the state-of-the-art tracking algorithms using the online object tracking benchmark (OTB). Experimental results validate the effectiveness and superiority of the proposed tracking algorithm.

Key words Scale and rotation, Fourier-Mellin transform, kernelized correlation filter, histogram and variance

Citation Xiong Dan, Lu Hui-Min, Xiao Jun-Hao, Zheng Zhi-Qiang. Robust long-term object tracking with adaptive scale and rotation estimation. *Acta Automatica Sinica*, 2019, 45(2): 289-304

目标跟踪作为计算机视觉领域的一个核心的并且极具挑战性的研究课题, 在多年之前科研人员就已经开始进行深入研究. 大量关于目标跟踪的参考

文献和算法已经发表和提出, 具体的这些文献和算法可以参考目标跟踪的一些综述^[1-5] 或视觉跟踪竞赛^[6-8]. 根据视觉外观的统计模型不同, 可将现有的跟踪方法分为基于生成模型的目标跟踪方法和基于判别模型的目标跟踪方法.

基于生成模型的跟踪方法一般首先构建目标的外观模型, 然后通过拟合该模型在图像中找到最相似的区域作为目标区域. 文献 [9] 提出了能够同时编码目标的颜色和颜色空间分布的空间-颜色混合高斯外观模型. 文献 [10] 提出了一种基于向量空间的统计模型对刚性和铰接式的跟踪目标进行建模. 文献 [11] 提出利用增量式的核主成分分析方法为跟

收稿日期 2017-06-29 录用日期 2017-11-17
Manuscript received June 29, 2017; accepted November 17, 2017

国家自然科学基金 (61403409, 61503401), 中国博士后基金 (2014M562648) 资助

Supported by National Natural Science Foundation of China (61403409, 61503401) and China Postdoctoral Science Foundation (2014M562648)

本文责任编辑 左旺孟

Recommended by Associate Editor ZUO Wang-Meng

1. 国防科技大学智能科学学院 长沙 410073

1. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073

踪目标构建具有稳定更新速度和恒定内存需求的非线性子空间模型. 基于生成模型的跟踪方法普遍对背景信息缺乏考虑, 跟踪子在背景凌乱的环境中效果不是很理想. 基于判别模型的跟踪方法将目标跟踪问题考虑为一个二分类问题, 通过构建分类器来区分跟踪目标和背景. 与仅利用目标外观信息的生成模型不同, 基于判别模型的跟踪子对目标和目标周围环境都进行了建模. Grabner 等利用在线 Boosting 算法训练级联分类器来区分目标和背景^[12]. 文献 [13] 利用压缩特征对目标进行描述, 并通过朴素贝叶斯分类器区分跟踪的目标和背景. Kalal 等将跟踪、检测和学习结合起来, 检测子能够恢复跟踪失败的跟踪子, 而跟踪子的跟踪结果能够为检测子提供具有结构约束的训练样本, 提高了检测子的鲁棒性和区分能力^[14].

近年来, 相关滤波方法在目标检测和跟踪中得到广泛应用, 显示出较好的检测和跟踪性能. 基于相关滤波的目标跟踪方法采用判别式统计模型为目标的视觉外观进行建模, 该方法将两个图像块之间的卷积操作转换为傅里叶域的点乘操作, 计算简单高效. 文献 [15] 提出了一种误差最小平方和相关滤波方法 (Minimum output sum of squared error filter, MOSSE), 在目标跟踪中得到了成功应用. MOSSE 跟踪算法对跟踪过程中的光照、局部非刚性扭曲等变化具有较好的鲁棒性, 并且计算非常简单, 跟踪帧速率甚至能够达到每秒几百帧, 但是 MOSSE 跟踪子并没有考虑目标的尺度和旋转变化. 文献 [16] 提出了一种核相关滤波方法 (Kernelized correlation filter, KCF), 并应用于目标跟踪中, 取得了较好的跟踪效果, 但是 KCF 同样没有考虑目标的尺度和旋转变化问题. 为了解决跟踪过程中的尺度变化问题, 文献 [17] 利用学习的滤波器在原始图像和该图像双线性插值缩放后的图像上检测目标, 通过响应最大的平移位置和响应所在的尺度估计出目标的位置和尺度. 文献 [18] 使用基于 HOG 特征的自适应多尺度相关滤波器 (Discriminative scale space tracker, DSST) 估计目标的尺度, 进一步考虑到基于相关滤波的目标跟踪算法一般都是利用周期性假设产生具有循环结构的训练样本, 这样不可避免在样本中引入了边界效应, 为此在学习滤波器时加入了空间正则项, 能够得到区分能力更强的滤波器^[19]. 基于相关滤波跟踪子对运动模糊和光照变化具有较好的鲁棒性, 但是对目标形变比较敏感, 而基于颜色统计的目标外观模型受到目标的形变影响较小, Bertinetto 等将两者结合起来提高了跟踪子的性能^[20]. 文献 [21] 利用具有循环结构的训练样本训练支持向量机, 得到具有更强区分能力的支持相

关滤波器, 在目标跟踪中得到了成功的应用. 文献 [22] 提出将空间上下文模型和时间上下文模型结合起来学习相关滤波器, 并且根据上下文中不同区域与跟踪目标相似度大小对目标区域进行了加权, 提高了跟踪算法的鲁棒性. Ma 等提出了一种长时间的相关滤波目标跟踪方法, 他们学习了三个相关滤波器, 分别用于估计目标的平移、尺度和跟踪结果的置信度, 除此之外还专门训练了一个随机森林分类器, 用于目标丢失时再检测^[23].

在目标跟踪过程中, 目标出现尺度和旋转变化是一种比较常见的情况. 以往的研究对目标尺度估计考虑较多, 很少会专门针对目标的旋转运动进行分析, 因此当目标存在旋转时, 如何实现鲁棒的视觉跟踪仍然是一个极具挑战的研究问题. 此外, 在目标跟踪过程中, 特别是长时间目标跟踪过程中, 因为存在长时间半遮挡或全遮挡等情况, 跟踪子有时不可避免地会跟踪失败. 如何让跟踪子从失败中恢复过来是实现长时间目标跟踪的前提. 文献 [14] 和文献 [23] 都设计有专门的检测模块用以恢复错误的跟踪子, 但是这两种方法均没有考虑目标的旋转情况, 而且检测子需要处理每一个滑动窗口来判断其是否包含目标, 整个过程比较耗时, 降低了跟踪算法的实时性能. 本文针对上述问题, 提出了具有尺度和旋转适应性的长时间目标跟踪方法. 傅里叶-梅林变换能够将原始图像上的目标尺度和旋转估计转换为频域幅值谱对数极坐标图像上简单的平移估计, 在图像配准领域已经得到了成功应用^[24-26]. 本文受到该方法启发, 将其与核相关滤波结合起来, 提出一种鲁棒的旋转和尺度参数估计方法. 为了让跟踪子具有从失败中恢复过来的能力, 本文提出了基于直方图和方差加权的目标搜索算法, 根据计算的权重采用蒙特卡罗随机抽样原理, 能够从大量的滑动窗口中抽取其中权重比较高的进行处理, 提高了整个检测的实时性能. 本文还训练了两个核相关滤波器专门用于估计目标平移运动和跟踪结果的置信度. 鲁棒准确的跟踪结果置信度估计能够用于激活基于直方图和方差加权的目标搜索模块, 当跟踪失败时, 通过激活搜索模块使得跟踪子能够从失败中恢复过来.

1 基于傅里叶-梅林变换和核相关滤波的尺度和旋转参数估计

大量图像配准方面的研究工作是通过估计图像之间的平移、旋转和尺度参数来得到像素点之间的变换关系^[24-26]. 在目标跟踪过程中, 短时间内可以将目标附近区域看作是近似静止的, 如此就可将图像配准领域得到成功应用的一些成熟方法应用到目标跟踪过程中, 估计前后帧目标之间的旋转和尺度

关系. 但是, 图像配准一般针对两帧独立的图像, 对目标跟踪而言, 仅仅利用前后两帧图像估计目标的尺度和旋转, 会使得估计的结果抗干扰能力不强, 鲁棒性不高. 本文将图像配准领域得到成功应用的傅里叶-梅林变换和核相关滤波方法结合起来, 提高了目标尺度和旋转参数估计的鲁棒性和准确性.

1.1 傅里叶-梅林变换

笛卡尔坐标系下尺度变化和旋转运动转换到对数极坐标下表现为纯粹的平移运动. 对数极坐标变换过程中在变换中心附近密集采样, 在远离中心附近稀疏采样, 因此对变换中心的改变非常敏感, 在目标跟踪过程中会表现出对平移较为敏感. 在目标跟踪过程中, 通过对原始图像进行对数极坐标变换可以估计出目标的旋转和尺度参数, 但是存在的目标平移运动会使变换中心可能发生改变, 使得估计出的尺度和旋转参数并不准确. 先估计出目标的平移, 然后估计目标的旋转和尺度参数, 一定程度上能够减小变换中心对旋转和尺度参数估计的影响, 但是目标存在尺度变换和旋转运动时, 估计的平移参数也可能不是很准确, 不能保证尺度和旋转参数估计的准确性. 但是在频率域内, 两幅图像仅仅存在平移时, 傅里叶变换的幅值谱图像是完全相同的, 仅仅存在相位上的差异, 变换中心几乎是固定不变的, 这就避免了平移对目标旋转和尺度参数估计的影响. 傅里叶-梅林变换将原始图像上的尺度和旋转估计转换到频率域内进行, 与 DSST 等依靠有限的尺度金字塔估计尺度信息不同, 该方法能够实现连续空间的尺度和旋转参数估计.

假设存在图像 $s(x, y)$, 进行旋转、平移和尺度缩放后得到图像为 $t(x, y)$, 变换中心为原点 $(0, 0)$, 其中旋转变化为 $\Delta\theta$, 尺度变化为 α , 平移向量为 $(\Delta x, \Delta y)$, 变换后的图像 $t(x, y)$ 可以表示为

$$t(x, y) = s(\alpha(x \cos \Delta\theta + y \sin \Delta\theta) - \Delta x, \alpha(x \sin \Delta\theta - y \cos \Delta\theta) - \Delta y) \quad (1)$$

将 $t(x, y)$ 和 $s(x, y)$ 进行傅里叶变换, 得到傅里叶谱 $T(u, v)$ 和 $S(u, v)$, 两者之间存在如下关系:

$$T(u, v) = \frac{e^{-j2\pi(u\Delta x + v\Delta y)}}{\alpha^2} S\left(\frac{u \cos \Delta\theta + v \sin \Delta\theta}{\alpha}, \frac{u \sin \Delta\theta - v \cos \Delta\theta}{\alpha}\right) \quad (2)$$

忽略 $\left|\frac{e^{-j2\pi(u\Delta x + v\Delta y)}}{\alpha^2}\right|$, 则两者之间的幅值谱 $|T(u, v)|$ 和 $|S(u, v)|$ 仅仅存在旋转和尺度变化的关系, 对幅值谱图像取对数, 然后进行极坐标变换, 得

$$M_t(\log(\rho_t), \theta_t) = M_s(\log(\rho_t) - \log(\alpha), \theta_t - \Delta\theta) \quad (3)$$

对数 \log 的底与幅值谱的长宽和变换后对数极坐标图像的距离轴的宽度有关. 假设幅值谱图像的长为 H , 宽度为 W , 变换后距离轴的宽度为 d , 则文中对数 \log 的底定义为 $10^{\frac{1}{d} \log_{10}(\frac{1}{2} \sqrt{H^2 + W^2})}$. 本文实验部分, 令 $d = \frac{1}{2} \sqrt{H^2 + W^2}$. M_t 和 M_s 分别为 $\log(|T(u, v)|)$ 和 $\log(|S(u, v)|)$ 的对数极坐标图像, $\rho_t = \sqrt{u^2 + v^2}$, $\theta_t = \arctan(v/u)$. 用 (ε, η) 表示 $(\log(\rho_t), \theta_t)$, 则式 (3) 可以表示为

$$M_t(\varepsilon, \eta) = M_s(\varepsilon - \Delta\varepsilon, \eta - \Delta\eta) \quad (4)$$

最终图像之间的旋转运动和尺度缩放转换成了幅值谱对数极坐标图像距离轴和角度轴上的平移运动: $\Delta\varepsilon = \log(\alpha)$, $\Delta\eta = \Delta\theta$.

傅里叶-梅林变换的具体过程如下: 1) 对输入图像进行傅里叶变换, 得到对数幅值谱图像; 2) 对对数幅值谱图像进行高通滤波; 3) 对数极坐标变换, 得到图像傅里叶-梅林变换之后的特征. 其中对对数幅值谱图像进行高通滤波, 是因为通过高通滤波能够突出图像的边缘信息和轮廓特征, 这些信息对尺度和旋转参数的估计更加重要, 最终使得估计的结果更加准确. 高通增强滤波器定义如下:

$$H(u, v) = (1.0 - X(u, v)) \times (2.0 - X(u, v)) \quad (5)$$

其中, $X(u, v) = \cos(\pi(u/H - 0.5)) \cos(\pi(v/W - 0.5))$, W 和 H 为待滤波图像的宽度和高度.

1.2 核相关滤波

核相关滤波实质是一种目标检测方法, 该方法将核函数脊回归与循环矩阵结合起来, 提高了训练和检测的效率, 增强了算法的实时性能. 基于核相关滤波的目标跟踪是一种基于检测的目标跟踪方法, 它在样本空间利用循环矩阵稠密采样, 越靠近目标区域为正样本的可能性越大, 该方法在目标跟踪中已经取得了较好的跟踪效果.

线性空间的脊回归和循环矩阵也能够结合起来训练相关滤波检测子, 但是在实际应用中, 并不能保证获得的样本是线性可分, 利用线性脊回归处理非线性样本效果并不是非常理想. 通过核策略将非线性样本映射到高维核空间后, 样本在映射后的空间一般是线性可分的, 因此使用核技巧使得学习的检测子更加鲁棒. 假设核函数为 $\phi(\mathbf{x})$, 核回归模型为 $f(\mathbf{z}) = \sum_{i=1}^n w_i \phi(\mathbf{x}_i, \mathbf{z})$, $\mathbf{x}_i, i = 1, 2, \dots, n$ 为训练样本, 对核脊回归求解最终得到^[16]:

$$\mathbf{w} = (K + \alpha I)^{-1} \mathbf{y} \quad (6)$$

其中, $K_{ij} = \phi(\mathbf{x}_i, \mathbf{x}_j)$. 为了保证 K 为循环矩阵, 必须满足下列两个条件: 1) 训练的样本集 $\mathbf{x}_i, i = 1, 2, \dots, n$ 应该具有循环结构, 即 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ 为循环矩阵; 2) 核函数必须满足条件: 假设 \mathbf{x} 和 \mathbf{x}' 满足条件 1) 时, 对于任意的排列矩阵 M , 有 $\phi(\mathbf{x}, \mathbf{x}') = \phi(M\mathbf{x}, M\mathbf{x}')$, \mathbf{x} 和 \mathbf{x}' 并不要求来自于同一循环结构样本集, M 是每一行每一列都恰有一个元素为 1, 其他均为 0 的方阵. 则求解式 (6), 得

$$\hat{\mathbf{w}} = \left[\frac{1}{\hat{\mathbf{k}}^{xx} + \alpha\delta} \right]^* \odot \hat{\mathbf{y}} = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}}^{xx} + \alpha\delta} \quad (7)$$

其中, 符号 \wedge 表示傅里叶变换后的结果, \mathbf{k}^{xx} 为循环矩阵 K 的第一行向量, 该向量是对称的, 它的傅里叶变换结果 $\hat{\mathbf{k}}^{xx}$ 为实数, $\hat{\mathbf{k}}^{xx}$ 的共轭为其本身.

在目标跟踪中, 训练样本由目标区域图像块 (或者图像块提取的特征) $\mathbf{x} = \mathbf{x}_1$ 和它循环移位的候选样本 $\mathbf{x}_i, i = 2, \dots, n$ 组成, 每个样本的对应标签根据移位的距离确定, 越靠近目标区域为正样本的几率越大. 在检测时, 待检测的样本也是由一个图像块 (或者图像块提取的特征) $\mathbf{z} = \mathbf{z}_1$ 和它循环移位的样本 $\mathbf{z}_i, i = 2, \dots, n$ 构成, 通过回归模型处理待检测样本, 得到

$$\mathbf{f}(\mathbf{z}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \phi(\mathbf{x}_i, \mathbf{z}_j) = (K^z)^T \mathbf{w} \quad (8)$$

其中, $\mathbf{f}(\mathbf{z})$ 为所有待检测样本的检测输出 (属于正样本的置信度) 组成的向量, $K^z_{ij} = \phi(\mathbf{x}_i, \mathbf{z}_j)$. $(K^z)^T$ 是一个循环矩阵, 对式 (8) 两边进行傅里叶变换, 得

$$\hat{\mathbf{f}}(\mathbf{z}) = \hat{\mathbf{k}}^{xz} \odot \hat{\mathbf{w}} \quad (9)$$

其中, $\hat{\mathbf{k}}^{xz}$ 为 K^z 的第一行向量进行傅里叶变换的结果.

在基于核相关滤波的学习和检测过程中, 仅仅需要通过脊回归模型计算循环矩阵第一行向量 \mathbf{k}^{xx} 和 \mathbf{k}^{xz} 即可, 然后进行傅里叶变换, 即可代入式 (7) 和式 (9) 进行求解. 在实际应用中, 高斯核函数

$\phi(\mathbf{a}, \mathbf{b}) = \exp(-\frac{1}{\sigma^2} \|\mathbf{a} - \mathbf{b}\|^2)$ 比较常用, \mathbf{k}^{xx} 和 \mathbf{k}^{xz} 的求解过程与 \mathbf{k}^{ab} 相同^[16].

$$\mathbf{k}^{ab} = \exp \left(-\frac{1}{\sigma^2} \left(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\mathcal{F}^{-1}(\hat{\mathbf{a}}^* \odot \hat{\mathbf{b}}) \right) \right) \quad (10)$$

1.3 尺度和旋转参数估计算法

基于相关滤波的目标跟踪算法一般利用图像块 $\mathbf{x}_{0,0}$ 训练的滤波器 $\hat{\mathbf{w}}$ 对视觉外观进行建模, $\mathbf{x}_{0,0}$ 的长宽分别为 M 和 N , 则通过循环移位一共可以得到 $M \times N$ 个样本 $\mathbf{x}_{m,n}, (m, n) \in \{0, 1, \dots, M-1\} \times \{0, 1, \dots, N-1\}$, 每个样本通过高斯函数分配一个标签 $y(m, n)$. 核函数脊回归模型为 $f(\mathbf{z}) = \sum_{m,n} w_{m,n} \phi(\mathbf{x}_{m,n}, \mathbf{z})$, 其中 ϕ 为核函数, 损失函数定义如下:

$$L = \min \left\{ \sum_{m,n} \|\phi(\mathbf{x}_{m,n}) \times \mathbf{w} - y(m, n)\|^2 + \alpha \|\mathbf{w}\|^2 \right\} \quad (11)$$

在目标尺度和旋转参数估计过程中, $\mathbf{x}_{0,0}$ 是图像块经过傅里叶-梅林变换之后的对数极坐标图像, 式 (11) 中的训练样本集 $\mathbf{x}_{m,n}$ 是由 $\mathbf{x}_{0,0}$ 循环移位产生. 基于傅里叶-梅林变换的核相关滤波模型学习如图 1 所示. 为了消除傅里叶变化过程中的窗口效应, 需要在对数极坐标图像上叠加具有空间权重的汉宁窗.

在估计旋转和尺度参数时, 首先得到目标区域的图像块, 进行傅里叶-梅林变换后得到对数极坐标图像 \mathbf{z} , 与样本集 $\mathbf{x}_{m,n}$ 产生过程类似, 对 \mathbf{z} 循环移位得到待检测的样本集. 快速检测方法参考式 (9), 对其进行傅里叶反变换得到位移的空间置信度.

$$\tilde{\mathbf{y}} = \mathcal{F}^{-1}(\hat{\mathbf{f}}(\mathbf{z})) = \mathcal{F}^{-1}(\hat{\mathbf{k}}^{xz} \odot \hat{\mathbf{w}}) \quad (12)$$

通过 $\tilde{\mathbf{y}}$ 的最大响应位置可以求得距离轴的平移 $\Delta\epsilon$ 和角度轴的平移 $\Delta\eta$, 最终计算出尺度 α 和旋转角

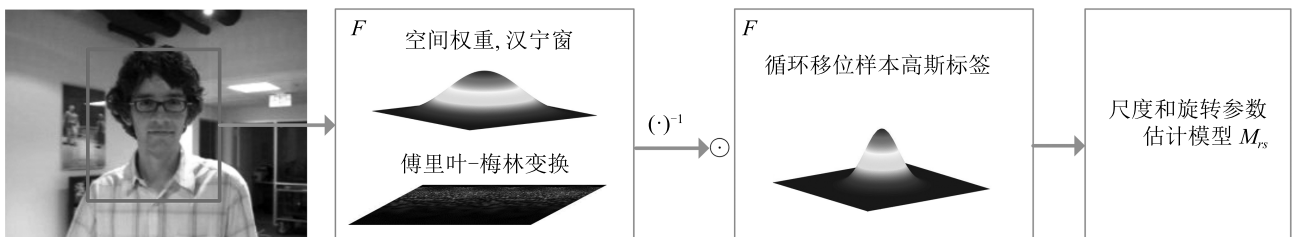


图 1 基于傅里叶-梅林变换的核相关滤波模型学习

Fig. 1 The kernelized correlation filtering model learning based on the Fourier-Mellin transform

度 $\Delta\theta$.

在目标跟踪中, 目标的视觉外观可能会发生变化, 模型应该具有适应目标的视觉外观变化的能力. 如果第 $t-1$ 帧训练的模型记为 $M_{rs}(t-1)$, 模型中包含两部分: 训练的样本 $\mathbf{x}(t-1)$ 和核脊回归的模型参数 $\hat{\mathbf{w}}(t-1)$. 在第 t 帧, 跟踪算法最终会得到目标新的位置、尺度和旋转参数, 根据这些参数剪切图像块, 然后进行傅里叶-梅林变换得到新的样本 $\mathbf{x}(t)$ 和核脊回归的模型参数 $\hat{\mathbf{w}}(t)$, 假设学习率为 β , 模型更新过程如下:

$$\begin{aligned} \mathbf{x} &= \beta\mathbf{x}(t-1) + (1-\beta)\mathbf{x}(t) \\ \hat{\mathbf{w}} &= \beta\hat{\mathbf{w}}(t-1) + (1-\beta)\hat{\mathbf{w}}(t) \end{aligned} \quad (13)$$

2 基于核相关滤波的目标位移和跟踪结果置信度估计

在目标跟踪过程中, 可能会遇到目标被遮挡、外部光照改变、目标快速运动和背景杂乱等情况, 跟踪算法的最重要任务就是在这些挑战性的情形下鲁棒地估计目标的位置. 在目标跟踪过程中, 准确判断目标当前的跟踪状态是一件非常有意义的事情, 能够给更上层的决策例如风险评估等提供支持. 一般的跟踪算法在估计目标位置的同时, 也会给出一个结果置信度值^[16], 但是在跟踪过程中, 目标模型的更新经常会存在缓慢漂移等现象, 如此会导致跟踪算法给出的置信度值始终比较高, 该值并不能够用于准确判断目标跟踪状态. 本文提出训练两个核相关滤波器分别用于估计目标在图像序列中的平移和目标跟踪结果的置信度.

用于目标位移估计和跟踪结果置信度估计的核相关滤波模型过程如图 2 所示. 图 2 中的两个核相

关滤波器, 一个用于估计目标在图像序列中的位移, 一个用于估计跟踪结果的置信度. 两个滤波器的学习、检测和模型更新过程与第 1 节介绍的目标尺度和旋转参数估计过程一样, 都是利用核相关脊回归方法来实现快速的模型学习和目标检测, 具体的过程参见图 1 和式 (11)~(13). 从图 2 可以看出, 两个核相关滤波器都使用了 HOG 特征, 这是因为 HOG 特征对外部光照变化和局部扭曲具有较好的鲁棒性, 而且在目标跟踪过程中已经取得了较好的跟踪效果^[16]. 在目标位移估计过程中, 在目标和它的周围区域内的图像块上提取 HOG 视觉特征作为训练的样本, 并且在提取的 HOG 特征上叠加具有空间权重的汉宁窗. 在目标跟踪结果置信度估计时, 应该减少周围环境对目标跟踪置信度估计的影响, 而且跟踪结果的所有像素点应该是同等重要的, 因此本文在训练置信度估计的核相关滤波器时仅仅考虑了跟踪算法定位的目标区域, 并没有叠加具有空间权重的汉宁窗, 如此可使得估计的置信度结果更加准确.

3 基于直方图和方差加权的跟踪目标搜索

直方图对目标幅度不大的平移和缩放等几何变化具有较好的鲁棒性, 对观察轴为轴心的旋转变换具有较好的不变性. 颜色直方图是一种比较简单的全局视觉特征, 计算非常高效. 对于图像 I , 它的颜色直方图特征是一个 n 维的向量 $\mathbf{H}_I = [h_1, h_2, \dots, h_n]$, 向量中的第 j 个元素 h_j 表示图像 I 落在组件 j 确定的颜色区间像素点的个数. 在目标跟踪过程中, 一般需要计算特定区域的直方图 (如目标区域), 图像 I 中特定区域 Ω 的直方图表示为 H_Ω^I . 区域 Ω 内像素点的数目为 $M \times N$, 直方图的组件数为 n , 对

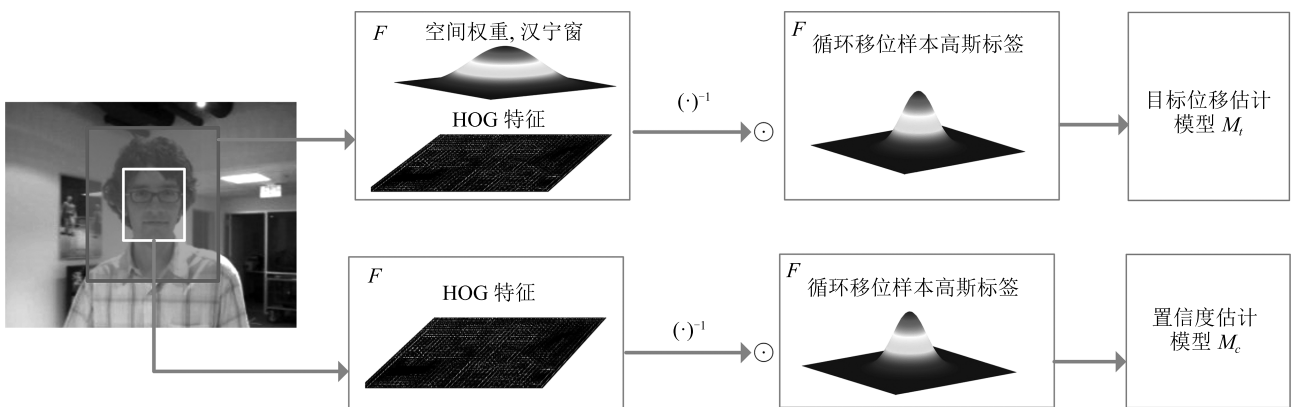


图 2 用于目标位移估计和跟踪结果置信度估计的核相关滤波模型

Fig. 2 The kernelized correlation filtering models for the estimation of object translation and the confidence of the tracking result

直方图进行归一化实质上反映了图像颜色空间的一种概率分布, 每个组件对应的概率为

$$P(b) = \frac{H_{\Omega}^I(b)}{M \times N}, \quad b = 1, 2, \dots, n \quad (14)$$

在目标跟踪应用中, Ω 表示的是跟踪目标在图像中的矩形区域. 本文根据 $P(b)$ 构建概率模型 M_h 对目标进行描述, 并且将其用于目标丢失时的再搜索. M_h 包含两部分: 目标区域的归一化直方图 $H_p = [P(b_1), P(b_1), \dots, P(b_n)]$ 和目标区域归一化概率和 S_{Ω} .

$$S_{\Omega} = \frac{\sum_{i=1}^{M \times N} P(b_{x_i})(\Omega)}{M \times N} \quad (15)$$

其中, x_i 为 Ω 区域内的像素点, 它最终分配到的 H_p 中的组件表示为 b_{x_i} .

在跟踪目标丢失时, 需要根据构建的模型 M_h 在图像 I_s 快速找到目标. 对于图像 I_s 上的任意像素点 x , 最终分配到的直方图组件表示为 b_x , 则 x 像素点属于目标的概率表示为 $P(x \in I_s) = P(b_x)$, 最终得到了与目标相关的似然图像 L , 其中 L 上的每个像素点 x 表示其对应的原始图像 $I_s(x)$ 属于目标的概率为 $P(b_x)$. 假设前一帧的目标位置矩形框记录为 w_t , 目标框的长宽分别 W_t, H_t , 使用大小和 w_t 相同的矩形框在当前帧图像 I_s 滑动可以得到大量的滑动窗口 $w_i, i = 1, 2, \dots, n$. 统计每个滑动窗口区域归一化的概率和 $S(w_i)$ 为

$$S(w_i) = \frac{\sum_{i=1}^{W_t \times H_t} P(b_{x_i})(w_i)}{W_t \times H_t}, \quad i = 1, 2, \dots, n \quad (16)$$

可以利用 L 的积分图像快速计算 $\sum_{i=1}^{W_t \times H_t} P(b_{x_i})(w_i)$, 最终得到每个滑动窗口是否包含目标的权重为

$$W_h(w_i) = \frac{|S(w_i) - S_{\Omega}|}{S_{\Omega}}, \quad i = 1, 2, \dots, n \quad (17)$$

利用直方图对滑动窗口进行加权, 主要考虑了目标区域颜色信息概率分布情况, 并没有反映出颜色信息的变化幅度, 即在均值上下的波动的剧烈程度, 因此对原始图像的像素分布情况考虑并不充分. 方差特征能够反映目标区域颜色数据的波动大小, 对于直方图特征而言是一种非常有效的补充. 本文根据目标区域的方差值 V_{Ω} 构建方差模型 M_v . 利用积分图像能够快速求解出每个滑动窗口对应图像块的方差值 $V(w_i)$ ^[14]. 通过方差求得每个滑动窗口是否包含目标的权重为

$$W_v(w_i) = \frac{|V(w_i) - V_{\Omega}|}{V_{\Omega}}, \quad i = 1, 2, \dots, n \quad (18)$$

将 $W_h(w_i)$ 和 $W_v(w_i)$ 线性加权最终得到每个滑动窗口包含目标的权重为

$$W(w_i) = \frac{1}{\alpha W_h(w_i) + (1 - \alpha) W_v(w_i)}, \quad i = 1, 2, \dots, n \quad (19)$$

对所有滑动窗口的权重进行归一化, 得

$$W(w_i) := \frac{W(w_i)}{\sum_{i=1}^n W(w_i)}, \quad i = 1, 2, \dots, n \quad (20)$$

其中, $W(w_i)$ 为 i 窗口被抽样的概率 $P(w_i)$, 故滑动窗口被抽样的概率分布已知. 根据蒙特卡罗离散型分布抽样原理, 可以从滑动窗口中随机抽取部分候选窗口进行处理, 能够提高跟踪目标的搜索效率.

在目标跟踪失败时, 短时间内目标很可能仍然在丢失位置附近, 因此并不需要在整帧图像上搜索目标, 通过限定搜索范围能够提高采样到目标区域的概率. 本文在跟踪失败时, 目标位置剪切 L 倍目标区域大小的图像块作为限定的搜索区域, 在剪切的图像块上计算权重, 采样待处理的候选样本. 在目标跟踪中, 模型需要实时更新以适应目标的视觉外观变化. 在基于直方图和方差的跟踪目标搜索过程中, 整个模型 M_{hv} 包含两部分: 直方图模型 M_h 和方差模型 M_v . 假设学习率为 γ , 模型更新过程如下:

$$\begin{cases} H_p = \gamma H_p(t-1) + (1 - \gamma) H_p(t) \\ S_{\Omega} = \gamma S_{\Omega}(t-1) + (1 - \gamma) S_{\Omega}(t) \\ V_{\Omega} = \gamma V_{\Omega}(t-1) + (1 - \gamma) V_{\Omega}(t) \end{cases} \quad (21)$$

4 具有旋转和尺度适应性的长时间目标跟踪算法

为了实现具有旋转和尺度适应性的长时间目标跟踪, 本文训练了三个核相关滤波器, 分别用于估计目标的尺度和旋转参数、目标在图像中位移以及跟踪结果的置信度. 同时本文提出一种基于直方图和方差加权的跟踪目标搜索方法, 能够快速选择置信度高的候选目标区域, 用于目标丢失时再检测. 本文提出的具有尺度和旋转适应性的长时间目标跟踪 (Robust long-term object tracking with adaptive scale and rotation estimation, RLOT) 算法框图如图 3 所示.

RLOT 算法的具体执行流程如下:

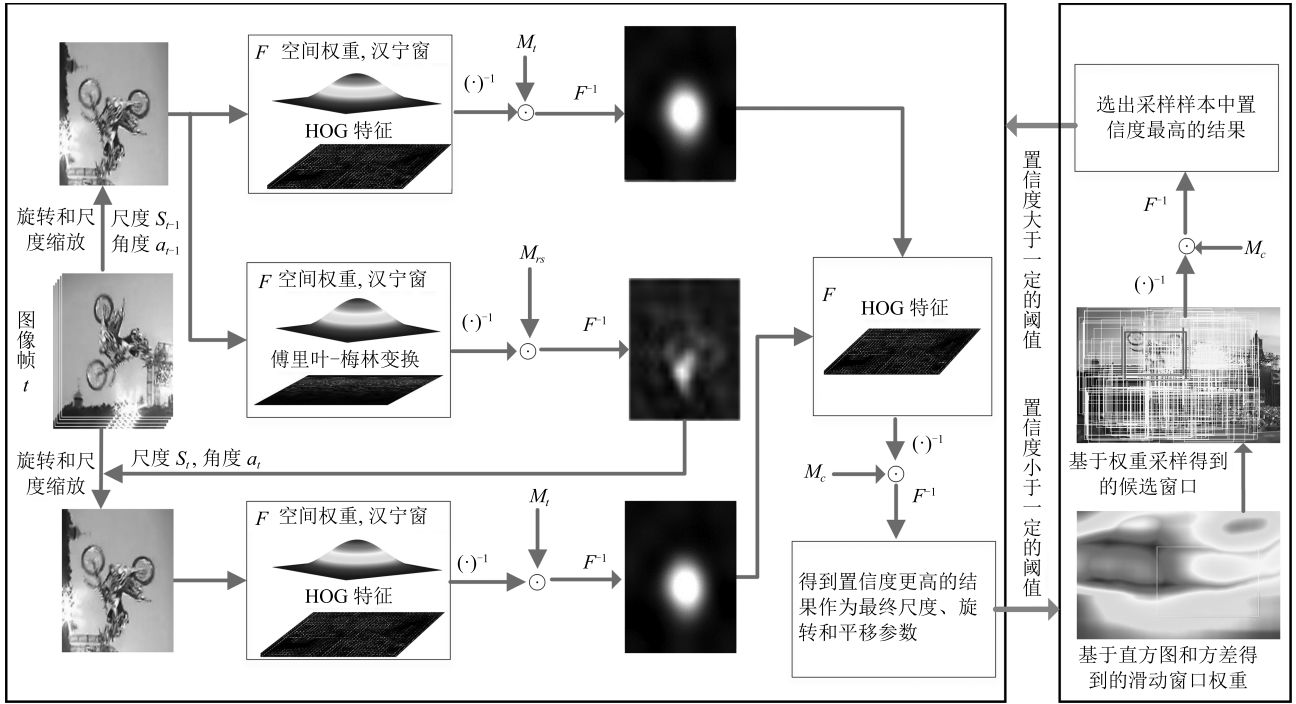


图3 具有尺度和旋转适应性的长时间目标跟踪算法框图

Fig. 3 The architecture of robust long-term object tracking with adaptive scale and rotation estimation

步骤 1. 初始化:

目标状态 $S_0 = (x_0, y_0, \alpha_0, s_0)$, 位移估计核相关滤波模型 M_t , 尺度和旋转估计核相关滤波模型 M_{rs} , 置信度估计核相关滤波模型 M_c , 直方图和方差模型 M_{hv} .

步骤 2. 对于第 t 帧图像 I_{th} , 根据 S_{t-1} 从图像中剪切块 p_i , 提取 HOG 特征, 通过 M_t 估计目标位置 (\hat{x}_t, \hat{y}_t) .

步骤 3. 在 p_i 上提取傅里叶-梅林变换特征, 通过 M_{rs} 估计旋转和尺度 $\hat{\alpha}_t, \hat{s}_t$.

步骤 4. 根据 $(x_{t-1}, y_{t-1}, \hat{\alpha}_t, \hat{s}_t)$ 在 I_{th} 上剪切新的图像块 p'_i , 提取 HOG 特征, 通过 M_t 估计目标位置 (\hat{x}'_t, \hat{y}'_t) .

步骤 5. 根据 $(\hat{x}'_t, \hat{y}'_t, \hat{\alpha}_t, \hat{s}_t)$ 和 $(\hat{x}_t, \hat{y}_t, \alpha_{t-1}, s_{t-1})$ 剪切两个目标区域图像块 p'_{ti} 和 p_{ti} , 提取 HOG 特征, 通过 M_c 估计两个图像块的置信度 c'_{ti} 和 c_{ti} .

步骤 6.

If $c'_{ti} > c_{ti}$

Then $S_t \leftarrow (\hat{x}'_t, \hat{y}'_t, \hat{\alpha}_t, \hat{s}_t)$

Else $S_t \leftarrow (\hat{x}_t, \hat{y}_t, \alpha_{t-1}, s_{t-1})$.

步骤 7.

If $\max(c'_{ti}, c_{ti}) < \mathcal{T}_r$

Then 执行步骤 7.1, 7.2, 7.3.

步骤 7.1. 根据 S_t 旋转 I_{th} , 然后利用 M_{hv} 加权采样得到候选的目标状态 X .

步骤 7.2.

For $x'_i \in X$

根据 M_c 计算每个 x'_i 候选区域的置信度, 得到 c'_i

Endfor.

步骤 7.3.

If $\max\{c'_i\} > \mathcal{T}_d$

Then $S_t \leftarrow x'_i$, 其中 $i = \arg \min_i(c'_i)$.

步骤 8. 更新模型 M_t, M_{rs} .**步骤 9.**

If $\max(c'_{ti}, c_{ti}) > \mathcal{T}_u$

Then 更新模型 M_c 和 M_{hv} .

步骤 10. 如果序列没有结束, 则转到步骤 2, 否则算法终止.

估计尺度和旋转参数的核相关滤波器的特征为幅值谱对数极坐标图像, 而估计目标置信度和目标位移的两个核相关滤波器采用的是 HOG 特征, 提取 HOG 特征的单元格大小定义为 4, 特征的维数为 31. 三个核相关滤波器都采用了高斯核函数. 利用式 (9) 和式 (12), 在频率域内计算样本之间的相关性, 运算效率有了较大提高. 在估计目标的位移和跟踪结果的置信度之前, 需要利用估计的旋转和尺度参数将目标及其周围区域进行旋转和缩放, 使得旋转和缩放后的目标区域与核相关滤波模型 M_t 有同样的旋转角和尺度, 可使得目标位移的估计更加准确. 在目标丢失后, 采用基于直方图和方差的跟踪目标搜索算法再次定位目标. 本文使用的彩色直方图在 Lab 颜色空间上计算得到, 而方差是在灰度图像上计算得到的. Lab 色彩空间是颜色-对立空间, 维度 L 表示亮度, a 和 b 表示颜色对立维度, Lab 颜色空间是描述人眼可见的所有颜色最完备的色彩模型, 在计算机视觉领域已经得到了广泛的应用. Lab 颜色空间每个通道离散化为 32 个组件, 则 Lab 颜色空间上得到的直方图组件数为 $32^3 = 32768$. 直

方图和方差的权重计算都是在积分图像上进行, 运算效率很高. RLOT 跟踪算法中, 求解得到的跟踪结果的置信度都不小于 0, 当 \mathcal{T}_r 参数设置为负值时, 步骤 7.1, 7.2, 7.3 不会运行, 相当于 RLOT 中的再检测模块被屏蔽, 此时无需更新模型 M_{hv} , 本文将没有再检测模块的 RLOT 算法称为 ROT.

5 目标跟踪实验

本文在目标跟踪标准数据集 OTB 上比较了提出的 RLOT、ROT 跟踪算法和目前主流的目标跟踪算法的性能^[6-7]. OTB 数据集包含 100 个测试序列, 这些序列包含 11 种不同类型的视觉跟踪挑战, 包括目标突然快速运动、凌乱背景、运动模糊、扭曲、光照变化、平面内旋转、平面外旋转、低分辨率、半遮挡、出相机视野和尺度变化. OTB 数据集上的所有图像序列都已经被人工标注, 标注的真值在图像上表现为包含有目标的矩形框. OTB 数据集提供了 29 个跟踪子的跟踪结果, 能够用于评估跟踪算法的性能. 本文使用 C++ 实现了提出的目标跟踪算法, 所有实验都在配备有 3.1 GHz i7-5557U CPU 和 8 GB RAM 的计算机上进行.

5.1 实验设置

OTB 数据集提供了精度和成功率两种度量准则来评估目标跟踪算法性能. 精度准则是基于中心的位置误差进行度量, 是指跟踪算法估计的目标中心和标注的目标区域中心之间的平均距离. 一个序列的所有帧的平均中心位置误差可表征该序列的总体跟踪精度性能. 精度图显示了对于给定的位置误差阈值, 跟踪算法估计的目标位置与真值之间距离落在阈值之内的帧数占整个测试数据集帧数的百分比. 成功率是利用重叠度进行度量, 实质是指跟踪算法估计的目标区域和标注的目标区域之间的重叠度. 跟踪算法得到的目标区域为 B_t , 标注的真实区域为 B_a , 则定义两者的重叠度为 $O = |B_t \cap B_a| / |B_t \cup B_a|$. 成功率图显示了对于给定的重叠度阈值, 跟踪算法估计的目标区域与真值区域之间的重叠度大于阈值的帧数占整个测试数据集帧数的百分比. 使用特定的阈值来判断跟踪算法的成功率并不公平, OTB 分别使用成功率图和精度图曲线下的面积 (Area under curve, AUC) 对跟踪算法进行排序^[6]. OTB 标注真值为标准矩形 (矩形的两边和图像的边界平行), 其重叠度和精度的计算方法针对的也是两个标准矩形. 本文通过估计目标的旋转参数得到了非标准的矩形, 根据 OTB 提供的方法需要得到非标准矩形的外接矩形的中心和区域来计算中心位置误差和重叠度. 目标标注真值、跟踪产生的非标准矩

形和非标准矩形的外接矩形示意图如图 4 所示. 从图 4 可以看出, 利用外接矩形计算重叠度时, 可能会降低重叠度的计算准确性, 外接矩形的中心和非标准矩形的中心比较接近, 该方法对中心位置误差的结果影响比较小. 为了提高估计结果的准确性, 本文直接求解非标准矩形和真值之间的中心位置误差和重叠度. 因为标注的真值也没有考虑目标的旋转情况, 标注的区域和实际目标区域还是存在一定的区别, 但是标注的区域中心和实际目标的中心会比较接近, 最终采用本文提出的估计方法在计算重叠度上还是存在着一定的偏差, 但是偏差较原始 OTB 方法已经有一定程度的减小, 在计算中心位置误差时则比较准确.

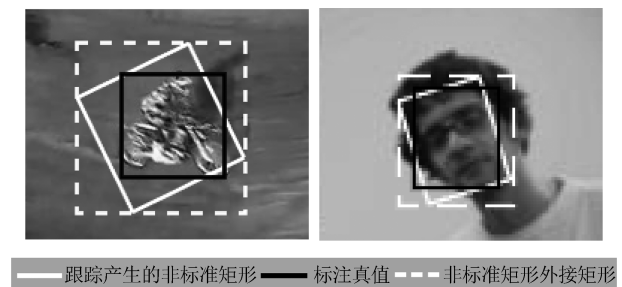


图 4 目标标注真值、跟踪产生的非标准矩形和非标准矩形的外接矩形示意图

Fig. 4 The diagram of target annotations, nonstandard rectangles from our trackers and external rectangles of nonstandard rectangles

本文设计了三个核相关滤波器, 相关的参数如表 1 所示, 其中对于高斯标签宽度参数 s , 位移估计和置信度估计核相关滤波器中 n 和 m 指的是目标区域宽和高方向上 HOG 单元格的数目, 尺度旋转核相关滤波器中 d 表示距离轴长度, a 表示角度轴长度, 两个轴上高斯宽度的参数不同. 对于目标位移估计相关滤波器和尺度、旋转估计相关滤波器, 需要将目标和它附近一定区域的背景剪切出来, 提取视觉特征, 训练相关滤波器、检测待跟踪的目标. 本文设置剪切的初始图像块大小为目标区域的 2.8 倍. 从 RLOT 算法流程可以看出, 有几个阈值参数需要设置. \mathcal{T}_r 设置为 0.25, 通过该阈值来判断是否需要使用基于直方图和方差加权的方法搜索目标; \mathcal{T}_d 设置为 0.4, 通过其判断搜索到的结果是否可信, 是否需要用该结果重新初始化目标; \mathcal{T}_u 设置为 0.4, 通过该阈值判断是否需要更新置信度估计核相关滤波模型 M_c , 直方图和方差模型 M_{hv} . 通过基于直方图和方差加权采样得到的候选滑动窗口数量设置为 150, 直方图权重和方差权重的比重是相等的, 即式 (19) 中的 α 设置为 0.5. 跟踪失败时, 限定搜索区域的参

数 L 设置为 4. 当运行 ROT 跟踪算法时, \mathcal{T}_r 设定为 -1 , 此时 M_{hv} 不更新.

表 1 三个核相关滤波器参数
Table 1 The parameters of three kernelized correlation filters

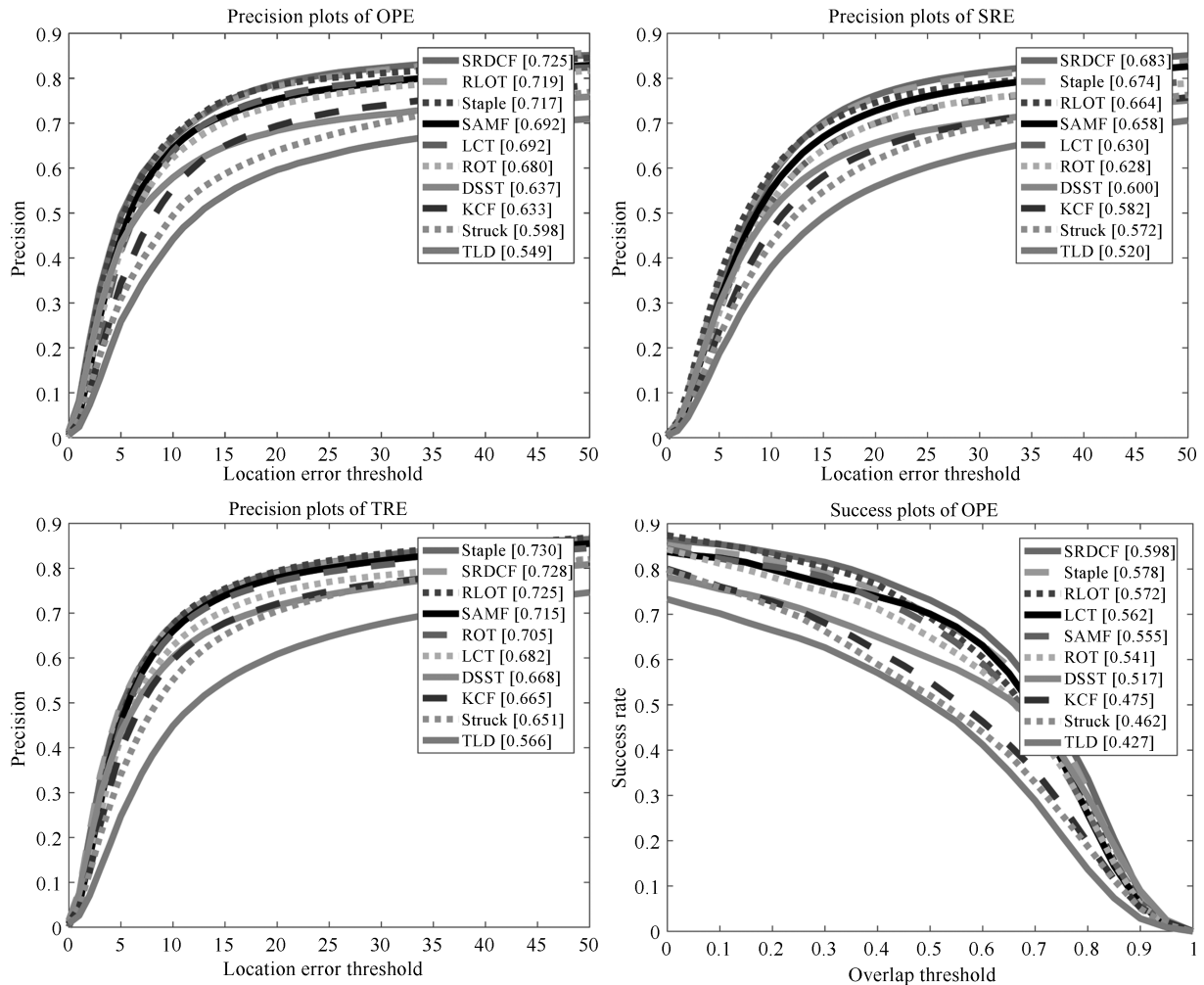
核相关滤波器	位移估计	置信度估计	尺度旋转估计
	核相关滤波器	核相关滤波器	核相关滤波器
高斯核宽度参数 σ	0.6	0.6	0.4
学习率 β	0.012	0.012	0.075
高斯标签宽度参数 s	$0.125\sqrt{mn}$	$0.125\sqrt{mn}$	$0.075(d, a)$
正则化参数 α	10^{-4}	10^{-4}	5^{-5}

5.2 定量分析

本文定量比较了提出的目标跟踪算法 RLOT, ROT 与目前主流的跟踪算法的性能. 其中 OTB 已经提供了 29 个跟踪子, 包括有 TLD^[14], Frag^[27], Struck^[28], CT^[13], SCM^[29], ASLA^[30], CXT^[31] 等的算法性能, 本文还比较了几个最新的跟踪算

法 KCF^[15], SAMF^[17], DSST^[18], SRDCF^[19], Staple^[20], LCT^[23]. 为了使得评估的效果更加公平有效, 更能反映出跟踪算法的鲁棒性能. 本文从传统的一次通过估算 (One-pass evaluation, OPE)、时间鲁棒性估算 (Temporal robustness evaluation, TRE) 和空间鲁棒性估算 (Spatial robustness evaluation, SRE) 三个方面给出了跟踪算法的精度图和成功率图^[6]. 其中 OPE 是指在整个序列第 1 帧用标注的真值进行初始化后, 得到跟踪算法的平均精度和成功率, 整个跟踪执行一次, 这是一种传统的跟踪算法性能评估方法. 然而, 跟踪子可能对初始人工给定的真值比较敏感, 并且在不同起始帧进行初始化时跟踪算法的性能也可能不同, 因此通过 TRE 和 SRE 在初始化时对跟踪子添加人工的干扰以分析不同初始条件下目标跟踪算法的鲁棒性.

在 OTB 数据集下, 本文提出的目标跟踪算法和主流跟踪算法的性能比较结果如图 5 所示, 为了使得结果清晰可见, 图 5 仅显示了前 10 个最好的跟踪结果. 从图 5 可以看出, SRDCF 和 Staple 跟踪算法的性能比较优越, 但是本文提出的 RLOT 算法性



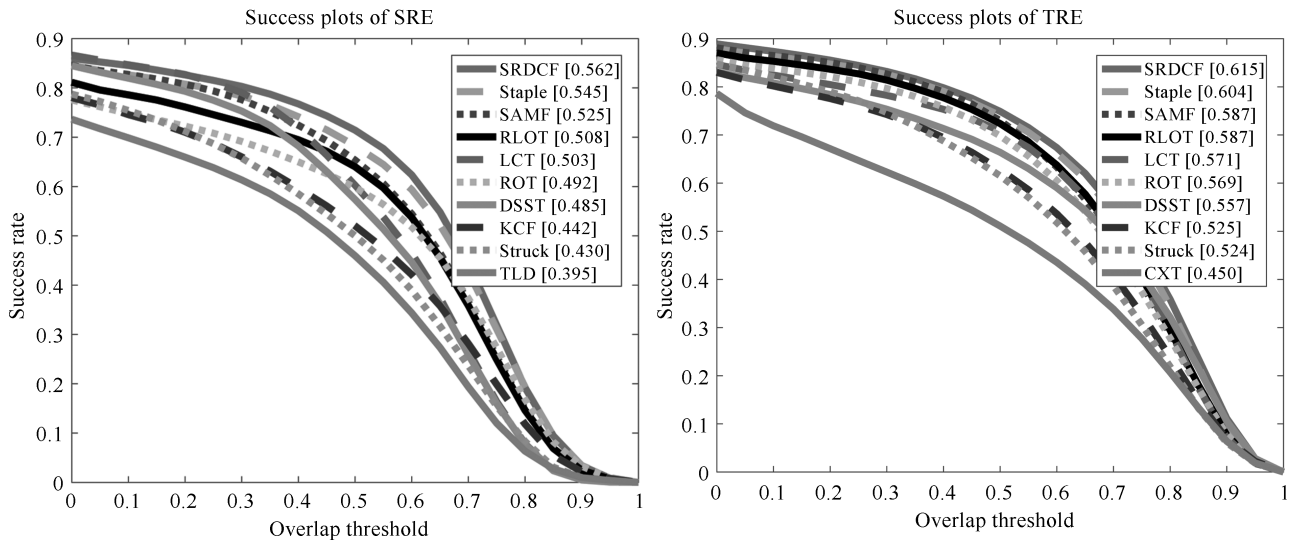


图5 通过 OPE, SRE 和 TRE 估算准则得到的跟踪算法精度图和成功率图

Fig. 5 Precision plots and success rate plots of tracking algorithms evaluated by OPE, SRE and TRE standards

能与两者比较接近,也取得了较好的跟踪效果.没有检测模块的 ROT 跟踪算法性能有一定的下降,与 LCT, SAMF 算法的性能比较接近,但是要优于 DSST, KCF 等跟踪算法.对于 OPE 评估准则, RLOT 算法的精度仅次于 SRDCF,仅相差 0.006 (0.8%),成功率与 Staple 非常相近,与 SRDCF 也仅相差 0.026 (4.3%). ROT 的精度性能和成功率性能略差于 SAMF 和 LCT.采用空间鲁棒性 SRE 评估时, RLOT 算法的精度与 SRDCF 和 Staple 相近,但是成功率性能略低于 SRDCF, Staple 和 SAMF,要高于 LCT. ROT 算法同样对空间干扰比较敏感.不准确的初始值会导致估计的旋转和尺度估计准确度下降,最终导致性能一定程度的降低.对于时间鲁棒性 TRE, RLOT 精度性能比较高,与 SRDCF 和 Staple 仅相差 0.005 (0.68%),成功率与 SAMF 算法相同,与 SRDCF 和 Staple 性能比较接近. ROT 算法性能与 LCT 比较接近. RLOT 较之 ROT, 算法参数完全相同,唯一区别在于 ROT 屏蔽了再检测模块,从两者的性能比较可以看出,目标跟踪过程中再检测模块是有意义的,能够提高跟踪子的性能.

OTB 数据集中包含 11 种不同类型的视觉跟踪挑战,每个序列可能遇到的视觉挑战都已经被人工标注,通过这些标注的序列可以分析跟踪算法针对不同挑战的性能.本文提出的 RLOT 目标跟踪算法主要有两个优点: 1) 目标的旋转和尺度估计; 2) 目标丢失时再检测.平面外旋转、平面内旋转以及尺度变换都已经被标注,通过标注的数据能够分析出算法旋转和尺度估计的有效性.在跟踪过程中,目标存在半遮挡或全遮挡时,跟踪经常会失败,通过标注的半遮挡序列能够评估再检测模块的性能.在平面

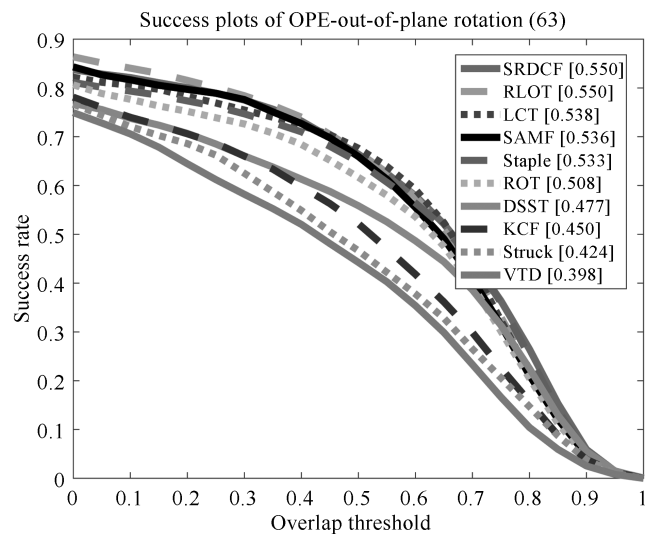
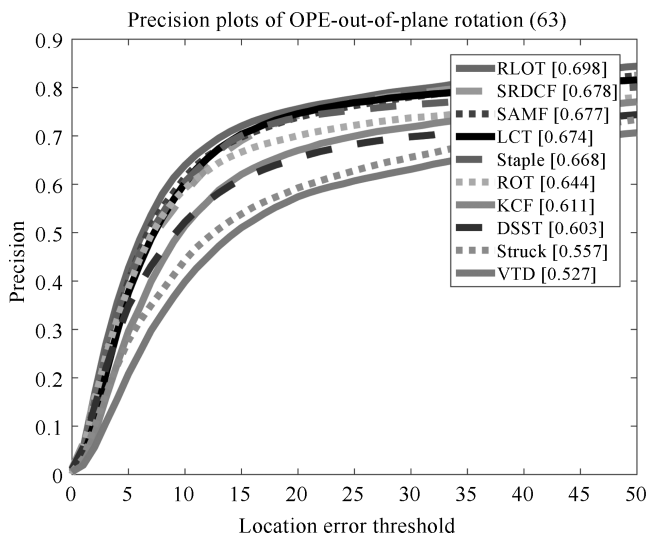
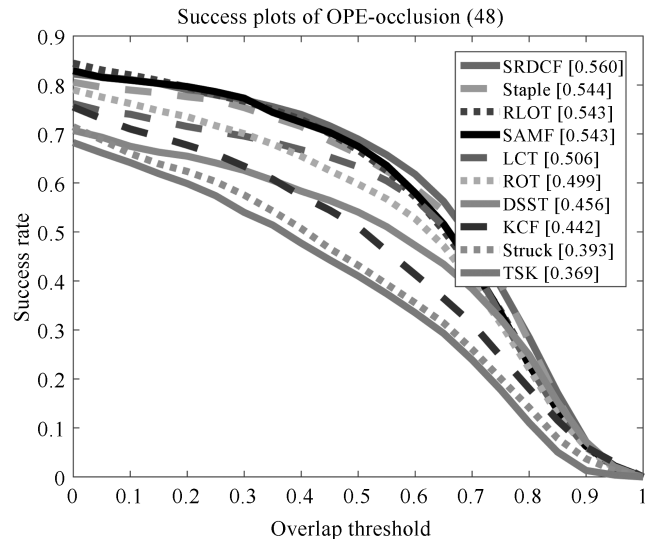
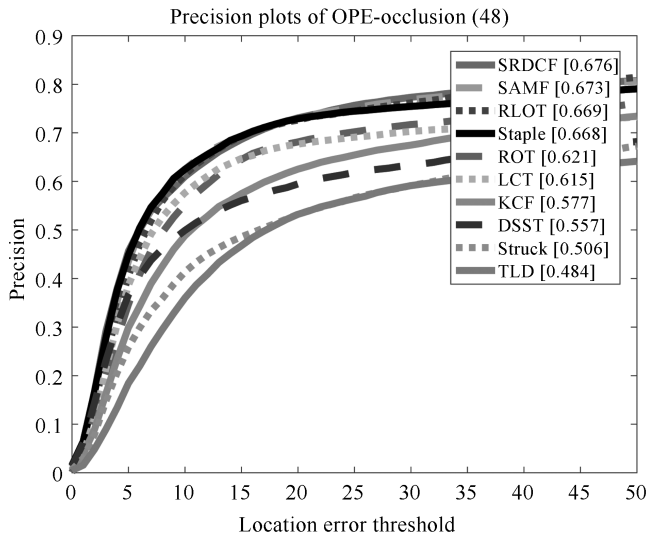
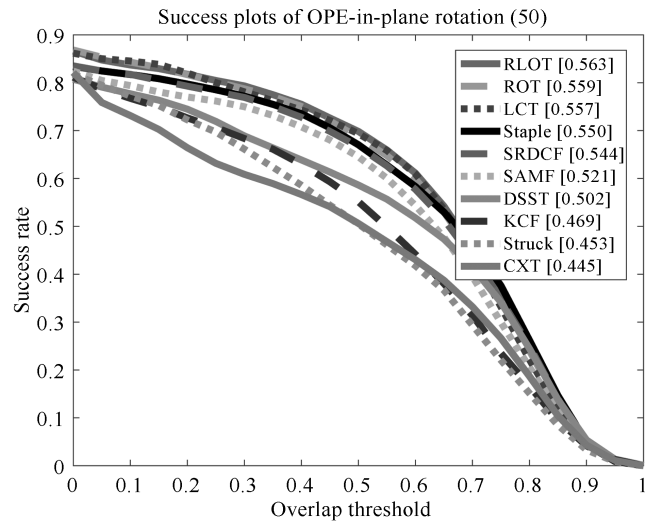
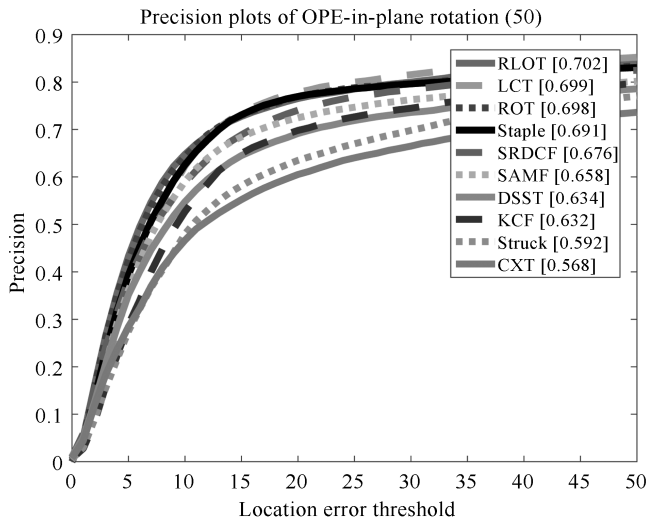
外旋转、平面内旋转、尺度变化、半遮挡视觉挑战情况下,通过 OPE 估算准则得到的跟踪算法精度和成功率如图 6 所示,为了使得结果清晰可见,图 6 仅显示了前 10 个最好的跟踪结果.从图 6 可以看出,在平面内旋转和平面外旋转时, RLOT 在精度和成功率上都取得了最好的性能. ROT 在平面内旋转时与 LCT 的性能非常相近,仅次于 RLOT 跟踪算法. ROT 在平面外旋转时,性能有一定程度的下降,但是较 KCF 和 DSST 也有较大提高.在目标存在尺度变化时, RLOT 算法取得了最好的精度性能,成功率性能仅次于 SRDCF.而 ROT 算法性能与 SAMF 非常接近.从上面分析可以看出,基于傅里叶-梅林变换和相关滤波方法的目标旋转和尺度估计是有效的,能够有效提高跟踪算法的性能.存在半遮挡时, RLOT 跟踪子和 SRDCF 跟踪子、SAMF 跟踪子和 Staple 跟踪子性能比较接近,要好于 ROT 跟踪子,验证了跟踪过程中再检测模块的有效性.

5.3 定性分析

本节定性分析 RLOT、ROT 跟踪算法与目前主流目标跟踪算法的性能,包括相关滤波跟踪算法 SRDCF,具有再检测模块的相关滤波跟踪算法 LCT,基于检测的目标跟踪方法 Struck 以及将跟踪、检测和学习结合的跟踪算法 TLD.同时本文从 100 个图像序列中选择 11 个具有代表性的序列进行着重分析,这些序列包含 OTB 中几乎所有的视觉挑战,如表 2 所示.表 2 中 0 表示该序列不包含该项挑战,1 表示序列包含该项挑战.跟踪算法 RLOT, ROT, SRDCF, LCT, Struck 以及 TLD 在这 11 个序列上的跟踪结果如图 7 所示.图 7 中,从上到下分

别为 David, CarScale, Dog1, FaceOcc2, Jogging-2, Lemming, MotorRolling, Shaking, Singer2, Tiger1 和 Soccer. SRDCF 跟踪子将 HOG 特征、颜色特征和灰度特征结合起来训练相关滤波器, 并且针对

循环移位的边界效应, 添加了空间正则项, 提高了跟踪子的性能. SRDCF 借鉴了 SAMF 的尺度估计方法, 在原始图像和该图像缩放后图像上检测目标, 最终得到最大响应值对应的尺度和位置. SRDCF 没有



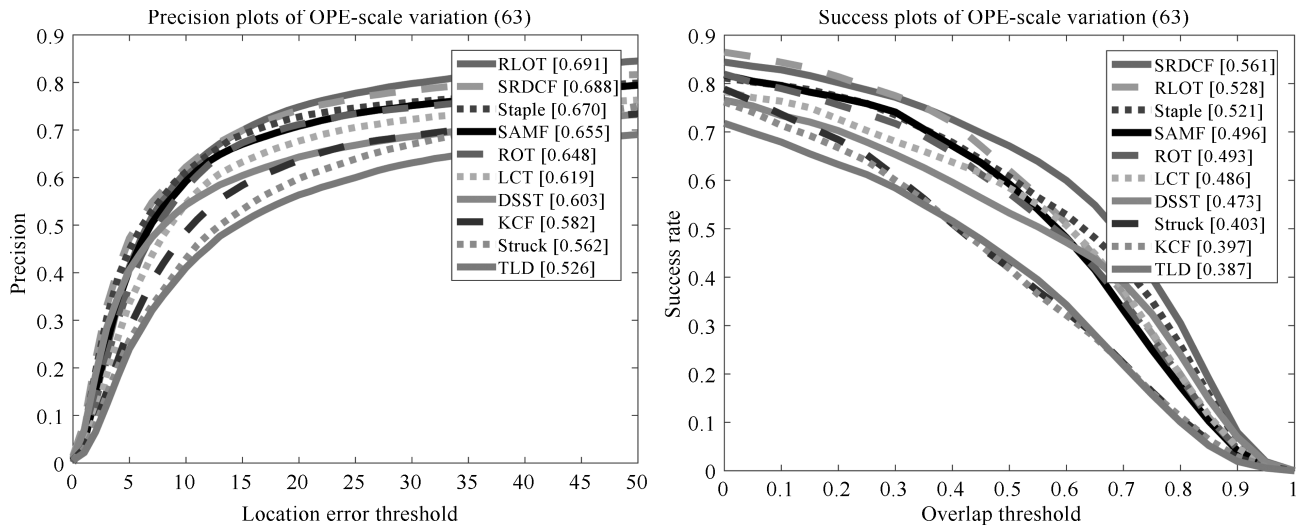


图 6 不同视觉挑战情况下通过 OPE 估算准则得到的跟踪算法精度图和成功率图
Fig. 6 Precision plots and success rate plots of tracking algorithms evaluated by OPE standard under different visual tracking challenges





图 7 RLOT, ROT, SRDCF^[19], LCT^[23], TLD^[14] 和 Struck^[28] 在 11 个 OTB 序列上的跟踪结果
 Fig. 7 Tracking results using RLOT, ROT, SRDCF^[19], LCT^[23], TLD^[14] and Struck^[28] on 11 OTB image sequences

表 2 OTB 数据集中选择的 11 个序列包含的视觉挑战

Table 2 The visual tracking challenges included in the 11 image sequences selected from the OTB datasets

图像序列	光照变化	平面外旋转	尺变变化	半遮挡	扭曲	运动模糊	快速运动	平面内旋转	出相机视野	凌乱背景	低分辨率
David	1	1	1	1	1	1	0	1	0	0	0
CarScale	0	1	1	1	0	0	1	1	0	0	0
Dog1	0	1	1	0	0	0	0	1	0	0	0
FaceOcc2	1	1	0	1	0	0	0	1	0	0	0
Jogging-2	0	1	0	1	1	0	0	0	0	0	0
Lemming	1	1	1	1	0	0	1	0	1	0	0
MotorRolling	1	0	1	0	0	1	1	1	0	1	1
Shaking	1	1	1	1	1	0	0	0	0	1	0
Singer2	1	1	0	0	1	0	0	1	0	1	0
Tiger1	1	1	0	1	1	1	1	1	0	0	0
Soccer	1	1	1	1	0	1	1	1	0	1	0

估计目标的旋转参数,除了具有快速旋转变换的 MotorRolling 序列外, SRDCF 在其他序列上都取得了较好的跟踪性能. LCT 跟踪子将直方图特征和 HOG 特征结合起来训练相关滤波器,并且还专门训练了一个随机森林分类器,用于目标丢失时再检测. LCT 并没有包含旋转估计模块,当目标具有旋转运动时,跟踪效果并不理想(如 MotorRolling). 当背景非常凌乱,整个目标与周围区域相似性又比较高时, LCT 检测模块也会发生漂移,导致跟踪失败(如 Soccer). 其他情况下, LCT 取得了较好的跟踪性能. Struck 对尺度变化的鲁棒性不好(如 CarScale, Dog1),当存在快速的平面内旋转时,跟踪会失败(如 MotorRolling),而且在目标存在大的平面外旋转时存在漂移(如 David),对凌乱背景的序列的跟踪效果也不是太理想(如 Singer2, Soccer). Struck 在其他条件下的跟踪性能都比较好,特别是对重度遮挡的目标具有较好的鲁棒性. TLD 算法将跟踪、检测和学习结合起来,当跟踪丢失时能够再次检测目标,而且检测子对样本的选择非常严格,使得检测子的漂移较跟踪子更慢. TLD 的检测模块在凌乱背景,光照较暗的情况下性能并不好(如 Singer2, Shaking, Soccer), TLD 整个检测框架并没有考虑目标大的旋转运动(如 MotorRolling). TLD 虽然具有从错误跟踪中恢复过来的能力,但是其跟踪模块采用简单的光流法跟踪目标,整体性能比较差,而检测子依赖跟踪子提供的目标运动线索更新检测模块,如此导致 TLD 检测子性能也受到较大影响,最后虽然能再次定位目标,但是中间可能存在比较多的图像帧跟踪失败的情况(如 Lemming, Tiger1). RLOT 和 ROT 算法利用傅里叶-梅林变换和核相关滤波器估计目标的尺度和旋转参数,在目标存在旋转和尺度变化时跟踪效果比较好(如 CarScale, Dog1, FaceOcc2, MotorRolling, Singer2). 但是 ROT 没有再检测模块,当目标存在严重的半遮挡时, ROT 跟踪失败后无法自行恢复,而 RLOT 能够再次检测到目标(如 Jogging-2, Lemming, Shaking, Tiger1). 与 LCT 类似,背景非常凌乱,整个目标与周围区域相似性又比较高时, RLOT 检测模块也会发生漂移(Soccer). 通过定性分析可以看出,当目标存在尺度变换和旋转运动时(特别是快速旋转运动),本文提出的 RLOT、ROT 算法的优势比较明显,定性分析结果也进一步验证了跟踪过程中再检测模块的有效性.

5.4 算法实时性分析

不同目标跟踪算法的平均处理帧速率如表 3 所示,包括 RLOT, SRDCF, LCT, KCF, DSST, TLD

和 Struck. 从表 3 可以看出, KCF 跟踪子的平均处理帧速率最高,其次是本文提出的 RLOT 跟踪算法, SRDCF, LCT, DSST, TLD 和 Struck 跟踪子处理帧速率都没超过 30 fps. SRDCF 帧速率最低,仅为 4 fps. RLOT 在 KCF 基础上整合了旋转和尺度估计、跟踪结果置信度估计和基于直方图和方差加权的目标搜索等模块,导致计算量增加,一定程度上降低了处理帧速率,但是其平均处理帧速率能达到 36 fps,仍然满足目标跟踪的实时性要求.

表 3 不同跟踪算法的平均处理帧速率

Table 3 The average frame rates of different object tracking algorithms

跟踪算法	帧速率
RLOT	36
SRDCF	4
LCT	27.4
KCF	167
DSST	27
TLD	20
Struck	28

6 结论

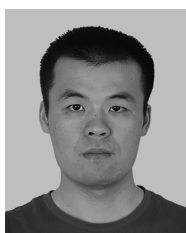
本文提出了一种具有尺度和旋转适应性的长时间目标跟踪方法(RLOT). 该方法可以分为三个部分: 1) 基于傅里叶-梅林变换和核相关滤波的尺度和旋转参数估计; 2) 基于核相关滤波的目标平移和跟踪结果置信度估计; 3) 基于直方图和方差加权的目标搜索. 本文提出的方法能够准确估计出目标的尺度、旋转和平移参数,并且能够对跟踪结果的置信度进行有效评估,以判断跟踪结果是否正确,根据跟踪结果的置信度选择是否激活基于直方图和方差加权的目标搜索模块,最终使得跟踪算法具有从失败中恢复过来的能力. 本文的尺度和旋转参数估计模块、跟踪结果置信度估计模块和再检测模块能够用于其他跟踪算法,提高跟踪算法的性能. 本文在 OTB 数据集上开展了实验研究,并与当前主流的目标跟踪算法进行了比较,实验结果验证了本文提出算法的有效性和优越性.

References

- 1 Yilmaz A, Javed O, Shah M. Object tracking: a survey. *ACM Computing Surveys (CSUR)*, 2006, **38**(4): Article No. 13.1-13.45
- 2 Hu W M, Tan T N, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Transac-*

- tions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2004, **34**(3): 334–352
- 3 Li X, Hu W M, Shen C H, Zhang Z F, Dick A, van den Hengel A V D. A survey of appearance models in visual object tracking. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2013, **4**(4): Article No. 58.1–58.48
 - 4 Yin Hong-Peng, Chen Bo, Chai Yi, Liu Zhao-Dong. Vision-based object detection and tracking: a review. *Acta Automatica Sinica*, 2016, **42**(10): 1466–1489
(尹宏鹏, 陈波, 柴毅, 刘兆栋. 基于视觉的目标检测与跟踪综述. 自动化学报, 2016, **42**(10): 1466–1489)
 - 5 Guan Hao, Xue Xiang-Yang, An Zhi-Yong. Advances on application of deep learning for video object tracking. *Acta Automatica Sinica*, 2016, **42**(6): 834–847
(管皓, 薛向阳, 安志勇. 深度学习在视频目标跟踪中的应用进展与展望. 自动化学报, 2016, **42**(6): 834–847)
 - 6 Wu Y, Lim J, Yang M H. Online object tracking: a benchmark. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA: IEEE, 2013. 2411–2418
 - 7 Wu Y, Lim J, Yang M H. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1834–1848
 - 8 Kristan M, Matas J, Leonardis A, Felsberg M, Cehovin L, Fernandez G, Vojir T, Hager G, Nebehay G, Pflugfelder R, Gupta A, Bibi A, Lukezic A, Garcia-Martin A, Saffari A, Petrosino A, Montero A S et al. The visual object tracking VOT2015 challenge results. In: Proceedings of the 2015 IEEE International Conference on Computer Vision Workshops (ICCVW). Santiago, Chile: IEEE, 2015. 564–586
 - 9 Wang H Z, Suter D, Schindler K, Shen C H. Adaptive object tracking based on an effective appearance filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(9): 1661–1667
 - 10 Black M J, Jepson A D. Eigen tracking: robust matching and tracking of articulated objects using a view-based representation. In: Proceedings of the 4th European Conference on Computer Vision. Cambridge, UK: Springer Berlin Heidelberg, 1996. 329–342
 - 11 Chin T J, Suter D. Incremental kernel principal component analysis. *IEEE Transactions on Image Processing* *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **16**(6): 1662–1674
 - 12 Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting. In: Proceedings of the 2006 British Machine Vision Conference. Edinburgh, England: BMVA Press, 2006. **1**: 47–56
 - 13 Zhang K H, Zhang L, Yang M H. Fast compressive tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(10): 2002–2015
 - 14 Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(7): 1409–1422
 - 15 Bolme D S, Beveridge J R, Draper B A, Lui Y M. Visual object tracking using adaptive correlation filters. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 2544–2550
 - 16 Henriques J F, Caseiro R, Martins P, Batista J. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(3): 583–596
 - 17 Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration. In: Proceedings of the 2014 European Conference on Computer Vision Workshops. Zurich, Switzerland: Springer Berlin Heidelberg, 2014. 254–265
 - 18 Danelljan M, Häger G, Shahbaz K F, Felsberg M. Accurate scale estimation for robust visual tracking. In: Proceedings of the 2014 British Machine Vision Conference. Nottingham, UK: BMVA Press, 2014. 65.1–65.11
 - 19 Danelljan M, Häger G, Shahbaz K F, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 4310–4318
 - 20 Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr P H S. Staple: complementary learners for real-time tracking. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 1401–1409
 - 21 Zuo W M, Wu X H, Lin L, Zhang L, Yang M -H. Learning support correlation filters for visual tracking. arXiv preprint arXiv: 1601.06032, 2016.
 - 22 Xu Jian-Qiang, Lu Yao. Robust visual tracking via weighted spatio-temporal context learning. *Acta Automatica Sinica*, 2015, **41**(11): 1901–1912
(徐建强, 陆耀. 一种基于加权时空上下文的鲁棒视觉跟踪算法. 自动化学报, 2015, **41**(11): 1901–1912)
 - 23 Ma C, Yang X K, Zhang C Y, Yang M H. Long-term correlation tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. 5388–5396
 - 24 Zokai S, Wolberg G. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations. *IEEE Transactions on Image Processing*, 2005, **14**(10): 1422–1434
 - 25 Reddy B S, Chatterji B N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 1996, **5**(8): 1266–1271

- 26 Sarvaiya J N, Patnaik S, Kothari K. Image registration using log polar transform and phase correlation to recover higher scale. *Journal of Pattern Recognition Research*, 2012, **7**(1): 90–105
- 27 Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2006, **1**: 798–805
- 28 Hare S, Golodetz S, Saffari A, Vineet V, Cheng M M, Hicks S L, Torr P H S et al. Struck: structured output tracking with kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(10): 2096–2109
- 29 Zhong W, Lu H C, Yang M H. Robust object tracking via sparsity-based collaborative model. In: Proceedings of the 2012 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1838–1845
- 30 Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of the 2012 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1822–1829
- 31 Dinh T B, Vo N, Medioni G. Context tracker: exploring supporters and distracters in unconstrained environments. In: Proceedings of the 2011 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 1177–1184

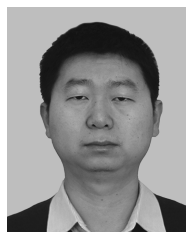


熊丹 国防科技大学智能科学学院博士研究生. 2012 年获得国防科技大学硕士学位. 主要研究方向为机器人视觉和视觉 SLAM.

E-mail: xiongdan@nudt.edu.cn

(**XIONG Dan** Ph.D. candidate at the College of Intelligence Science and Technology, National University of Defense Technology. He received his master degree from National University of Defense Technology in 2012. His research interest covers robot vision and visual SLAM.)

He received his master degree from National University of Defense Technology in 2012. His research interest covers robot vision and visual SLAM.)



卢惠民 国防科技大学智能科学学院副教授. 2010 年获得国防科技大学博士学位. 主要研究方向为机器人视觉, 视觉 SLAM 和机器人足球. 本文通信作者.

E-mail: lhmnew@nudt.edu.cn

(**LU Hui-Min** Associate professor at the College of Intelligence Science and Technology, National University of Defense Technology. He received his Ph.D. degree from National University of Defense Technology in 2010. His research interest covers robot vision, visual SLAM and robot soccer. Corresponding author of this paper.)

He received his Ph.D. degree from National University of Defense Technology in 2010. His research interest covers robot vision, visual SLAM and robot soccer. Corresponding author of this paper.)



肖军浩 国防科技大学智能科学学院讲师. 2007 年获得国防科技大学学士学位, 2013 年获得德国汉堡大学博士学位. 主要研究方向为移动机器人三维感知和多机器人协同控制.

E-mail: junhao.xiao@ieee.org

(**XIAO Jun-Hao** Lecturer at the College of Intelligence Science and Technology, National University of Defense Technology. He received his bachelor degree from National University of Defense Technology in 2007, and Ph.D. degree from University of Hamburg, Germany in 2013. His research interest covers 3D perception for mobile robots and multi-robot coordination.)

He received his bachelor degree from National University of Defense Technology in 2007, and Ph.D. degree from University of Hamburg, Germany in 2013. His research interest covers 3D perception for mobile robots and multi-robot coordination.)



郑志强 国防科技大学智能科学学院教授. 1994 年获得比利时列日大学博士学位. 主要研究方向为多机器人协同控制, 飞行器控制.

E-mail: zqzheng@nudt.edu.cn

(**ZHENG Zhi-Qiang** Professor at the College of Intelligence Science and Technology, National University of Defense Technology. He received his Ph.D. degree from University of Liege, Belgium in 1994. His research interest covers multi-robot coordination control and flight control.)

He received his Ph.D. degree from University of Liege, Belgium in 1994. His research interest covers multi-robot coordination control and flight control.)