

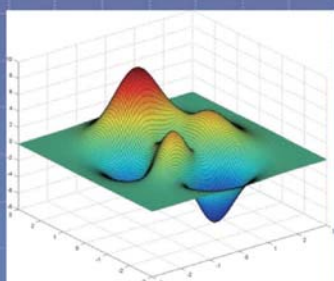


主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

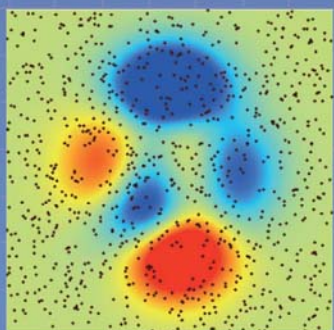
中国图象图形学报

2014
02
VOL.19

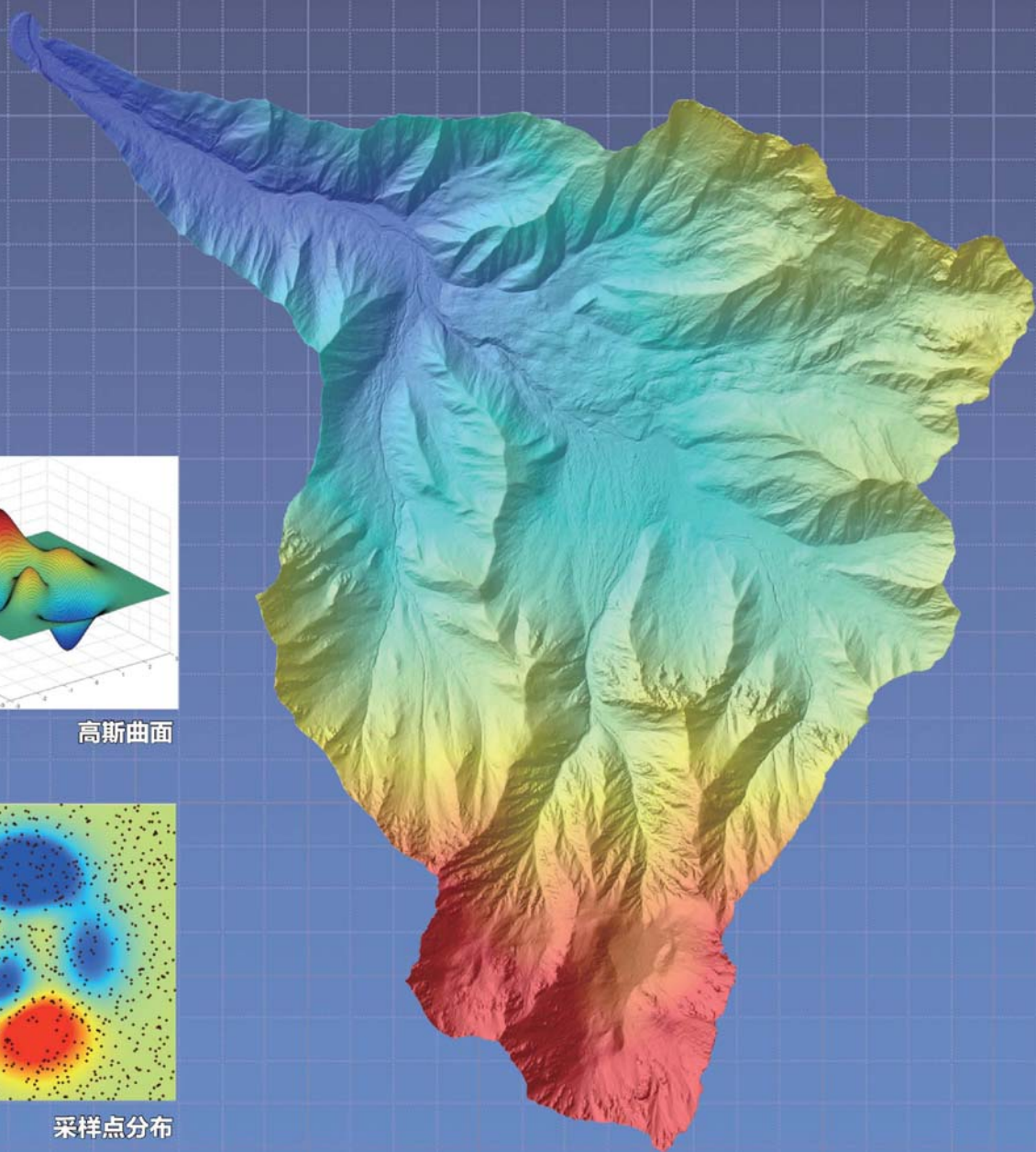
ISSN1006-8961
CN11-3758/TB



高斯曲面



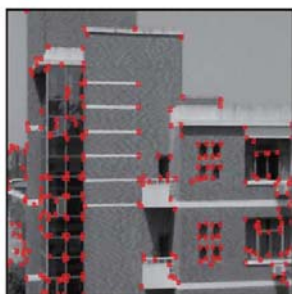
采样点分布



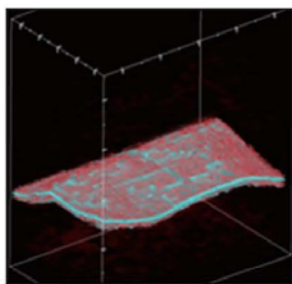
HASM地形模拟 P290



邻域最短距离法寻找最佳拼接缝(第227页)



曲率尺度空间与链码方向统计的角点检测(第234页)



目标特征指导的多分辨率体绘制算法(第283页)

计算机视觉前沿论坛

行为理解的认知推理方法

陶霖密, 杨卓宁, 王国建 0167

深度学习及其在目标和行为识别中的新进展

郑胤, 陈权崎, 章毓晋 0175

图像处理和编码

分布式视频编码中基于改进FCM聚类的相关噪声模型估计

杨春玲, 吴娟 0185

结合一阶自回归滑动平均和压缩感知的视频模型

王教余, 徐小红, 沈仁明, 廖重阳, 杨勋 0194

信息量加权的梯度显著度图像质量评价

徐少平, 杨荣昌, 刘小平 0201

优化加权TV的复合正则化压缩感知图像重建

费选, 韦志辉, 肖亮, 李星秀 0211

消除图像伪轮廓的各向异性自适应滤波

倪婧, 王朔中, 廖纯, 曾兴 0219

邻域最短距离法寻找最佳拼接缝

郑悦, 程红, 孙文邦 0227

图像分析和识别

曲率尺度空间与链码方向统计的角点检测

曾接贤, 李炜烨 0234

基于方向特征的手掌静脉识别

周宇佳, 刘娅琴, 杨丰, 黄靖 0243

复杂环境下高效物体跟踪级联分类器

江伟坚, 郭躬德 0253

图像理解和计算机视觉

均值规范化对比度的局部特征描述符

颜雪军, 赵春霞, 袁夏, 徐丹, 刘凡 0266

计算机图形学

三角域上Said-Ball基的推广渐近迭代逼近

张莉, 李园园, 杨燕, 檀结庆 0275

目标特征指导的多分辨率体绘制算法

郭思奇, 鲁才, 聂小燕 0283

地理信息技术

高精度曲面建模优化方案

赵明伟, 岳天祥, 赵娜 0290

医学图像处理

应用于人体关节缺损修复的建模与可视化

邢慧君, 杨健, 李勤 0297

小邻域统计信息核磁共振医学图像分割模型

张建伟, 方林, 陈允杰, 詹天明 0305

遥感图像处理

结合暗通道原理和双边滤波的遥感图像增强

周雨薇, 陈强, 孙权森, 胡宝鹏 0313

改进NSCT和IHS变换相结合的遥感影像融合

刘慧, 周可法, 王金林, 王珊珊 0322

分段2维主成分分析的超光谱图像波段选择

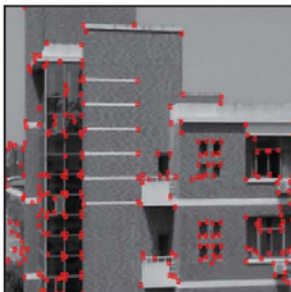
张婧, 孙俊喜, 阮光诗, 刘红喜 0328

CONTENTS

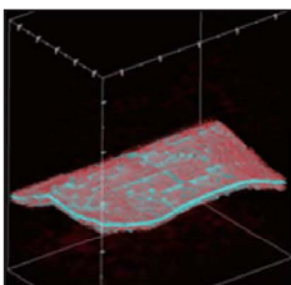
JOURNAL OF IMAGE AND GRAPHICS



Finding an optimal seam-line through the shortest distance in the neighborhood(P227)



Corner detection based on curvature scale space and chain code direction statistics(P234)



Target feature based multi-resolution volume rendering(P283)

Forum: Forefront of Computer Vision

Cognitive reasoning method for behavior understanding

Tao Linmi, Yang Zhuoning, Wang Guojian 0167

Deep learning and its new progress in object and behavior recognition

Zheng Yin, Chen Quanqi, Zhang Yujin 0175

Image Processing and Coding

Correlation noise modeling based on improved Fuzzy C-Means clustering in distributed video coding

Yang Chunling, Wu Juan 0185

Compressed sensing video model based on a first-order autoregressive moving average model

Wang Jiaoyu, Xu Xiaohong, Shen Renming, Liao Chongyang, Yang Xun 0194

Information content weighted gradient saliency structural similarity index for image quality assessment

Xu Shaoping, Yang Rongchang, Liu Xiaoping 0201

Compound regularized compressed sensing image reconstruction based on optimal reweighted TV

Fei Xuan, Wei Zhihui, Xiao Liang, Li Xingxiu 0211

False contour suppression with anisotropic adaptive filtering

Ni Jing, Wang Shuozhong, Liao Chun, Zeng Xing 0219

Finding an optimal seam-line through the shortest distance in the neighborhood

Zheng Yue, Cheng Hong, Sun Wenbang 0227

Image Analysis and Recognition

Corner detection based on curvature scale space and chain code direction statistics

Zeng Jiexian, Li Weiye 0234

Palm-vein recognition based on oriented features

Zhou Yujia, Liu Yaqin, Yang Feng, Huang Jing 0243

Efficient cascade classifier for object tracking in complex conditions

Jiang Weijian, Guo Gongde 0253

Image Understanding and Computer Vision

Local feature descriptor based on mean normalized contrast

Yan Xuejun, Zhao Chunxia, Yuan Xia, Xu Dan, Liu Fan 0266

Computer Graphics

Generalized progressive iterative approximation for Said-Ball based on triangular domains

Zhang Li, Li Yuanyuan, Yang Yan, Tan Jieqing 0275

Target feature based multi-resolution volume rendering

Guo Siqi, Lu Cai, Nie Xiaoyan 0283

Geoinformatics

HASM optimization based on the improved difference scheme

Zhao Mingwei, Yue Tianxiang, Zhao Na 0290

Medical Image Processing

Method of modeling and visualization for repairing of osteoarticular defect

Xing Huijun, Yang Jian, Li Qin 0297

Magnetic resonance medical images segmentation based on a local statistical information model

Zhang Jianwei, Fang Lin, Chen Yunjie, Zhan Tianming 0305

Remote Sensing Image Processing

Remote sensing image enhancement based on dark channel prior and bilateral filtering

Zhou Yuwei, Chen Qiang, Sun Quansen, Hu Baopeng 0313

Remote sensing image fusion based on an improved NSCT and IHS transformation

Liu Hui, Zhou Kefa, Wang Jinlin, Wang Shanshan 0322

Segmented 2DPCA algorithm for band selection of hyperspectral image

Zhang Jing, Sun Junxi, Ruan Guangshi, Liu Hongxi 0328

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2014)2-0253-13

论文引用格式: 江伟坚, 郭躬德. 复杂环境下高效物体跟踪级联分类器[J]. 中国图象图形学报, 2014, 19(2): 253-265. [DOI: 10.11834/jig.20140211]

复杂环境下高效物体跟踪级联分类器

江伟坚^{1,2}, 郭躬德^{1,2}

1. 福建师范大学数学与计算机科学学院, 福州 350007;

2. 福建师范大学网络安全与密码技术福建省重点实验室, 福州 350007

摘要: **目的** 传统跟踪算法在复杂环境下容易发生漂移(drift)现象,通过改进TLD(tracking learning detection)跟踪技术算法,提出了基于Sliding-window的局部搜索和全局搜索策略、积分直方图过滤器和随机Haar-like块特征过滤器。**方法** 首先,采用积分直方图过滤器可以有效地过滤大量非目标子窗口块,从而减少后续过滤器特征匹配数;其次,利用随机Haar-like块特征过滤器能够解决跟踪算法在复杂环境(多物体、部分或较大区域遮挡、快速运动等)跟踪过程易发生漂移而导致跟踪精度的不足。**结果** 结合TLD原始过滤器与新提出的两个过滤器组合而成的级联分类器,通过与主流的跟踪算法实验进行对比表明,级联分类器在稳定的背景或复杂环境的跟踪鲁棒性强、跟踪精度高,并且采用了局部和全局搜索策略提高了计算速度。**结论** 提出的方法在诸多背景环境变化,跟踪物体形变等情况下,能够精确地多尺度跟踪待测目标;结合全局和局部搜索跟踪策略能够有效地克服级联分类器所带来的时间复杂度过高的问题,从而实现实时目标跟踪。

关键词: 视觉追踪; Haar-like 特征; 级联分类器; TLD 算法; 积分直方图

Efficient cascade classifier for object tracking in complex conditions

Jiang Weijian^{1,2}, Guo Gongde^{1,2}

1. School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, China;

2. The key laboratory of Network Security and cryptographic technology, Fujian Normal University, Fuzhou 350007, China

Abstract: **Objective** We improve the TLD algorithm and propose local and global search based on the sliding-window method, the Integral Histogram Filter, and Random Haar-like Feature Filter to solve the drift problem of traditional tracking algorithms in complex conditions. **Method** First, we use the Integral Histogram Filter to reject the Sliding-window patches as quickly as possible to release the feature matching in the following filters. Then, we use Random Haar-like Feature Filter to overcome the drift problem, which causes a loss of accuracy during the object tracking under complex conditions (multi-object, occlusion, fast movement). **Result** We ultimately combine filters of the TLD algorithm and two new filters of our proposed. The experimental results show that the proposed approaches compared with the traditional tracking algorithms not only presents robustness and tracking accuracy in stable background or complex conditions, but also obtains the best computing speed with the use of the local and global search. **Conclusion** The proposed method is able to detect the multi-scale object accurately both in different environment and tracking object deformation. Combining the global and local search strategy can overcome the time consuming effectively to achieve the real-time object tracking.

Key words: computer vision tracking; Haar-likes feature; cascade classifier; TLD algorithm; integral histogram

收稿日期: 2013-06-17; 修回日期: 2013-09-03

第一作者简介: 江伟坚(1988—), 男, 福建师范大学数学与计算机学院在读硕士研究生, 主要研究方向为视觉跟踪、模式识别和机器学习。E-mail: kongwaigain@gmail.com

通信作者: 郭躬德, E-mail: ggd@fjnu.edu.cn

0 引言

视觉追踪技术是计算机视觉领域的一个重要课题。在过去 30 年间,研究人员不断提出了许多跟踪算法^[1-10]应用到诸如手势识别、人机交互、车辆导航、人脸识别和安全监控等方面。但是,现有的技术仍然面对一系列的挑战。图像从 3D 现实场景到 2D 的投影导致空间信息流失,跟踪的物体客观有存在噪点的可能,物体在跟踪过程中的半遮挡或者全遮挡,背景信息的干扰,物体的运动模糊,光照的变化,物体的不同尺度和不同视角等都给视觉追踪技术带来了挑战^[11-12],对于要设计一个好的追踪系统,评价系统的优劣需要考虑以下 3 个方面^[12]:1) 鲁棒性;2) 准确性;3) 快速性。

近年来,随着研究工作的深入,在某种程度上能够解决复杂变化环境下的物体追踪。从不同体征表现框架来分,跟踪模型可以分为两类: Generative Models^[4-5,13] 和 Discriminative Model^[6-8]。Generative Models 须事先构建一个模型用来表示物体,然后通过对于每一帧视频流进行相似性匹配从而达到物体追踪的目的。Ross 等人^[14]提出了 IVT(增量视觉追踪)跟踪算法,其本质上是一个在线学习模型,但是对于遮挡物体的跟踪效果不好。Adam 等人^[4]提出了一种方块集合用来解决物体的部分遮挡而丢失跟踪物体的弊端。Discriminative Models 则是把追踪物体任务看做是二元分类任务,目的是把跟踪目标从背景信息中分离出来。一般来说,这些方法需要利用前几帧视频流样本训练出分类器,再通过训练出的分类器对后续的视频帧进行分类并更新当前的分类器用于下帧跟踪。Boris 等人^[9]提出了一种新颖的跟踪算法 MILTrack。其实质是监督学习的一种改进形式,训练数据不是以单个样例进行标注,而是采用多样例包袋(bag of instances)标注,对训练样本标签进行模糊化处理达到弱化监督效应的目的。Zhang 等人^[6]提出了一种简单高效的基于压缩感知的跟踪算法,利用压缩感知 RIP(restricted isometry property)条件的随机感知矩对多尺度图像进行特征降维,对降维后的特征采用简单的朴素贝叶斯分类器进行分类。其优点是跟踪算法十分简单,实验结果有很好的鲁棒性,但是对于多尺度问题还是未能得到根本的解决。Kalal 等人^[1-3]提出了一种单目标长时间跟踪算法,与传统跟踪算法的显著区别在于

它融合了传统跟踪算法和传统检测算法来解决跟踪目标在被跟踪过程所发生的形变和部分遮挡等问题,并且通过一种改进的在线学习机制不断修正其目标模型和跟踪参数以达到重定位后期物体发生形变的目的。但是在解决对于复杂环境的精度较低,检测速度较慢。本文旨在解决此问题提出了改进方法,不仅在低分辨率和较高分辨率下能够有效地提高检测速度,而且对于稳定环境或者复杂环境下有较高的鲁棒性。

1 TLD 跟踪算法

TLD(tracking learning detection)跟踪算法^[1]由 3 个部分组成,即跟踪器(tracker)、检测器(detection)和在线学习(online learning)算法。其中跟踪器和检测器两者并行运行,所产生的结果皆参与学习过程,学习后的模型反作用于跟踪器和检测器,学习过程对跟踪器和检测器进行实时更新模型和跟踪参数,从而保证了目标外观在发生变化的情况下能够持续被跟踪。3 个组成部分的描述如下:

1) 跟踪器 采用层叠块式跟踪方法,即指定跟踪的目标用矩形框标出,采用 Lucas-Kanade 光流法^[15]且结合相似性判断和前向-后向误差(forward-backward error)法^[3]来估计目标运动位置,用此局部跟踪方法可以有效解决部分遮挡的问题。

2) 检测器 跟踪器无法解决跟踪大运动物体或者物体临时丢失后重现的重新定位问题,检测器旨在解决此类问题而设计,用来弥补跟踪器的不足。主要流程如下:

(1) 每一帧图像通过 Sliding-window 方法^[16-17]扫描图像,形成上万个图像块(patch)。

(2) 每个图像块被标记为正例或负例需要经过 3 层级联分类器^[17]的过滤:

① 方差过滤器(variance filter) 设 $P_i, i = 0, 1, \dots, n$ 为滑动窗口图像块,其中 P_0 为手动标定的跟踪目标块, $V(i)$ 表示第 i 个图像块的方差,若 $V(P_i) > 0.5 \times V(P_0)$, 图像块 P_i 将会顺利通过方差过滤器。

② 随机森林分类器(random ferns filter) 由几组二值化特征组合而成的 Fern 特征^[17],其特征选取为区域边缘方向,它具有不受光线干扰的特性。为了解决多窗口特征块的计算速度,采用积分图像^[17]的方法。

③ 最近邻分类器(nearest-neighbor classifier) 用于过滤度量图像块 P 同目标模型 M 之间的相似程

度低的图像块。如果相似度 $S(P, M) > \theta$ (θ 为待更新的阈值, 初始阈值 $\theta = 0.65$), 那么该图像块将 P 分类为正样例, 否则将划分为负样例。

3) 在线学习算法 该过程是一个动态模型, 随着视频序列的扩张或缩小, 其发展由两个事件进行驱动, 分别为扩张事件 (growing event) 和修剪事件 (pruning event)。样本空间的扩张来自于跟踪器, 它将符合目标样本的样本加入正样本池, 扩张正样例。修剪事件假设每帧只有一个目标, 若跟踪器和检测器确定目标位置, 其余检测的图像块皆认为是错误样本, 从在线模型剔除。

2 改进算法

TLD 算法对于复杂环境的鲁棒性低, 且容易发

生漂移现象 (drift) 导致检测精度较低以及检测速度较慢。对于此类问题, 提出的改进算法能够有效地解决。1) 在原有算法分类器的基础上引入积分直方图过滤器和随机 Haar-like 块过滤器。积分直方图过滤器可以有效地过滤 Sliding-window 方法所产生的与目标块相似度较低的块窗口, 减少待匹配子窗口块从而提高特征提取速度; 随机 Haar-like 块过滤器基于 Haar-like 特征提取待测目标的纹理信息, 能够提高大尺度下的物体跟踪精度。2) 检测器方面的搜索策略采用局部搜索和全局搜索相结合的方式。由于物体帧间运动发生“小运动”概率较大, 采用局部搜索可以有效地提高检测速度; 如果物体发生快速运动, 全局搜索则能够弥补局部搜索的物体丢失问题。图 1 所示为本文改进算法 (ETLD) 的流程框架图, 蓝色的方框表示改进的方法。

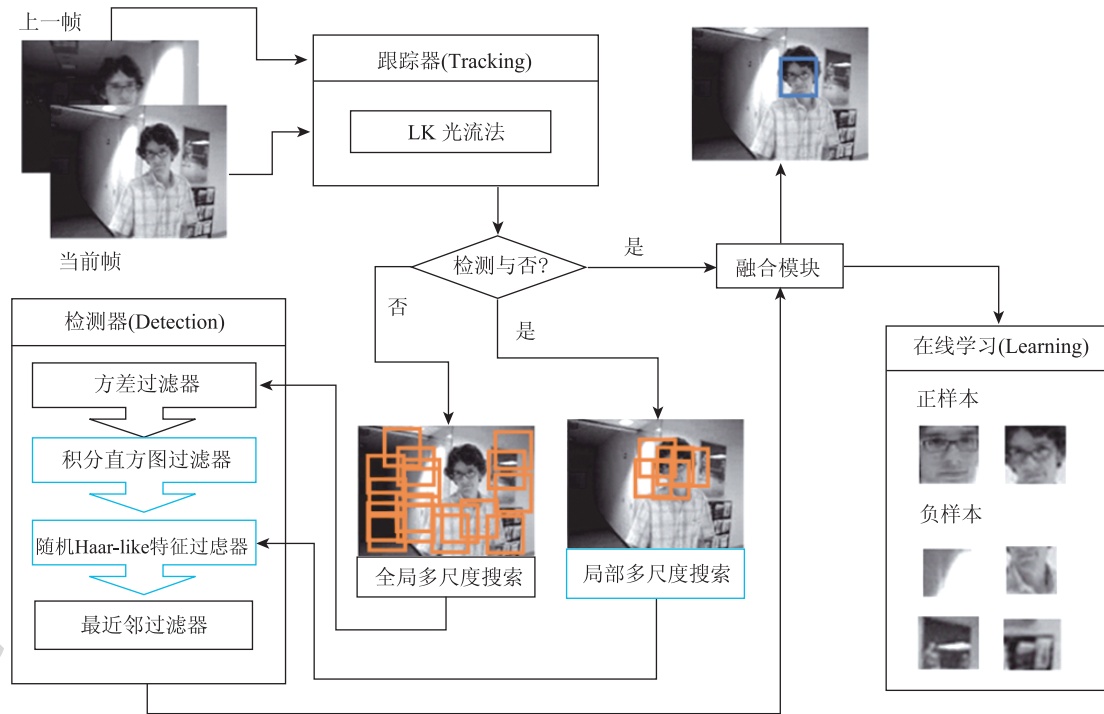


图 1 改进算法 (ETLD) 的框架图

Fig. 1 The framework of improved algorithm (ETLD)

2.1 过滤器

2.1.1 积分直方图过滤器

TLD 算法的方差过滤器 (variance filter) 能够过滤掉与目标模板相似度低的图像块, 并且采用积分图像^[19]有效提高了块方差的计算速度。图 2 中红色矩形框计算出的方差低于目标 (绿色矩形框) 的方差, 大于某个阈值的方差的图像片才能通过方差

过滤器 (蓝色矩形标记框)。但是通过实验发现, 方差过滤器在复杂的环境中 (多物体, 复杂背景) 能够过滤的窗口数较少, 而且被锁定的前景目标和背景目标区分度不高。如图 3(a) 所示, 对于类似多物体的环境下, 方差过滤器过滤掉方差相似度低的图像块较少, 导致进入到集合分类器的样本数目十分庞大, 严重影响随机森林分类器的识别速度。因此, 为了

解决在复杂环境下过滤器过滤效果差的问题,提出了一个积分直方图过滤器,它能够有效地过滤与目标相似度低的图像块。

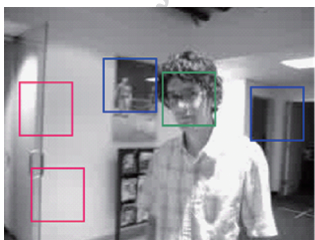


图2 矩形块方差示意图

Fig. 2 The rectangular variance of the diagram



图3 Sliding-window 块集合通过过滤器

Fig. 3 Filters reject the Sliding-window patches

积分直方图^[18]的计算与积分图像十分相似,给定一个矩形图像块 $I(x, y, w, h)$,就可以通过简单的加减运算出对应图像块的直方图,而且不管图像块的位置大小,其计算复杂度皆为 $O(1)$ 。对于给定一个目标图像块 O ,其积分直方图值为 H_o ,那么对于待测图像块 P_i ,其积分直方图值为 H_{p_i} ,若 $S(H_o, H_{p_i}) > \theta$,即两个图像块的直方图相似度大于 θ ,那么 P_i 的似然区域将被用作下一层过滤器评估。但是对于灰阶图像块255 bins的直方图似然计算复杂度较高,为了进一步提高过滤速度,采取了16 bins直方图的相似度度量方式,采用了文献[4]中所提到的EMD (earth mover's distance) 相似度度量方法。图3(b)是Sliding-window所产生的图像块集合经过方差过滤器后所得到的图像块,虽然滤掉了部分子窗口块,但是依然存在大量的块窗口。图3(c)是图像块经过积分直方图过滤器的结果,虽然采用了16 bins的直方图相似度匹配,其目标匹配精度有所降低,但还是能够得到比较满意的过滤效果。从实验结果来看,绝大部分的待筛选块圈定在待测目标的周围,剩下图像块的分类任务就由下一层较复杂的过滤器来进行分类。

2.1.2 随机 Haar-like 块过滤器

近年来,在特征提取方面许多学者提出了具

有一定鲁棒性的局部特征描述子。Lowe 等人^[13,19]提出了 SIFT (scale-invariant feature transform) 描述子,该特征具有尺度不变性和旋转不变性的特点,并且具有一定的抗光照、抗噪等优势,但是计算复杂度代价较高。紧接着, Bay 等人^[20]提出了 SURF (speeded up robust feature) 特征, Rublee^[24]提出了 ORB 特征, Alahi^[21]提出了 FREAK (fast retina keypoint), 一定程度上解决了 SIFT 特征的高计算复杂度。但是在复杂环境下的跟踪上,以上描述子还是难以达到实时性。Ozuysal 等人^[22]提出了一种基于灰度的随机比对点串,即随机蕨。其计算十分简单且抗干扰能力强。随机蕨实质是非层次结构特征,该特征在基于朴素贝叶斯分类算法能够对局部图像进行快速识别和定位。它是由 m 组随机树组成的 Fern 特征,该特征由二进制组合而成的二元特征串,串长为 s ,组数设为 m ,那么 Fern 特征记为 $f_i, i = 1, 2, \dots, m$;二元特征 f_i 的值取决于子窗口块中分类器随机产生的两个像素位置 $d_{i,1}$ 、 $d_{i,2}$ 的灰度大小 $I(d_{i,1})$ 、 $I(d_{i,2})$ 的灰度比较所得的结果(如图4中随机蕨虚线框内的不同像素点比对示意图),即

$$f_i = \begin{cases} 1 & I(d_{i,1}) < I(d_{i,2}) \\ 0 & \text{其他} \end{cases} \quad (1)$$

因此,采用了多组比对结果进行正负例投票,取最大投票值作为目标。为了保证随机蕨过滤器计算的实时性和多尺度下窗口块匹配的准确性, TLD 算法中设定的参数为 $M = 10, S = 13$,即表示有10棵随机树,其中随机蕨二元特征串长为13,存储量为 $M \times 2^S$ 字节,有利于工程的实现。但是在较高分辨率下(640 × 480),要识别较大尺度的目标物体,需要更多的随机蕨特征组用来提高目标的识别率。然而,纯粹增加 M 和 S 的组数将会增加计算复杂度,而且通过手动调节参数去应对不同分辨率下的多尺度追踪问题将会导致算法的适应力较差。Haar-like 特征^[19]能够提取具有一定纹理特征的目标,提出了一种基于 Haar-like 特征的块比对与原始随机蕨分类器组成线性分类器,不仅能够自适应于不同分辨率下的目标跟踪,而且两者相结合的方法在大尺度下的目标跟踪精确度比原来单一的随机蕨分类器的精确性高。其中,线性分类器 w 表示权重,并进行归一化,初始值均为0.5。随机 Haar-like 块比对方法的思想类似于随机蕨方法,图4阐述了改进算法的主要思路。

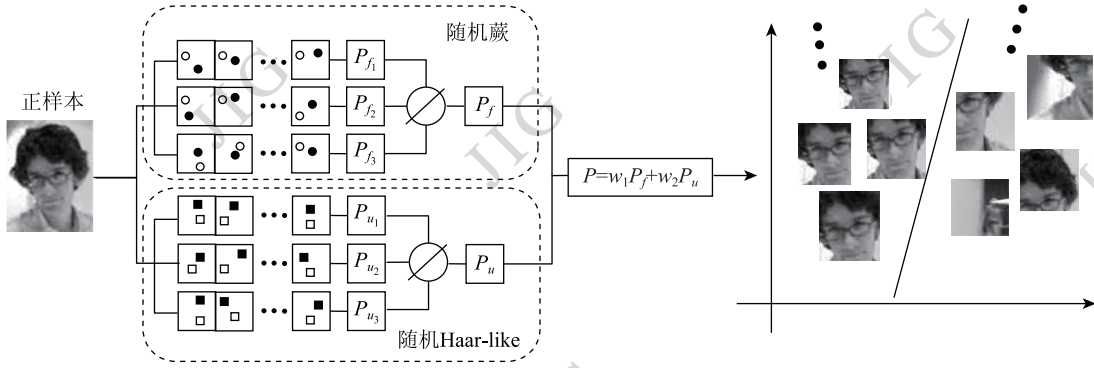


图 4 随机蕨和随机 Haar-like 块过滤器组合而成的分类器示意图

Fig. 4 The combined classifier used by random ferns and random Haar-like feature filters

随机 Haar-like 块的算法思想: 设 $c_i, i = 0, 1$ 表示特征区域所属的类别 (本文为二分类问题, c_0 表示负样本, c_1 表示正样本), 目标图像的特征块为 T , 查找待测图像特征块 O 在目标图像上的对应特征可以视为对图像块的分类问题。那么, 设 $f_j, j = 1, \dots, u$, 为 $P = w_1 P_f + w_2 P_u$ 组基于 Haar-like 的二进制特征串, 图像块所属类别, 即

$$\hat{c}_i = \arg \max_{c_i} P(C = c_i | f_1, f_2, \dots, f_u) \quad (2)$$

式中, c_i 为类的随机变量, 由贝叶斯公式可得

$$P(C = c_i | f_1, f_2, \dots, f_u) = \frac{P(f_1, f_2, \dots, f_u | C = c_i) P(C = c_i)}{P(f_1, f_2, \dots, f_u)} \quad (3)$$

式(3)中分母与分类无关, 那么式(3)可简化为

$$\hat{c}_i = \arg \max_{c_i} P(f_1, f_2, \dots, f_u | C = c_i) \quad (4)$$

式中, f_j 的值取决于分类器在训练阶段随机生成的两相等随机矩形块 $D_{i,1}$ 和 $D_{i,2}$ 所在位置的灰度大小比较的结果, 即

$$f_j = \begin{cases} 1 & I(D_{i,1}) < I(D_{i,2}) \\ 0 & \text{其他} \end{cases} \quad (5)$$

式中, 矩形块的大小由对应尺度空间的大小所决定, 具体设定大小参见 2.2 节的分析。

随机矩形块的灰度值可以通过积分图快速计算得出, 且计算复杂度为 $O(1)$, 可以加快随机矩形块的计算。那么, 采用多组的随机矩形块来提取特征, 然后计算后验概率 (类似随机蕨)。设 f_j 有 N 个矩形组比对结果, 即 N 对矩形所在位置的灰度比较大小, 则求解式(4)中联合分布概率的表达形式的求解需要存储量为 $U \times 2^N$ 字节。为了降低存储量, 假设 f_j 之间相关, 即每组包含 $k = N/U$ 个二元特征 (类似随机蕨串长), 且不同组的二元特征之间相互独立, 那么式(4)中的条件概率近似为

$$P(f_1, f_2, \dots, f_u | C = c_i) = \prod_{u=1}^U P(F_u | C = c_i) \quad (6)$$
 式中, $F_u = [f_{\sigma(u,1)}, f_{\sigma(u,2)}, \dots, f_{\sigma(u,S)}]$, $u = 1, \dots, U$, 表示第 u 个蕨, $\sigma(u, j)$ 表示范围为 $1, \dots, U$ 的随机函数, 那么图像块所属类别即为

$$\hat{c}_i = \arg \max_{c_i} \prod_{u=1}^U P(F_u | C = c_i) \quad (7)$$

为了求解式(7), 仅需要估计每个蕨 F_u 和类 c_i 的类条件概率 $P(F_u | C = c_i)$, 即计算图像块的 Haar-like 块组特征的后验概率, 那么式(7)的存储量为 $U \times 2^k$ 字节。求解块特征串的计算伪代码如算法 1 所示。其中, I 为输入图像, F 为待测随机 Haar-like 块特征。

算法 1

Input: I

Output: F

$F \leftarrow 0$

for $i = 1 : H$ do

$F \leftarrow F < 1$

if $I(D_{i,1}) < I(D_{i,2})$ then

$F \leftarrow F + 1$

end if

end for.

2.2 搜索策略

待跟踪目标在帧与帧之间往往发生较小的运动, 因此跟踪目标可以在局部的小范围内进行 Sliding-window 搜索块搜索。采用局部滑动搜索框可以有效地节省跟踪算法的执行时间。当目标发生大运动或者在视野丢失重现时, 全局滑动搜索

的策略才会有更好的鲁棒性(参见图1搜索策略改进方案)。因此在搜索策略方案的选择采用两者相结合的方式,即假设跟踪目标是小运动,当检测器检测结果无效时,采取全局搜索方式(已经搜索的局部区域不再进行搜索)。假设一幅分辨率为 640×480 的图像,那么就会生成 23 507 020 800 个滑动子窗口块,显然特征匹配计算达不到实时性,对此进行了以下限定:设滑动子窗口块的放大因子 $s = 1.2^b$, $b \in \{-10, -9, \dots, 9, 10\}$ 。子窗口块 x 轴和 y 轴的偏移步长 $dx = dy = 2$ pixel,那么生成的滑动窗口数 R 可以用下式给出,即

$$R = \sum_{s \in 1.2^{\{-10, -9, \dots, 9, 10\}}} \left\lceil \frac{n - (w + dx)}{dx} \right\rceil \left\lceil \frac{m - (h + dy)}{dy} \right\rceil \quad (8)$$

式中, n 和 m 表示图像 I 的长和宽, w 和 h 表示初始待测目标块的长和宽。那么假设初始目标块 $w = 80, h = 60$,那么式(8)中可以算出生成子窗口数为 160 809 个,其计算量大大降低且基本可以满足一般的精度需求。那么在目标为“小运动”的前提下,采用局部搜索策略后,将进一步减少子窗口特征匹配数。

设局部窗口块集合 P' , P_T 为当前跟踪到的目标块, P_i^s 表示的是 s 尺度第 i 个窗口块,那么 $P' \in \{P_1^s, P_2^s, \dots, P_k^s | R_{P_i^s} - R_{P_T} < r\}$ 。 $R_{P_i^s}$ 为 P_i^s 窗口块的半径, R_{P_T} 为 P_T 目标块的半径。以目标块中心为圆心,半径为 r 的圆内所有子窗口块即为局部窗口块集合 P' 。其中, r 初始值为 20 个像素点。搜索策略算法伪代码如算法2所示。其中,步骤3)~5)为局部搜索过滤策略。由于局部存在目标的概率较大且子窗口块较少,直接采用随机蕨和随机 Haar-like 块过滤器以及最近邻分类器组成的级联分类器;步骤8)~10)为全局搜索的级联过滤器。

但是随机 Haar-like 矩形块的大小随着不同尺度大小而变化。矩形块长宽相等,记为 a ,初始目标块宽为 W ,高为 H 。那么矩形块的大小为

$$a = \left\lfloor 1.1^b \times \frac{u \times \max(W, H)}{\min(W, H)} \right\rfloor \quad (9)$$

式中, u 初始参数为 5 个像素点。假设初始目标块宽 $W = 60$,高 $H = 50$,那么在尺度为 $s = 1.2^2$ ($b = 2$) 下的矩形块大小为 7。

算法2

for all of the image sequence

initial $P = \{P_1, P_2, \dots, P_n\}$; $Patch \leftarrow \emptyset$

$Patch \leftarrow P' = \{P'_1, P'_2, \dots, P'_k | R_{P_i} - R_{P_T} < r\}$

$Patch \leftarrow \text{EnsembleFilter}(Patch)$;

$Patch \leftarrow \text{NearestNeighbor}(Patch)$;

if $Patch \in \emptyset$ then

$Patch \leftarrow P - P'$

$Patch \leftarrow \text{VarianceFilter}(Patch)$;

$Patch \leftarrow \text{HistogramFilter}(Patch)$;

$Patch \leftarrow \text{EnsembleFilter}(Patch)$;

$Patch \leftarrow \text{NearestNeighbor}(Patch)$;

end if

end for.

3 实验结果

3.1 参数设定

本文算法是在 TLD 的基础上进行改进,因此对应的一些参数设定做了以下分析和限定:TLD 算法中随机蕨分类器的随机树初始 $M = 10$,蕨(Fern)串长 $S = 13$;另外 ETLD 的随机蕨分类器的参数值同 TLD 算法一样,而新引入的随机 Haar-like 块过滤器初始随机树 $H = 10$,组成的二元串长 $K = 13$ 。

3.2 数据集测试结果

对本文算法和 3 个主流跟踪算法^[2,4,9]在不同测试集上进行对比。其中 animal, shaking, football 序列参见文献[23]; girl, David, faceocc1, faceocc2 序列参见文献[9]; coke11, sylvester 序列参见文献[24]; jumping 序列参见文献[25];与本文算法的对比算法分别为: fragment tracker (Frag)^[4], MILtrack (MIL)^[9], TLD 算法^[1]。微机环境配置为 Intel Core (TM) i7 CPU I640@2.13 GHz,内存 4 GB。为了更好地验证 ETLD 算法的鲁棒性,将上述所罗列的测试集序列划分为静态稳定环境(girl, David, faceocc1, faceocc2, coke11, sylvester)和复杂环境(jumping, animal, shaking, football)两类分别对各算法进行测试。为了保证检测的公平性,各个算法初始的位置皆按照对应集 Ground Truth 的第 1 帧矩形块位置进行初始化,然后对各算法的运行检测结果同 Ground

Truth 对应帧矩形块位置进行跟踪精度误差分析,各序列 Ground Truth 参见文献[26]。

3.2.1 静态稳定环境

1) girl 序列 在该序列中,女孩的人脸绕水平轴旋转(俯仰)、绕视平面法向轴旋转(倾斜)、绕垂直轴旋转(摇摆)、人脸的部分遮挡和人脸多尺度变化。主要测试算法的对物体旋转、半遮挡和多尺度的鲁棒性。由图 5(a)序列可以看出,TLD 和 ETLD 算法对于多尺度跟踪定位都有较好的表现,跟踪块大小随着人脸远离摄像头而变

小,靠近摄像头而变大(见图 5(a) #71, #91, #325, #431),而 Frag 和 MIL 算法均无对多尺度进行检测,始终维持初始块大小直至跟踪结束。由于首帧标定的是人脸信息,当人脸发生丢失,ETLD 算法默认不进行标定位置,因此第 102 帧(见图 5(a) #102)无法圈定目标。

2) David 序列 该序列集描述了 David 从光线较暗的室内到较亮的室外的行走过程,其中人脸也相应发生旋转,人物表情变化以及行走中摘掉眼镜等姿态变化。主要用来测试算法对较慢的光线变

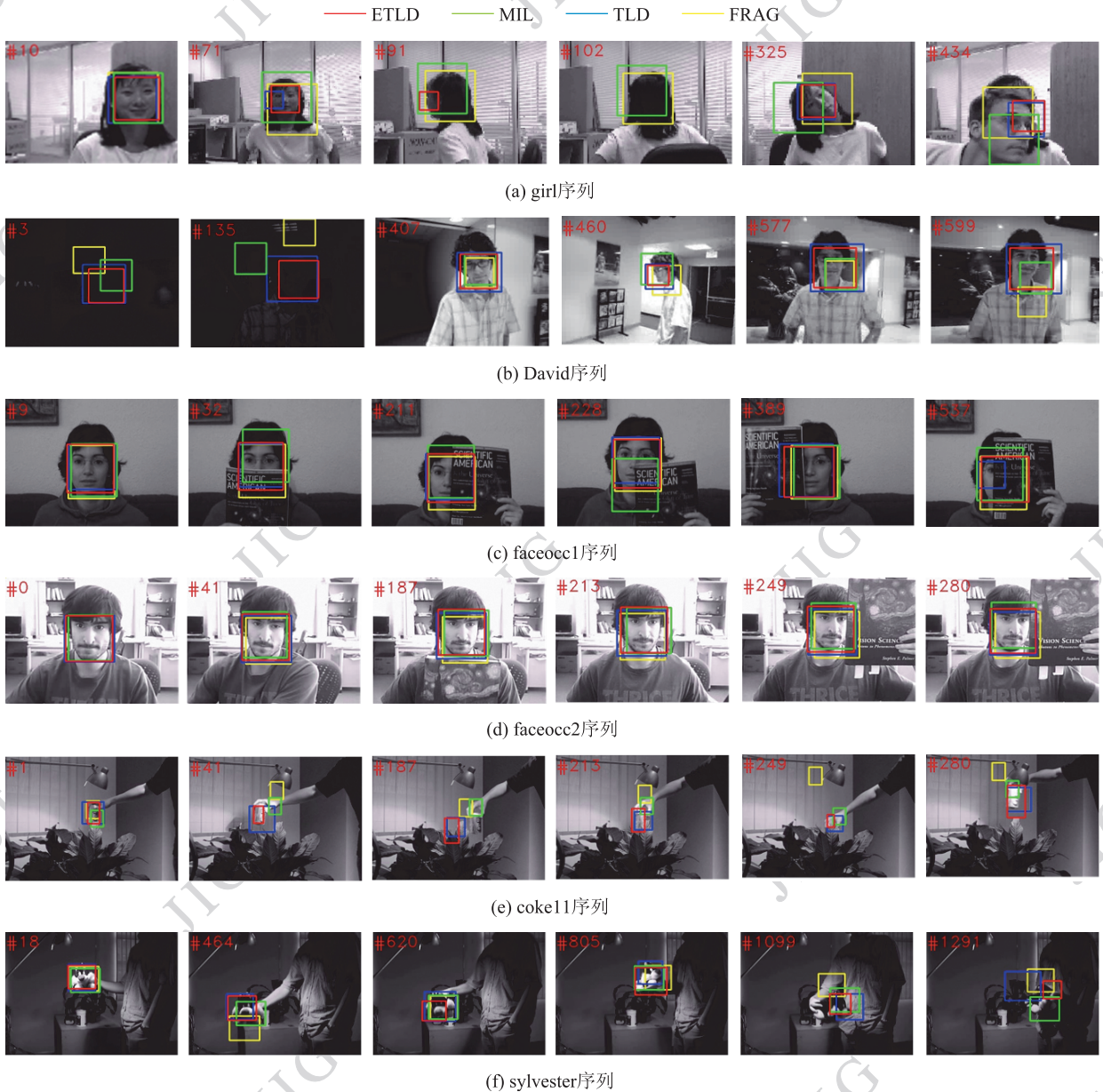


图 5 静态环境下各跟踪算法检测结果

Fig. 5 Tracking object location results in stable condition

化、物体旋转、尺度变化的鲁棒性表现。从图 5 (b) 序列可以看出,在初始窗口一致的前提下,第 3 帧(图 5(b)#3)的 Frag 和 MIL 算法已经发生了较大的漂移现象(drift),而且跟踪后期无法再次正确跟踪。可见这两个算法在对于光线较暗的特征提取方面表现较差;TLD 和 ETLTD 在光线较暗环境下跟踪稳定性都较好(图 5(b)#3, #135)。在第 407 帧光线较亮的区域,对 Frag 和 MIL 算法进行重新初始化后与 TLD 和 ETLTD 算法进行对比(图 5(b)#407)。在跟踪的后期阶段, Frag 跟踪精度最差,而 MIL 算法跟踪漂移不是很大, TLD 和 ETLTD 表现效果依旧很好且对于尺度变化检测效果较好(图 5(b)#460)。

3) faceocc1 和 faceocc2 序列 这两个序列集主要针对静态背景下,人脸的半遮挡、大面积遮挡进行算法的鲁棒性评估。对于该类型的测试集, Frag 算法对于不发生旋转的物体跟踪精度较好(图 5(c) #32, #211, #228, #389; 图 5(d) #138, #269, #631)。从 faceocc1 序列的跟踪误差可以看出, Frag 算法跟踪精度误差最小, ETLTD 次之,而 MIL 表现最差。但是在 faceocc2 序列中, Frag 算法在人脸有旋转的情况下跟

踪误差较大。综合两个静态背景的鲁棒性来看, ETLTD 算法的综合表现最好,跟踪精度和鲁棒性最好。

4) coke11 和 sylvester 序列 这两个序列集主要描述物体在运动过程下的姿态变化等情形,用于测试算法的尺度变化,光照变化和物体形变的鲁棒性测试。从图 5(e) 序列可以看出, Frag 算法在一开始发生了较大的漂移现象(见图 5(e) #41)。同样的在 sylvester 集中物体发生了较快运动也发生了漂移(见图 5(f) #464)。ETLTD 在后期检测中出现漏检情况,但是整体来看, ETLTD 的平均跟踪误差最小, TLD 次之,最后 Frag 算法表现最差。

3.2.2 复杂环境下的跟踪效果

1) jumping 和 animal 序列 该序列集描述了目标物体快速运动引起物体模糊拖影等现象,摄像机拍摄过程也处于运动状态,即动态背景。该序列集主要用来测试算法对于模糊拖影以及较快运动的目标物体的跟踪鲁棒性。Frag 算法对于快速运动物体的跟踪前期就发生了漂移(见图 6(a) #38, #62; 图 6(b) #13, #30, #33)。ETLTD 跟踪精度和跟踪检测率都为最好,其他算法都发生了不同程度的漂移以及

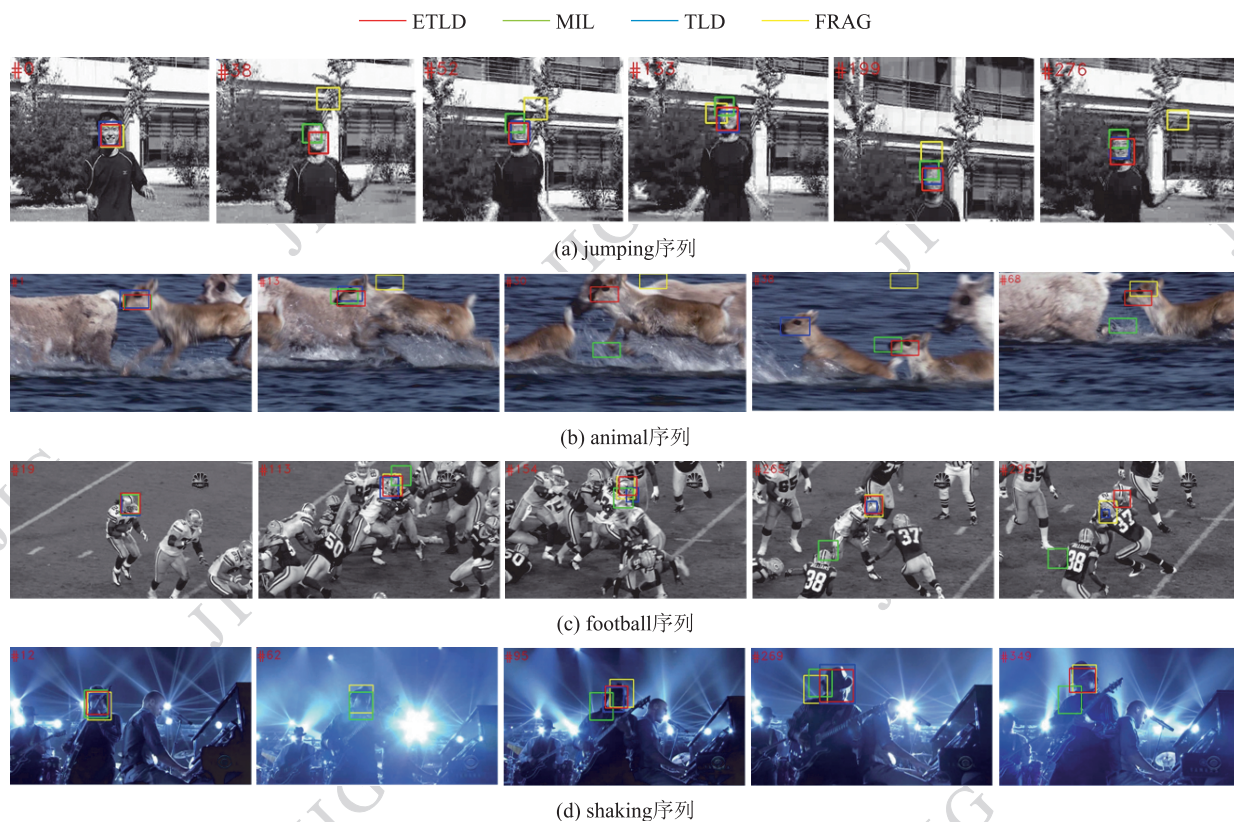


图 6 复杂环境下各跟踪算法检测结果

Fig. 6 Tracking object location results in complex condition

误判。其中,在 animal 序列集中 TLD 算法由于采取全搜索方式,容易导致几帧目标误判(见图 6(b)#38)。因此 ETLD 在加入局部搜索策略以及随机 Haar-like 过滤器提高了跟踪精度,并且有效地防止与目标模板类似而导致的误判问题。表 1 可以看出该序列集下 ETLD 算法的检测率明显高于 TLD 算法。

表 1 不同算法在测试集的检测率
Table 1 The detection rate of the algorithm in different datasets

序列	ETLD	MIL	TLD	Frag
girl	72	81.8	67	80.2
David	100	77.51	99.71	60.6
faceocc1	100	97.25	100	100
faceocc2	100	99.25	98.89	92.33
coke11	84.32	91.93	85.26	81.75
syvester	95.95	96.63	95.28	92.13
jumping	92.67	86.67	91.33	40.13
animal	87.8	43.9	68.29	7.1
football	98.61	28.06	79.44	80.33
shaking	84.85	49.64	74.66	57.85

注:黑体为最高值,斜体次之

2) football 序列 该序列集描述了橄榄球运动员奔跑过程和其他运动员交汇,碰撞和遮挡分离后的运动过程。主要用来测试各个算法对于多物体交汇过程出现背景交汇和遮挡后重定位的鲁棒性。跟踪前期 Frag 算法就开始出现漂移现象,而其他算法在多物体交汇依然可以跟踪到目标(图 6(c)#113, #154, #265)。但是在后期运动员发生碰撞遮挡后,ETLD 算法能够进行重定位,其他算法皆发生不同程度的漂移(图 6(c)#295)。

3) shaking 序列 该序列集描述了演唱会灯光效果急剧变换以及摇滚吉他手的头部频繁摇动。主要用于测试各算法对于光照突变下物体形变的跟踪鲁棒性。Frag 算法和 MIL 算法在灯光短暂突变下还能得到较好的目标跟踪,而 TLD 和 ETLD 检测失败(图 6(d)#62)。其主要原因是因为光照突变导致算法在线学习样本得不到及时更新,缺乏先验知识。但是后期光线缓和,姿态重现的情况能够进行重新定位(图 6(d)

#95), 而 Frag 和 MIL 算法后期跟踪出现了较大的漂移(图 6(d)#349)。

3.2.3 综合性能分析

1) 鲁棒性

ETLD 算法和 TLD 算法从实验来看,对缓和的光照变化、物体模糊、半遮挡、旋转、仿射变换和尺度变化具有较好的鲁棒性。而 MIL 和 Frag 在稳定环境下的跟踪效果较好,特别对于光线突变变化适应性较好。ETLD 对于两物体似然度较高的情况下误判率明显低于 TLD 算法。主要取决于 ETLD 算法对物体进行局部跟踪,排除与目标似然相差太远的跟踪物体,但是在对于光照的突变和快速旋转物体的鲁棒性较低。

2) 准确性

ETLD 算法首先采用了局部多尺度密集搜索策略,使得跟踪目标在前期能够更精确地得到定位,防止早期过多的引入背景等干扰引起后期跟踪发生漂移现象。TLD 算法在复杂环境跟踪容易发生漂移,导致跟踪模板偏差过大,从而容易发生后期跟踪失效(见图 5(f)、图 6(b)(d));其次,引入随机 Haar-like 块过滤器能够有效地提取具有纹理特征的跟踪物体,表 2 中 ETLD 的跟踪精度明显优越于 TLD 算法,同时在一些测试集上也明显高于 MIL 和 Frag 算法。跟踪精度误差在各序列集的对比参见图 7 和图 8。

表 2 算法在测试集的平均偏移
Table 2 The mean distance error of the algorithm in different datasets

序列	ETLD	MIL	TLD	Frag
girl	30.6	20.13	31.58	25.87
David	10.23	15.82	12.89	30.48
faceocc1	4.45	10.72	8.42	3.68
faceocc2	3.05	6.04	6.14	11.63
coke11	6.12	10.42	7.27	23.43
syvester	5.61	7.64	5.95	8.31
jumping	5.86	11.87	5.92	26.98
animal	3.33	20.3	5.26	40.64
football	6.77	45	10.14	9.81
shaking	20.21	22.25	18.26	17.92

注:黑体为最高值,斜体次之

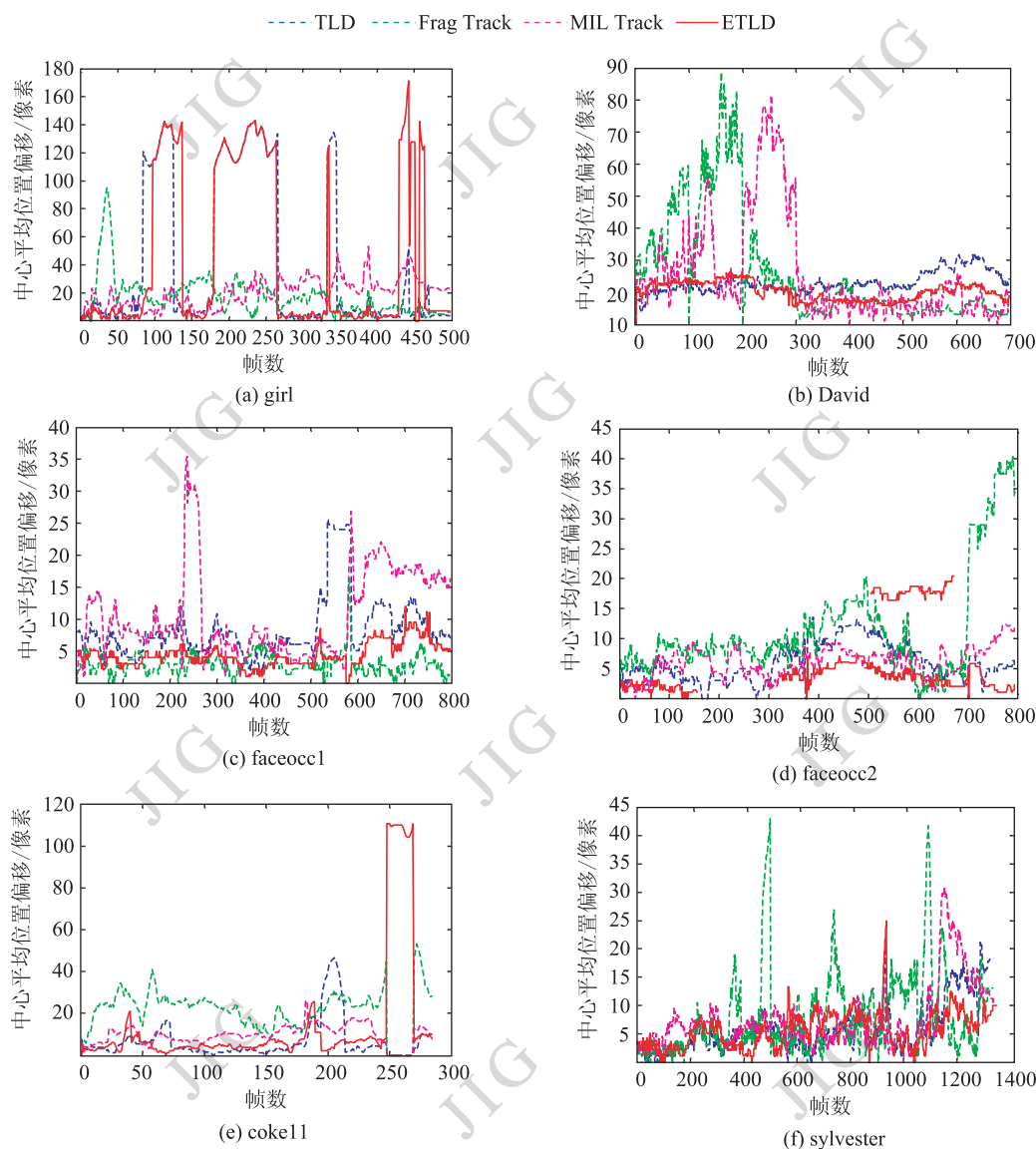


图7 各算法在稳定环境测试集的跟踪误差示意图

Fig. 7 Location error plots in stable condition datasets

3) 快速性

ETLD 算法的检测实时性能够适应于低分辨率和较高分辨率。(1) 该算法在早期子窗口块过滤添加了积分直方图过滤跟目标模板差异较大的窗口, 能够大量减少后面过滤器对非目标子窗口的特征匹配, 使得过滤器更加专注于区分较难分类的子窗口块; (2) 进行了局部尺度密集搜索策略, 从而针对小运动的物体避免了全局搜索的过度耗时。但是当物体在视频帧序列消失或者跟踪失败, 检测速度相对于 TLD 算法较耗时, 主要原因是需要对全局子窗口进行随机 Haar-like 块特征匹配和随机蕨特征匹配过程。图 9(a) 描述了 320×240 分辨率下 girl 序列

集(从第 90 帧到第 135 帧)发生人脸丢失后, ETLD 算法的检测帧数明显低于 TLD 算法的检测帧; 而 MIL 和 Frag 对于丢失目标无进行特殊的处理, 因此检测时间相对稳定。 640×480 分辨率(animal、football、shaking 集)下在不发生目标丢失的前提下, ETLD 检测速度也表现非常稳定, 而 TLD 检测速度明显检测非常耗时(参见图 9(b))。可见, ETLD 对于较高分辨率的检测速度不会下降很多。通过实验表明, ETLD 在分辨率为 320×240 图像序列集上, 其检测速度平均能够达到 15 帧/s, 在 640×480 分辨率下平均可以达到 14 帧/s, 基本能够达到实时检测。

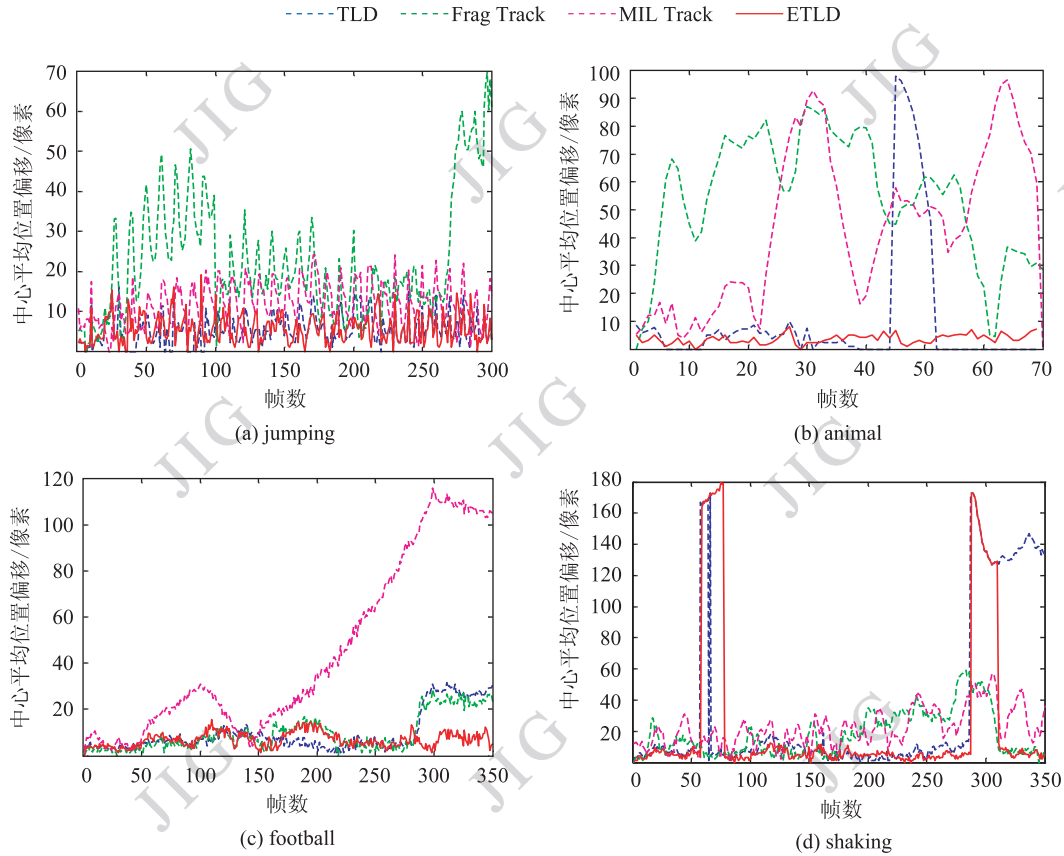


图 8 各算法在复杂环境测试集的跟踪误差示意图

Fig. 8 Location error plots in complex condition datasets

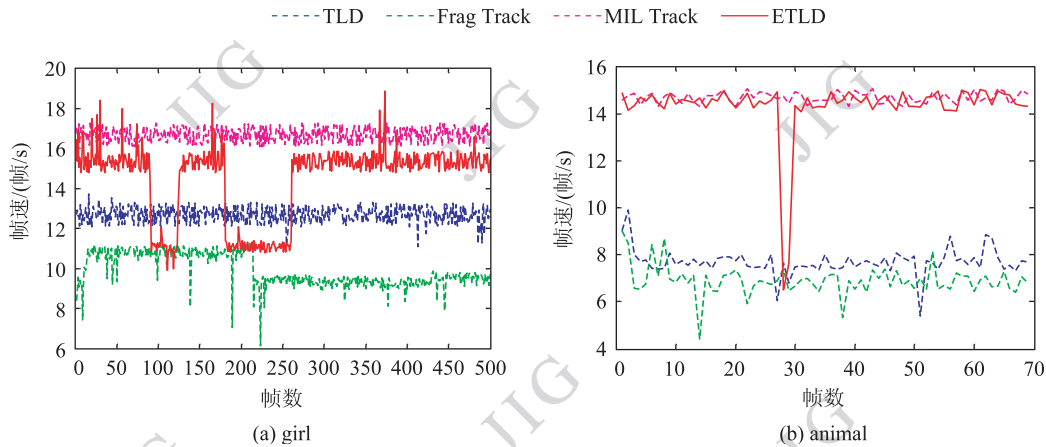


图 9 各算法对应测试集检测帧速

Fig. 9 Frame rate of different algorithms in corresponding datasets

3.3 参数分析

为了进一步分析随机 Haar-like 块过滤器对提高跟踪精度有一定的贡献,下面对级联过滤器进行限定:级联分类器 1) 仅采用随机滤波器(RF);2) 仅采用随机 Haar-like 块过滤器(RH);3) 采用随机 Haar-like 块组合而成的线性过滤器(RF&RH)。测

试集序列采用 David 序列,主要原因是该序列包含了光照变化、尺度变化和姿态变化。跟踪误差对比曲线如图 10。可见在较暗环境下(David 序列前 250 帧),若跟踪物体的纹理不明显,RH 的过滤性能表现较差,但是在尺度变化和纹理明显的物体跟踪 RH 的跟踪精度效果较好(参见图 10 David 序列 400 帧

以后的跟踪精度)。因此,RF&RH 组合而成线性分类器可以弥补 RF 在大尺度下的跟踪精度不足以及 RH 在小尺度、弱光、物体纹理性较差下的跟踪精度不足,从而综合两者的优点提高了跟踪精度。

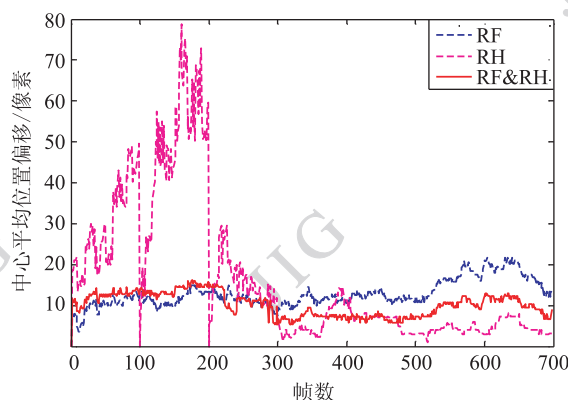


图 10 David 序列不同级联分类器的跟踪精度

Fig. 10 The accuracy of different cascade classifiers on David sequence

4 结 论

基于 TLD 跟踪算法的不足,提出了基于 Slding-window 方法的积分直方图过滤器,能够快速过滤掉非目标物体块,提高了其他过滤器的特征提取速度。同时采用了随机 Haar-like 特征块提取目标物体的纹理信息,特别对于大尺度下的物体具有相当高的检测精度,并结合随机藏分类器形成一个线性分类器,有效地解决了运动物体在复杂环境下的快速运动,旋转形变所引起的漂移现象等鲁棒性问题。通过实验对比了 3 个主流跟踪算法,从跟踪算法的鲁棒性、快速性和准确性 3 个方面进行了全面的分析,实验结果表明了 ETLD 算法级联分类器的有效性;搜索策略采用了局部和全局搜索方式,一定程度上提高了计算速度。但是本文算法在对于以下两个方面的检测效果较差:1)光线出现较大突变。主要原因是因为本文采用了在线学习的方法,其先验知识是由首帧确立待测目标样本且帧间进行样本提取进行学习,假如帧间出现严重的反光或者环境光变化(参见图 6 shaking 序列#62)将会导致帧间的目标的差异性增大,从而出现检测目标丢失的情况。相反地,假如帧间光线变化较慢,ETLD 算法的检测效果十分突出(参见图 5 david 序列#1~#400)。2)目标出现丢失后检测速度较耗时。鉴于本文采用了随机

Haar-like 块过滤器和随机藏过滤器相结合的方法,若目标出现丢失,全局搜索势必会加大计算的复杂度(参见图 9(a)(b)在 ETLD 算法检测失效下,检测帧数急剧下降)。今后将对此类问题的解决方案进行深入的研究。

参考文献 (References)

- [1] Kalal Z, Matas J, Mikolajczyk K. Online learning of robust object detectors during unstable tracking [C]//Proceedings of 12th International Conference on Computer vision workshops. New York: IEEE Xplore, 2009: 1417-1424.
- [2] Kalal Z, Matas J, Mikolajczyk K. Pn learning: Bootstrapping binary classifiers by structural constraints [C]//Proceedings of Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2010: 49-56.
- [3] Kalal Z, Mikolajczyk K, Matas J. Forward-backward error: Automatic detection of tracking failures [C] //Proceedings of 20th International Conference on Pattern Recognition. New York: IEEE Press, 2010: 2756-2759.
- [4] Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram [C]//Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2006: 798-805.
- [5] Klein D A, Schulz D, Frintrop S, et al. Adaptive real-time video-tracking for arbitrary objects [C]//Proceedings of International Conference on Intelligent Robots and Systems. New York: IEEE Press, 2010: 772-777.
- [6] Zhang K, Zhang L, Yang M-H. Real-time Compressive Tracking. Computer Vision-ECCV 2012 [M]. New York: Springer, 2012: 864-877.
- [7] Gu S, Zheng Y, Tomasi C. Efficient visual object tracking with online nearest neighbor classifier [M]. New York: Springer, 2011: 271-282.
- [8] Grabner H, Grabner M, Bischof H. Real-time tracking via online boosting [J]. Proceedings of the British Machine Vision Conference. Edinburgh, UK: British Machine Vision Association 2006, 1(5): 4-7.
- [9] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632.
- [10] Hare S, Saffari A, Torr P H S. Struck: Structured output tracking with kernels [C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 263-270.
- [11] Yilmaz A, Javed O, Shah M. Object tracking: a survey [J]. ACM Computing Surveys, 2006, 38(4): 13.
- [12] Yang H, Shao L, Zheng F, et al. Recent advances and trends in

- visual tracking: a review [J]. *Neurocomputing*, 2011, 74(18): 3823-3831.
- [13] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [14] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking [J]. *International Journal of Computer Vision*, 2008, 77(1-3): 125-141.
- [15] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [C]//*Proceedings of the 7th International Joint Conference on Artificial Intelligence*. Vancouver, Canada: Citeseer, 1981: 674-679.
- [16] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//*Proceedings of Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*. San Diego, CA: IEEE 2005, 1: 886-893.
- [17] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C]//*Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Washington DC: IEEE, 2001, 1: 511-518.
- [18] Porikli F. Integral histogram: a fast way to extract histograms in cartesian spaces [C]//*Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington DC: IEEE 2005, 1: 829-836.
- [19] Lowe D G. Object recognition from local scale-invariant features [C]//*Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra: IEEE, 1999, 2: 1150-1157.
- [20] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up Robust Features. *Computer Vision-ECCV 2006* [M]. New York: Springer, 2006: 404-417.
- [21] Alahi A, Ortiz R, Vandergheynst P. Freak: Fast retina keypoint [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence RI: IEEE 2012: 510-517.
- [22] Ozuysal M, Fua P, Lepetit V. Fast keypoint recognition in ten lines of code [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN: IEEE 2007: 1-8.
- [23] Kwon J, Lee K M. Visual tracking decomposition [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, CA: IEEE 2010: 1269-1276.
- [24] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C]//*Proceedings of 2011 IEEE International Conference on Computer Vision*. Barcelona, Spain: IEEE, 2011: 2564-2571.
- [25] Santner J, Leistner C, Saffari A, et al. Prost: parallel robust online simple tracking [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* San Francisco, CA: IEEE 2010: 723-730.
- [26] Tomas Vojir. Tracking Dataset [EB/OL]. 2013-02 [2013-06-06] <http://cmp.felk.cvut.cz/~vojirtom/dataset/index.html>.